

# 基于 ESP 的多模态语料库构建与应用

——以医学英语为例

蒋云磊, 康光明

(山东农业大学, 山东 泰安 271000)

**摘要:** 多模态语料库作为一种新型计算机辅助语言研究与学习工具, 可以客观真实反映语言在使用过程中的全貌。对于促进专门用途英语教学、教师教育、多模态话语分析研究具有巨大应用价值。以医学英语为例, 综述了语料库的相关理论和多模态语料库已有的研究成果, 探索分析了借助 Elan 软件构建多模态语料库的过程, 最后对这一小型自建多模态语料库的应用进行了多维探讨。

**关键词:** 多模态语料库; 专门用途英语; Elan; 医学英语教学

中图分类号: H319

文献标识码: A

文章编号: 1671-931X (2017) 06-0053-05

53

## 一、引言

第一代计算机语料库出现于上世纪 60 年代, 半个多世纪以来现代意义的语料库语言学取得了长足发展, 世界语言学界的研究方向也逐步转向在语料库的基础上进行语言研究。然而, 文本型语料库存在诸多弊端, 比如时效性差、真实性不足和内容不丰富等, 很难达到信息时代语言研究对语料来源的要求。因此, 基于计算机、网络和新型语言研究工具的多模态语料库为数据驱动语言学习的研究带来了新的契机。

专门用途英语 (English for Specific Purposes, ESP) 教学是目前大学公共英语课程改革的方向。专门用途英语既包含一定的英语语言知识和运用该语言知识的技巧, 还包括某一特定领域的专业英语和专业知识。ESP 教学语篇具有专业性和多样性的特点, 很多教学信息是以图解注释、产品介绍、市场推广的语篇形式出现, 这些特点使得多模态教学更具

实用性。同时也对英语教育从业者提出更高的要求。

## 二、建设多模态语料库的理论基础

### (一) 系统功能语言学理论

从语言学的意义上来说, 多模态语料库多使用于多模态话语分析。多模态话语分析是现代篇章语言学的研究重点, 虽然受到皮尔斯等人符号学理论的影响, 吸取了其中媒介、对象、解释三位一体等理念, 但其主要理论基础还是韩礼德创立的系统功能语言学 (朱永生, 2007)。系统功能语言学理论 (system-functional linguistics theory) 认为语言除了意义潜势 (meaning potential) 和社会符号 (social semiotic) 以外, 其他符号系统也同样意义丰富而重要; 因此多模态话语也具有多功能性, 即概念功能、人际功能和语篇功能。

### (二) 建构主义学习理论

建构主义学习理论 (constructive learning

收稿日期: 2017-11-21

作者简介: 蒋云磊 (1981-), 男, 山东莘县人, 硕士, 泰山医学院讲师, 研究方向: 语言学理论与应用、文学与翻译学; 康光明 (1962-), 男, 山东荣成人, 山东农业大学教授, 研究方向: 理论语言学及应用、语篇分析。

theory)认为,学习是一个与情境联系紧密的、积极主动的活动,在学习过程中,知识、内容、能力等不是被动地吸收或训练,而是被建构,学生主动建构知识,而非被动接受,教师的角色也发生了根本变化,从教学的中心转变成学习的指导者和学习环境的创设者(温彭年,2002:18)。传统意义上的外语教学以传统的教具支持,缺乏现代多媒体信息技术的支持,教师们在教学过程中创设外语学习环境,会遇到诸多困难。如果运用多模态语料库来支持语言的教学和学习,真实客观的语境重新复现,有利于提高语言学习的自主性、建构性和兴趣性。

### (三)情景认知理论

知识与活动是不可分割的,学习者在真实的情境中获得知识和解决实际问题的经验和技能,从而不断地建立起自己的新的知识体系,这是情景认知理论(situated cognition theory)的主要观点。也就是说知识是情境化的,学习者获取与应用知识取决于真实的情境。对于语言学习来说,真实语境能够有效地促进语言的认知与学习,这也是构建多模态语料库驱动语言学习的重要宗旨。

