

# 地方性旅游景区翻译语料库的创建与应用

谭兴 石婕妤

(贺州学院 外国语学院, 广西 贺州 542800)\*

**摘要:**论述以广西贺州市为例的地方性旅游景区翻译语料库的设计思路 and 具体创建步骤,包括语料的收集、语料的数字化、语料的对齐和语料的标注等,探讨地方性旅游景区翻译语料库在推动旅游翻译教学与研究的应用前景。

**关键词:**地方性;旅游景区;翻译;语料库

**doi:**10.3969/j.issn.2095-5642.2014.010.079

**中图分类号:**H059

**文献标志码:**A

**文章编号:**2095-5642(2014)10-0079-04

## 1 引言

进入新世纪以来,中国各地区的旅游业有了迅猛的发展,其相应的旅游翻译研究也随之增长。地处广西东部的贺州市是一座正在崛起的新兴旅游城市,有着“粤港澳后花园”的美誉。其著名景点有诸如黄姚古镇、姑婆山、大桂山国家森林公园、玉石林、十八水、紫云洞景区等。鉴于此,以贺州市为例的地方性旅游景区翻译语料库的创建与应用在提升其对外形象和影响力方面可以发挥其应有的作用,为贺州族群文化遗产的保护和对外宣传做出贡献,为岭南的人文风情研究添砖加瓦。

## 2 语料库的设计思路

广义地说,思路就是人们思考某一问题时思维活动进展的线路或轨迹。具体到语料库的创建,也就是创建者为了达到其预定目标而遵循的思维活动的线路。它影响到语料库所设计的研究目的或者相关教学作用是否能有效实现。就本旅游景区翻译语料库而言,其设计目的首先是翻译教学用途,也即本语料库将为地方性本科院校的旅游管理专业的双语教学或旅游英语专业的翻译教学提供具有服务于地方特色的第一手教学材料,并为设计多媒体辅助教

学提供真实和多样化的资源;其次是研究用途,包括对旅游景区源语言特点的研究和旅游景区翻译目的语特点的研究;再次是实用性,即有实际使用价值的,能够为在地方从事涉外旅游的英汉双语工作者提供便捷的旅游景区翻译参考。

标准的本质是统一,在设计思路上严格把守其标准,为语料库创建奠定坚实的基础与框架。<sup>[1]</sup>为了实现以上三个方面的总体目标,本语料库在设计时必须严肃谨慎,确保语料库的检索的结果反映所想要考察内容的程度和结果的一致性、稳定性及可靠性。

## 3 创建语料库的具体步骤

语料库具体创建的考虑因素包括语料的收集、语料文本的数字化、语料的标注、多语种语料的对齐、语料库的调试等实际操作步骤。

### 3.1 语料的收集

语料库根据收录语料的内容和属性的不同,通常可以分为异质型语料库(heterogeneous corpus)、同质型语料库(homogeneous corpus)、系统型语料库(systematic corpus)和专用型语料库(specialized corpus)。<sup>[2]</sup>由此可见,本语料库属于专用、同质型语料

\* 收稿日期:2014-04-10

基金项目:2012年度广西高等学校立项科研项目(201204LX455)和2011年贺州学院校级科研项目(2011SKZC05)的阶段性成果

作者简介:谭兴(1982—)男,湖北利川人,讲师,硕士,研究方向:语料库翻译学;

石婕妤(1978—)女,广西岑溪人,副教授,硕士,研究方向:二语习得、翻译学。

库,只收集与旅游景区介绍相关的中英文原生性文本(spontaneously sourced texts),<sup>[3]</sup>这就局限了该语料库的文本收集不能像异质语料库那样采取随遇性的原则,而只能是采用目的性取样标准。为了确保采集到的文本在内容上与语料库创建目的相关,在源语言和译入语都符合标准,本语料库建设中除了会实行研究小组多个成员把关制之外,还会邀请多个旅游管理领域的专业人士来审核文本的适合性和代表性。同时,对于某些介绍比较多的旅游景点文本材料,在决定采纳这些材料入库时亦会附加以三角验证法(triangulation),多渠道、多方式地反复核对入库文本。<sup>[4]</sup>在吸取前人宝贵经验和自发创新的基础上,这些举措将会确保该旅游景区翻译语料库的原始语料文本具备较强针对性和较好的准确性。

### 3.2 语料文本的数字化

一般而言,创建语料库所需要的原始材料绝大部分取自于网络在线的电子文本,以纸质形式出现的各类附带有语言文字的印刷品或者用各种相机拍摄下来的以图片格式存在的有关旅游景区的源语言介绍和对应的目的语翻译。总之,所有的这些原始材料都需转化为txt格式的文档备用。

