

农业学术英语语料库建设构想

范晶晶, 李丽霞* (西北农林科技大学外语系, 陕西杨凌 712100)

摘要 近年来, 语料库语言学已成为语言学发展的热点之一, 在各个领域呈现快速发展趋势, 建设一些小型专业化语料库也是近年来研究者的焦点。农业英语语料库作为 ESP 语料库的分支, 也应被广泛关注。提出农业学术英语语料库的建设构想, 并从语料收集、文本清洁、检索和标注等方面进行介绍, 以期为进一步的深层次研究作铺垫。

关键词 农业学术英语; 语料库语言学; 专门用途英语

中图分类号 S126 **文献标识码** A **文章编号** 0517-6611(2014)07-02169-02

The Construction of Agricultural Academic English Corpus

FAN Jing-jing, LI Li-xia (Department of Foreign Languages, Northwest A&F University, Yangling, Shaanxi 712100)

Abstract Nowadays, corpus linguistics has been one of the hot topics in the world, improving rapidly in many fields. Many researchers focus on constructing some small and specific corpora. As a branch of English for special purposes corpus, agricultural English corpus is believed to arouse more and more attention. The possibility of the construction of agricultural academic English corpus was discussed, the collection of texts, retrieval and tagging of them were introduced, so as to make preparation for the further study.

Key words Agricultural academic English; Corpus linguistics; ESP

语料库语言学是以计算机为载体存放真实语言材料的集合, 它是收集数据的重要手段。自 20 世纪 80 年代语料库语言学被引入我国以来, 国内已有大批学者参与语料库建设并开展语料库相关研究。自 2006 年开始, 外研社在北京开展高校英语教师暑期研修班, 每年举办两期语料库与外语教学培训班, 并且专门聘请梁茂成、李文中和许家金老师进行现场讲授。这些推动了语料库在我国的普及和发展, 也推动了小型专门用途语料库建设的热潮。

专门用途语料库是指出于某种特定研究目的需要, 只收集某特定领域的语料样本建成的语料库^[1]。因其特殊性、针对性强、专业化程度高、语料加工深入等特点, 专门用途语料库比通用语料库等大型语料库更适用于专业化的英语研究。为此, 笔者提出农业学术英语语料库的建设构想, 以期为进一步的深层次研究奠定基础。

1 专门用途英语语料库的建设

专门用途语料库作为一种和特定专业或者学科相关的语料库, 具有重要的地位, 而其发展与专门用途英语的发展是紧密相关的。在 Hutchinson 的语言教学树型图上, 它和通用英语占有相同地位, 主要分支有科技英语、经贸英语及社科英语等。1964 年, 韩礼德在《语言科学与语言教学》一书中提到关于“专门用途英语”的概念, 指出其范围包含公务英语、警察英语、法官英语、护士英语、农业专家英语和工程师英语等。香港理工大学将专门用途语料库进行分门别类, 细化为学术类、商务类、新闻类、法律类、文学类等。不同类别的专用语料库在词汇使用、语法规则以及语篇构建等方面都存在或多或少的差异, 这也是将建设农业学术英语语料库提上日程的原因之一。

国外很少提及专门用途英语语料库, 主要有英国的 The

Reading Academic Text Corpus 和 British Academic Written English Corpus; 美国的 Michigan Corpus of Academic Spoken English 等。从语料库语言学引入我国至今, 我国已建成的专门用途语料库主要包括最早“替代”国内建成并用于研究的上海交通大学 JDEST 语料库、解放军外国语学院建成的“军事英语语料库”、黑龙江大学创建的“商务英语语料库”、大连海事大学建成的“海事英语语料库”, 以及研究者为开展研究所建立的很多小型语料库, 如“雅思与六级阅读材料语料库”和“小型教学型翻译语料库”等。基于已建成的专门用途语料库所开展的语言研究主要涉及商务、医学、航空、水利、法律、石油、海事英语、IT 英语等领域, 并在翻译研究、词汇和语法研究方面有重要发展。其中, 基于语料库的实证研究目前已经成为二语习得和外语教学研究领域的一大亮点^[2]。