### (四)多元识读教学理念

多元识读理论(multiliteracies theory)认为当社会文化相互渗透、科学技术快速发展的今天,通讯渠道和多媒体呈现多重性,文化和语言的多样化也日益突出。人们相互之间交流是通过一系列丰富的符号进行的,而每一种符号都具有一定的规律性,它们在表达意义时与语言是相近似的(Kress,2000:199)。语言学家胡壮麟教授认为多元识读能力包含两种,即文化识读和技术识读,而其中多模态识读就是指后者,识读参与者能在网络信息环境下学习、研究、工作并构建内在意义等(胡壮麟,2007:8)。多模态语料库支持学生根据特定学习任务或自身兴趣进行多模态信息检索(concordance),来提高语言的产出能力,与此同时也提升跨文化交际能力。

## 三、语料库研究综述

### (一)中外传统语料库技术发展现状与应用

#### 1. 中外传统语料库技术发展现状

语料库(Corpus)作为一种语言资料库,通过概率统计的方法,把大量悉心采集而来的真实文本为研究素材,来得出研究结论,为研究者提供了一个分析、研究、描述和应用语言的载体。现代语料库的起步于十九世纪中期的“英语用法调查”(Survey of English Usage)语料库,此语料库由伦敦大学创建,开辟了语言研究的新时代。随着计算机技术的飞速发展,规模更大的语料库也相继诞生。与此同时,大量语料库检索软件的出现为语言研究提供了多种工具,比如 Concordance、VocabProfile、MicroConcord、TACT、Lexa、X-tract、Wordsmith Tools 等,研究者可

根据实际情况和使用习惯选择不同的检索软件。

中国最早的英语语料库是上海交通大学科技英语语料库(JDEST),由黄人杰和杨惠中于1986年共同主持构建。随后,我国相继建设了多个各类语料库,满足了英语教学、语言研究等需要。同时,符合我国语言研究特色的语料库辅助教学软件也陆续开发出来,如 Cooltomatoes 和 PatCount 等。

#### 2. 传统语料库的应用

上面提到的中外语料库都属于文本型语料库,通过借助宏大的计算机及互联网技术,为编写辞典、语法书及各种教材提供了重要的语料源,而且为语言研究和教学提供了新的途径。

在1991年Tim Johns提出数据驱动学习(Data-Driven Learning,DDL)的理论和方法,语料库用于外语教学的研究和实践中取得了较大进展,研究的内容包括DDL和语料库在语言教学中的优越性和潜在作用(Flowerdew,1993;Karpati,1995)、语料库支持下的英语教学新模式探讨(Boulton,2010)以及应用效果的评价与讨论(Perez-Paredes et al.,2011)等。与此同时,我国的语料库辅助外语教学应用研究也取得了一定的成就,比如介绍国外研究理论(李文中、濮建忠,2001;甄凤超,2005;崔维霞,2013)、英语教学模式改革尝试(施称、章国英,2015)及小型语料库的构建与应用(傅顺、罗永胜、杨劲松,2013;熊亭玉,2013;赵丽珠,2014;章国英,2014)等,有一定的推广价值。

### (二)专门用途英语(ESP)语料库综述

语言规律的复现及词汇、语法特点是传统语料库研究重点,这一特点有助于研究构建通用语法和作用于词典编纂,但是它社会学属性的不足,在语篇分析层次的表现差强人意。而专门用途英语语料库则弥补了这一不足,不仅包括特定的英语语言知识,还包括使用该语言技巧能力,同时能够充分把握所处的语言环境。另外,ESP语料库在词典编纂和研究领域越来越受到认可和重视。因为它不仅能够为编纂词典者提供大量真实例句及其语境,而且能大大提高词典编纂的工作效率,也极大地方便了词典的修订与增补。

### (三)建设和应用多模态语料库研究

有关多模态语料库的建设与研究始于上个世纪末期,欧洲多国于2008年共同建成SACODEYL(system for the assisted compilation and open distribution of European youth language)多模态教学语料库,这是全球首个将数据驱动学习(Data Driven Learning)付诸实践的多语种多功能的多模态语料库(潘璠,2012:179)。

我国的多模态语料库建设和研究虽然起步较晚,但发展较快。基于多模态语料库的多模态语言学研究(李战子,2003;朱永生,2007;代树兰,2013;胡

壮麟,2006,2007,2011;顾曰国,2010;张德禄,2009,2011;王立非、文艳,2008;汪燕华,2011;李斌,2012;黄国文,2005)、教学研究(张德禄,王璐,2010;王彤,2001)等尚在可行性探讨和模型设计阶段。