具体到本语料库的语料主要来源,对于已经出版的关于贺州本地区的中英双语版的旅游景区方面的书籍,则采用较为先进的扫描仪以及配套的文字识别软件(如汉王文本王文豪7600系列),即从扫描仪录入的书本页面中把图像文字变为编码文字;以拍摄相片形式收集的中英双语旅游资料可以使用扫描仪生成高清图片,随后制作成PDF文档,接下来借助文本转换器,将图片上的文字装换成文本。但如果错误率偏高的话则需要手工矫正;还有一些以网页形式出现的材料,我们在浏览网页的时候或在电脑屏幕的某些区域有一些文字材料需要摘取下来,但可能某些网页设限不允许复制,就可以用文字抓取软件把他们抓取下来,常见的譬如Aqua Desktop;以音频格式或者视频格式出现的贺州双语宣传材料,则直接采用键盘手工录入,最终保存后缀名为txt的文档。

### 3.3 语料的对齐

要想使用Paraconc或者Anticonc等软件对语料进行检索和分析还必须进行双语或多语语料的对齐处理(alignment)。语料对齐是指将源语语料和译语语料分别保存在不同文本,并使两个文本中的语料

按段落与段落或句子与句子的关系进行对齐。<sup>[5]</sup>其中,语料的句子与句子自动对齐是目前平行语料库创建方面研究的重点。由于印欧语系语言之间的相似性,欧美地区已有比较成熟的软件如Multiconcord能对其进行精确度很高的自动句对齐处理。但由于英汉语之间的标点、句式和语篇差异明显,尤其是旅游文本的翻译很多时候并非完全遵照句子对应句子的原则,省译、扩译或摘译情况比较多见。因此,目前来说依靠软件自动句对齐处理准确度很低,这一工作仍需人工完成。人工进行语料对齐处理的时候,为了省时省力,可以借助Word和Excel等办公软件的一些特定功能进行批量处理。

### 3.4 语料的标注

语料标注影响到语料库建成后的检索应用范围,也是实现原始语料机读化的重要环节,其主要包括语料标注的原则、模式以及类型。语料标注的原则概括起来只有一点,即最大可能地方便标注者和使用者。<sup>[6]</sup>语料标注的模式种类繁多,本语料库采用更具形式化机读语篇信息编码能力的TEI标注模式(Text Encoding Initiative)。

语料标注的类型主要囊括语料背景信息、词性、词形、语义、句法分析和语篇结构等。其实,语料中的每个txt文件名也部分地反映了语料的背景信息,例如hy\_eng就表示该文档是关于黄姚古镇景区的英文翻译。诚然,文件名只能反映极少量的语篇背景信息。完整的语篇背景还需包括标题、作者(如年龄、性别、国籍等)、语篇的完成时间、口语还是书面语、内容类别(如科技、宗教、历史等)、语料来源(如出版物、标识语等)等各种信息。本语料库对这些信息的标注采取上文所述的TEI标注模式。词性标注的目的是为了明确地标记出语料中各个单词的词性,它是整个语料标注过程的最基础阶段,标注的结果为语料库具体应用中的句法分析和语义分析奠定基础。本语料库的英文语料采用德国斯图加特大学计算语言学研究所Helmut Schmidt设计的全免费工具Tree Tagger,可以在互联网自行下载;中文语料则采取中科院中文分词程序ICTCLAS进行词法切分和词性赋码。词形、语义、句法分析和语篇结构在本语料库的语料标注暂略,今后如有需要再进行补充和添加。

### 4 研究应用的前景

目前,本旅游语料库已经基本建成,以后将逐渐

由主要收集贺州地区的语料扩展为收集整个广西的中英文旅游语料,并考虑将语料库上传到互联网或制成可供检索的光盘数据文件。同时不断完善现有的标注体系,根据新的研究和教学需求,进行新的词形、语义、句法和语篇的标注,以最大程度地发挥本语料库的应用潜能。

#### 4.1 旅游景区翻译语料库在外语教学和外语学习中的应用

从20世纪80年代开始,随着语料库的建设及相关学科研究的发展,希望借助语料库检索结果所提供的频率信息,更好地描写其词汇特点,并衍生出相应的教学法,也即所谓的“词汇大纲”(Lexical Syllabus)、“词汇中心教学法”(Lexical Approach)以及与前二者紧密相关的“数据驱动学习”(Data-Driven Learning, DDL)<sup>[7]</sup>。

数据驱动学习的主要理念符合“以学生为中心和以教师为主导”的思想,即课堂活动或者练习都以学生自主学习为主。教师将设计好的教学材料交给学生,让其自己观察,或假设或推断,然后分小组进行讨论,最后归纳总结。这与交际语言教学中的发现问题、解决问题、互动讨论的做法相吻合。尤其在贺州学院旅游管理的双语教学或者旅游英语专业的旅游翻译教学环节中可以借助该语料库完成数据驱动学习中最典型的“单词项、多语境”练习。

因为语料库对于真实语境、短语和语体差异、量化分析的关注,以及搭配、类联接、语义倾向、语义韵等不同的分析思路都可以被运用到外语教学或者翻译教学中,例如词汇讲解、翻译训练、写作讲评、课后练习和试题的编制等。鉴于此,根据语料特点开发出一系列可供教师和学生课堂内外随时可以使用的教学材料,为开展旅游翻译数据驱动教学提供最为直观和真实的语料源泉。另外,可以建立一个该语料库的类比料库——学习者旅游翻译语料库,语料来源于大学生旅游翻译作业。把它与本语料库的两个子语料库,即相对独立的旅游景区汉语语料库和英语语料库作对比分析,找出学生常犯的错误,分析和探寻产生这些错误的原因,进而开展教学反思,最终促进教学。