2 农业学术英语语料库的构建

国内现已建成的专门用途语料库多围绕医学、商务、法律等, 鲜有涉及农业的学科专业语料库, 更没有农业学术英语语料库。该研究根据农业院校的专业特点, 以期建成一个农业学术英语语料库。下面从几个部分对其构建进行阐述。

2.1 建设农业学术英语语料库的目的 建设一个农业学术英语语料库将填补国内涉农学科专业语料库的空白, 丰富专门用途语料库的研究内容, 对提高农林院校专业英语教学水平, 提高师生和科技人员的学术英语写作水平, 提升国际学术交流水平, 特别是提高在外文期刊发表论文的成功率, 以及服务于农业科技英语翻译等方面都具有重要的现实意义和应用价值。

2.2 农业学术英语语料库的语料收集 在语料库建设成型之前, 语料的收集十分重要, 其中包括文本、格式、字符和标点的获取等。语料库的建成不是简单的语料相加或累计, 而是依靠各种计算机软件进行整理和加工, 还需删掉干扰信息, 从而确保语料的准确。建设语料库需注意语料库的代表性、规模和内容。通常情况下, 语料应尽可能地涵盖跨度广的范围。从理论上讲, 语料库越大则得到的数据越可靠, 但是很多情况下建成一个规模庞大的语料库是不切实际的。

基金项目 2013 年西北农林科技大学人文社科重点项目 (2013-RWZD05)。

作者简介 范晶晶 (1990-), 女, 陕西杨凌人, 硕士研究生, 研究方向: 二语习得。* 通讯作者, 教授, 硕士生导师, 从事二语习得、应用语言学及语料库语言学研究。

收稿日期 2014-02-10

“语料库的选取通常是在研究目的、语料库的代表性以及语料收集的可操作性之间找一个平衡点来进行的^[3]。

按照建库的3个原则,农业学术英语语料库拟建设园艺专业和动物科学专业两个子库。学术英语语料库的语料主要来源于网络资源,利用国内外数据库下载相关文献。所选文本为英文国际知名期刊全文摘录,文章作者需为英语本族语者,文本选取年限为2007~2013年,拟定各子库文本数量约100篇,库容量超过100万词。所下载文本需按照所占比例进行合理统计,并且按照主题、期刊名称和时间范围进行整理归类。最终还需去除文本中的摘要、表格、公式、致谢等干扰信息,保存为纯文本格式。

2.3 农业学术英语语料库的文本重命名和元信息添加 文本的命名可依照固定模式用字母或数字区分。例如,在该语料库中,园艺专业的文章可记为A类,动物科学专业的文章记为B类。而园艺专业中的果树方向、蔬菜方向和园艺设施方向分别用A、B和C表示,后面添加期刊代码(可用期刊名称缩写代替)以及文章编号。例如,选自刊物《Hortscience》中的第35篇有关苹果嫁接的文章可以重命名为AAH035。元信息主要包括作者姓名、文章标题、文本来源、涉及主题等。以文章题目的元信息添加为例,按照<title> </title>的格式附加在文本开头,利用前后两个尖括号将文本的主要信息括起来。开展语言研究时,文本元信息会辅助文本的查找和检索。

2.4 农业学术英语语料库的文本标注 只有经过标注,有了附加信息的语料库才能称为有研究价值的语料库。利用标注软件Treetagger对清洁后的文本进行词性附码标注:导入未经附码的生语料库并选择Treetagger词性附码集,由计算机批量完成附码过程并最终将语料保存为3种格式,分别是生语料格式(.txt)、经过附码的语料(.pos)和添加元信息和附码后的语料(with_metadata.txt),以方便后期研究开展。为及时发现机器批量标注产生的错误,确保数据的准确性,利用BSFU Qualitative Coder 1.1软件进行检查和手工标注。

2.5 语料库的工具选择 构建语料库需要用到很多工具和软件,其中最重要的是“文本处理器”清洁文本软件和Treetagger标注工具。前者能够剔除文本从PDF转换为纯文本格式之后出现的全角半角混用,多余空格空行以及其他格式问题;而后者能够批量添加词性附码,相当于增加文本的研究价值,为深层次的研究奠定基础。