本文基于 ESP 多模态语料库的探讨将以医学英语为例,以针对医学英语的教学特点、内容和要求,构建了支持医学术语关键词、语音、主题、语境等多维检索的小型多模态语料库,并初步尝试其在医学英语教学中的应用,探索构建多模态语料库支持语言研究和教学。

四、建设医学英语多模态语料库

(一)医学英语多模态语料的筛选标准与库的规模探讨

医学英语多模态语料库的使用者多为医学英语教师、医学相关专业教师、基础医学研究者和高年级医学相关专业研究生、本科生等。根据医学英语的教学内容、医学涵盖领域及学生的学习基础与兴趣,在构建小型多模态语料库的语料选择中除了根据语料库建库的基本原则(针对性、代表性和规模性)之外,又需要兼顾以下 2 个特点:

- 1.全面性:兼顾医学相关尽可能多的学科,比如生理学、病理学、生物学、解剖学、影像诊断学、医学伦理学等;
- 2.适用性:临床医学语言尽量做到实用地道,考虑语言交际的具体场景和文化内涵。

依照前述建库原则,笔者共选取了总长约为 120 小时左右、约 220G 的视频材料和音频材料,建设医学英语多模态语料库,其中美国知名大学的医学相关教学公开课约占 40%,40%为医学题材的影视作品如《豪斯医生》、《扪心问诊》、《实习医生格蕾》、《嗜血法医》、《急诊室的故事》等,当然还包括语音类的预料,约占总语料的 20%,比如 VOA 健康报道、BBC 纪录片中的生物医学部分等。

(二)多模态预料标注与检索方法

目前常用的多模态语料库标注与检索软件主要包括 MCA、Elan、DRS、ANVIL 等,EUDICO Linguistic Annotator(ELAN)这款多模态话语分析软件,由位于荷兰奈梅亨市的马普心理语言学研究所设计开发的,完全免费开源,自 2011 至今已发布多个版本,目前最新版本为 Elan4.6.4,用户可根据自己不同的操作系统到 <https://tla.mpi.nl/tools/tla-tools/Elan/> 免费下载。对于多种视频音频格式,如 \*.mpg、mpeg、wav、mp4、mpg4、mov、qt、wmv、wma、asf 等,Elan 软件都可以进行播放、转写和多层结构标注。因此,本文选用 Elan 软件对多模态语料进行标注和检索。

Elan 软件中共划分了 5 种语言学类型:NONE 层、Time Subdivision 层、Symbolic Subdivision 层、Included In 层和 Symbolic Association 层。这样可以

依据多模态语料库的应用环境灵活地选择标注和检索,进而扩大了应用范畴。根据自身需要自定义不同类型的语料单元,大大提高了标注准确性,同时对语料库标注信息的管理也更加便捷。Elan 产出扩展名为.eaf 的标注文件,内含视频音频片段和标注信息,同时支持标注信息的多种显示方式,还可以随时回放并可控制播放速度。

(三)基于 Elan 建设小型多模态语料库

1. Elan 软件的安装

登录 Elan 官方网站 <https://tla.mpi.nl/tools/tla-tools/Elan/download/> 下载。安装程序和其他主流软件类似,依次点击“next”进行安装。软件使用前,建议下载 ELAN -Linguistic Annotator version 4.9.4 manual,阅读并熟悉 Elan 软件的详细使用方法。最新版本的 Elan 4.9.4 已经由中国湖南师范大学国际汉语文化学院李斌教授进行了简体中文汉化,安装时会自动选择简体中文。

2.载入多模态语料

激活并运行主程序文件 Elan.exe,笔者发现此软件界面友好、提示清晰、菜单层次明确,由菜单栏、标注区域和视频播放窗口组成,菜单栏和其他主流软件高度契合,主要由文件(File)、标注(Annotation)、编辑(Edit)类型(Type)、查找(search)、层级(Tier)、浏览(View)、窗口(Window)、选项(Options)、帮助(Help)等组成。点击“新建”按钮,载入未标注的多模态语料,完成多模态语料的载入。