3 农业学术英语语料库的应用研究

学术词汇的掌握情况是衡量学习者能否顺利完成学术英语学习的一个重要指标^[4]。2005年,吴瑾等对上海交通大学研究生在阅读学术类文章时遇到的困难进行调查,结果显示,24%的困难源自学生对学术类词汇掌握不够。农业学术

英语语料库作为涉及农林专业的专门用途语料库,利用有效的计算机软件对搜集到的语料进行加工,不仅可以从宏观角度对农业英语的语言特点进行描述,而且可以研究微观角度中词汇的使用频率、涉及范围以及分布情况等,利用生成的农业专业词表辅助教材编写、大纲设计和教学测试,从真实语料库中获得的文本信息对学生掌握教材中的专业词汇十分有用。在语言教学应用中,可以将农业学术英语语料库应用到课堂范围中,让学生作为活动和研究的主体,探究国外文献中的常用词块、语法及语篇特点,从而更好地把握专业英语特点,服务于学术写作和翻译。

4 农业学术英语语料库建设中可能出现的问题

(1)基于语料库语言学的发展和各种检索软件的出现,个人通过自建语料库来研制词表成为潮流和趋势^[5]。而国内众多专门用途语料库中没有涉农的语料库,因此研究者只能根据自身研究目的进行建设。但自建语料库不仅耗时费力,还可能和已建成的专业语料库存在很大差距,这些差距主要体现在语料库建设中文本转化和清洁过程中。

(2)要建立语料库,仅仅有语料还不够,海量的文本信息必须依靠特定的检索软件进行分析处理才能够得到有用的信息,否则只是文本的简单集合而已。在语料处理中,文本从PDF格式转化为TXT纯文本格式时所出现的乱码以及格式错误是否会影响所得结论的准确性仍是要面对的难题之一。

(3)在进行机器标注和手工标注时,利用Treetagger等软件进行的赋码是否正确,手工检查和修改之后是否还存在明显错误等或标注集使用不一致等。这些在构建农业学术英语语料库过程中可能存在的问题都需要逐步解决。

5 农业学术英语语料库的前景展望

农业学术英语语料库的建设是一项十分庞大的任务,与此同时它也能给语言与教学研究带来极大便利。Sinclair在2003年的语料库语言学国际会议上就指出大型语料库建设的势头已缓,取而代之的应该是大批小型语料库的兴起^[5]。那么农业学术英语语料库的建设很可能会推动农业英语的深入学习,成为优化农林院校大学英语教学课程设置的催化剂。虽然在语料库建设中可能会遇到文本处理不彻底或计算机软件操作不当、文本选择不当等种种问题,但是却无法忽视该语料库可能为农业英语研究所带来的重大作用。

参考文献

- [1] 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.
- [2] 王利莉,李丽霞.中国语料库语言学研究的最新动态[J].湖南科技学院学报,2013(7):170-172.
- [3] 梁燕,冯友,程良坤.近十年我国语料库实证研究综述[J].解放军外国语学院学报,2004(6):53-57.
- [4] KUHN P. Assessment of academic literacy skills: Preparing minority and limited english proficient (LEP) students for post-secondary education [M]. Fresno, CA: California State University, 1996.
- [5] 崔维霞,王均松.基于语料库的词表创建原则及方法研究[J].中国科技术语,2013(2):15-19.

(上接第2119页)

参考文献

- [1] 朱颜.中药的药理与作用[M].北京:人民卫生出版社,1958:231.
- [2] 黄冠文,薛卫萍.分光光度法测定维生素B₂片含量方法改进[J].药物分析杂志,1995,15(5):167.
- [3] 李徽,帕丽达·买买提.维生素B₂片含量的测定方法[J].中国临床医

- 药研究杂志,2003(11):95.
- [4] 李来生,张江.反相HPLC测定食品中水溶性维生素含量[J].分析测试学报,1997,16(3):257.
- [5] 江苏新医学院.中药大辞典(下册)[K].上海:上海人民出版社,1997:1606.
- [6] 张淑云,王志平,张继征.莜麦中维生素B₂的分子荧光法测定[J].集宁师专学报,2008,30(4):35-37.