3.多模态语料的分割及字幕转写

多模态语料载入后保存文件,默认层名“Default”,更改为“医学英文字幕”。然后在“选项”下拉菜单中找到“分割模式”,使用回车键对多模态语料进行分割标记,然后使用“选项”下拉菜单中的“转写模式”,对进行多模态语料的英文字幕转写。

4.多层标注

多层标注是多模态语料处理的核心环节。对多模态语料进行标注,首先选择“选项”下拉菜单中的“标注模式”。本文根据医学英语多模态语料库的教学应用设计,对多模态语料进行了主题层、情境层、音标层、词汇层等标注,如表 1 所示。

表 1 多层标注多模态语料

Tier/层	Linguistic Type/语言学类型	Annotation Info./标注信息
1	None	EN-Subtitle/英文字幕
2	Time Subdivision	Vocabulary/词汇
3	Symbolic Association	IPA/国际音标
4	None	Cultural Scene/文化场景
5	None	Theme/主题

5.在 Elan 软件中对多模态语料库检索与播放

对多模态语料进行标注结束后,使用“搜索”项可进行检索,Elan 的查找功能比较强大,可以检索



单个或多个.eaf文件,也可以检索单词的各种屈折形式,即正则表达式搜索,也可以进行速度自主控制的播放、停顿和回放,同时支持多种形式结果输出。

## 五、基于 ESP 的多模态语料库的应用

### (一)应用于语言教学

在医学英语专业术语教学、文献阅读、中英医学

术语翻译以及专业论文写作中,多模态语料库可以有力支撑教师们的课堂教学和同学们的自主学习,进而提升受教育者医学英语综合应用能力和课堂授课效果。反过来,教学实施者在课堂教学等实践中不断总结 ESP 多模态语料库的使用情况,并对语料库进行扩充和更新,从而更好地服务于语言教学(如图1所示)。

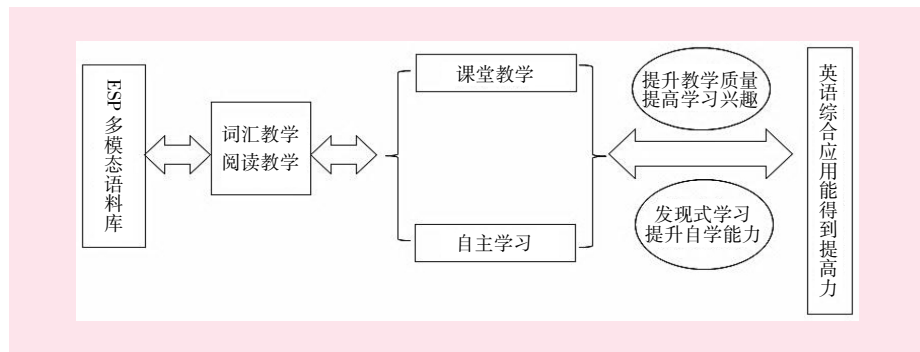


图1 多模态语料库辅助医学英语教学的模式

医学英语在词汇方面有其非常明显的特点,比如专业词汇多、日常词汇的医学释义、缩略语运用广泛、冠名术语应用广泛(熊亨玉,2013),在医学英语词汇教学中运用多模态语料库符合上述特点。多模态语料库也支持检索的音视频播放,使得语言使用情景得以再现,教学语境因此变得真实生动。还可以帮助受教育者对不同语料中词汇的使用情况进行比较,增进其对词汇语义、搭配和语用等的理解。

培养学习者的医学相关文章的阅读能力是医学英语教学的一个重要目标。多模态语料库中有丰富的语料,如医学公开课、剧集、纪录片等音视频资料,能够帮助医学英语学习者熟悉相关医学主题下的常用词汇、句式等。

### (二)应用于教师教育

基于 ESP 多模态教学语料库在教师教育中也有着重要的应用价值。帮助教师们在教学方法上从实践中教学,在教学中实践;在教学过程上,从教师讲授转变为任务性教学;学生和教师的角色也会发生变化,学生从被动学习到主动实践及反思,教师从主讲到主导;多模态语料库中有的语料来源于学生教学又应用于学生教学,既是结果又是材料,反复循环,多次验证。

### (三)应用于多模态话语分析

多模态话语分析使得话语意义的解读更准确全面,以文本为基础的话语分析研究并不能反映话语的全貌,Saussure 在 1916 年明确指出,语言学真正须要研究的对象是实际使用的话语。由于以前科技水平的局限性,话语分析主要集中在文本以内。随着计算机技术、信息加工及存储、多媒体网络技术、多模态语料库以及话语分析软件的更新发展,对自然状态下话语进行多模态分析成为可能,且是将来研

究的重要趋势。胡壮麟(2007)指出,多模态话语分析既可以加深对语言学的研究,也可以增加对语言符号学的认识。

## 六、结语

多模态语料库在 ESP 教学和多模态话语分析等领域中有广阔的研究和应用空间。多模态语料库能够为专门用途的英语教学提供了真实客观的话语材料和多元语境,在教育教学中,切实做到了“学生为主,教师为辅”,对于推进专门用途英语教育教学改革具有重大意义。本文通过分析专门用途英语和多模态语料库理论,综述了国内外的语料库、多模态语料库的建设和语言教学应用;以医学英语为例,分析讨论了如何借助 Elan 软件建设基于 ESP 的多模态语料库;然后多维探讨了这一小型自建多模态语料库的应用。希望更多教育者和语言研究者关心关注多模态语料库辅助的专门用途英语教学和语言研究。

### 参考文献:

- [1] Boulton, A. Data-driven learning: Taking the computer out of the equation[J]. Language Learning, 2010, 60(3): 534-572.
- [2] Halliday, M. A. K. An Introduction to Functional Grammar [M]. London: Arnold, 1994.
- [3] Kress, G. Multimodality [A]. in Cope, B. & Kalantzis, M. (eds.). London: Routledge, 2000: 199.
- [4] Perez-Paredes, P., M. Sanchez-Tonel, J. M. A. Claeo, & P. A. Jimenez. Tracking learners' actual uses of corpora: Guided vs. non-guided corpus consultation [J]. Computer Assisted Language Learning, 2011, (24): 233-253.
- [5] 顾曰国. 多媒体、多模态学习剖析[J]. 外语电化教学, 2007, (2): 3-12.

- [6] 胡壮麟,董佳.意义的多模态构建[J].外语电化教学,2006,(6):3-12.
- [7] 李战子.多模式话语的社会符号学分析[J].外语研究,2003,(5):1-8.
- [8] 李耕砚.新形势下的大学英语教学改革:挑战与对策[J].遵义师范学院学报,2016,(2):143-145.
- [9] 张德禄.多模态话语分析综合理论框架探索[J].中国外语,2009,(6):24-30.
- [10] 张振虹,何美,韩智.大学公共英语多模态语料库的构建与应用[J].山东外语教学,2014,(3):50-55.
- [11] 章国英.专门用途医学英语语料库构建的意义及实践[J].中国医学教育技术,2014,(4):146-149.
- [12] 郭慧敏.大学英语词汇教学行动研究[J].遵义师范学院学报,2016,(1):131-133.

[责任编辑:许海燕]

## Construction and Application of Multimodal Corpus Based on ESP—Taking Medical English as an Example

JIANG Yun-lei KANG Guang-ming

(Taishan Medical University, Tai'an 271000, China)

**Abstract:** As a new type of computer aided language research and learning tool, multimodal corpus can objectively and truly reflect the whole picture of language in use. It has a great application value for promoting special-purpose English teaching, teacher education and multi-modal discourse analysis. This paper takes medical English as an example, summarizes the relevant theories of corpus and the existing research results of multimodal corpus, explores and analyzes the process of constructing multimodal corpus with Elan software, and finally analyzes the progress of this small self-built multi-modal corpus Application of multi-dimensional discussion.

**Key words:** multi-modal corpus; ESP; Elan; medical English teaching

(上接第 52 页)

## Implementation of SPOC Teaching Mode in the Teaching of Higher Vocational Microcomputer Technology Courses

HUANG Hai-yan

(Zhejiang Guangsha Construction Vocational and Technical College, Dongyang 322100, China)

**Abstract:** By adopting SPOC teaching mode and using blue ink cloud class as the platform of SPOC online teaching, the reform of higher vocational SCM teaching has been carried out. The results show that both the enthusiasm and initiative of students in new teaching mode significantly improved.

**Key words:** SPOC teaching; microcomputer technology courses; effective classroom