当然,词汇大纲、词汇中心教学法以及数据驱动学习作为独立的语言教学形式难以与其他主流教学法相提并论。不过,词汇中心教学法在词汇教学、大纲词汇的选择和分级方面还是有其优势的。另外,

在听、说、读、写、译各项技能的教学中,也可适时地结合语料库方法辅助教学。

#### 4.2 旅游景区翻译语料库与研究应用

语料库翻译学在汲取普通语言学、传统译学以及其他相关学科营养的同时,不断充实和探寻自己新的研究内容和研究方法。该领域不仅颠覆了传统译学的研究手段,而且也深化和拓展了翻译学研究的内涵和外延。由此可见,本旅游景区翻译语料库有着较好的研究应用前景。

旅游翻译语言特征研究。比如,可以从翻译的共性入手,以词汇、句法和文体等为视角探讨旅游翻译是否具有显化、隐化、简化和范化等。除此之外,也可以从具体语言入手,探究翻译语言词汇特征、句法特征和搭配特征等。其中,词汇层面包括翻译语言词汇应用的总体特征和具体词汇或短语结构的应用特征,词汇应用的总体特征可以通过形符比、词汇密度和高频词的相关数据来分析;句法层面,从句法总特征和典型语句结构的应用特征两个层面来理解;翻译语言的搭配研究是一个相对比较复杂的工作,首先需要确定节点词(node)和搭配词(collocates),其次选择跨距(span)的长度,接着就是提取搭配词的和统计搭配显著性等。当然,上述这些研究前景都需在进一步提升本语料库研制程度之后进行。

译者风格研究。旅游景区的译者风格研究包括两个对象,分别是笔译和口译人员的翻译风格。但由于条件的限制,目前口译人员在景区的翻译材料收集还不充分。但是,旅游景区口译人员的风格研究是本语料库译学今后拓展方向之一。通常来说,基于语料库的译者风格研究主要从翻译语言特征分析和翻译策略与方法运用的研究等两个方面入手。关于前者翻译语言特征分析前文已有论述,不再多言。而后者涉及不同译者在翻译某个特定语言单位时所采用的翻译方法与策略的比较。首先,可以利用相关语料库软件(如Wordsmith 5.0)检索和统计不同翻译文本的类符或形符比、高频词、具体词汇、平均句长或者语句结构或者标点符号的使用频率,并用原创文本的相关数据为参照来进行比较;其次,依据语言学、文学和翻译学理论,详细分析和阐释译者风格形成的原因。<sup>[8]</sup>

旅游翻译规范研究。翻译规范是个抽象的社会科学概念,体现在具体社会行为当中,代表的是翻译

的社会性,即译者与客观世界和社会世界的相互关系。然而,译者本人在具体翻译实践时还要面对自己的主观世界,如何将这种主客观世界恰当地融合在一起,需要译者发挥自己的主体性和创造性。<sup>[9]</sup>

目前,基于语料库的翻译规范研究步骤如下:首先分别从词汇、句法和语篇等方面对翻译文本的具体特征进行描写和分析;其次,利用语料库所提供的关于翻译文本的出版时间和文本、译者性别及其他相关语料背景信息,依据相关翻译学理论归纳翻译规范,或对源于翻译规范方面的假设进行验证或修正,并从社会文化的角度对翻译规范的成因进行分析。

地方性译学语料库建设研究。以贺州市为例的地方性旅游景区翻译语料的创建与应用可以说是国内东南地区的旅游翻译研究开辟了一条新路径,对于系统研究区域性旅游翻译具有重要的意义。同时,还可以把该研究的成功经验推广至与其类似或相近的翻译研究,如地方性商务翻译语料库的研制与应用、地方性外宣翻译语料库的研制与应用、广西饮食文化翻译语料库的研制与应用和广西少数民族歌曲翻译语料库的研制与应用等等,不一而足。

## 5 结语

在借鉴前人自建小型翻译语料库的宝贵经验基础之上,无论是在设计思路,还是在语料的收集、语料的数字化、语料的对齐和语料的标注等每一环节都细心慎重。但由于财力和物力的局限,本旅游景区语料库还存在诸多不足,尤其是语料的标注方面

有待进一步完善和提高。另外,语料库只是一个工具,为了人尽其才和物尽其用,研制者需积累语言学、外语教学和翻译等多方面知识,树立实证思想,对英汉词性、词形、语义、句法、语篇对比、翻译共性等如有不解之处可经常求证语料库,以促进教、学、研的持续发展。

## 参考文献:

- [1] 李德超,王克非. 新型双语旅游语料库的研制和应用[J]. 现代外语, 2012, (2): 47.
- [2] 语料库的分类[EB/OL]. (2013-09-022) [2014-03-05]. <http://www.yygrammar.com/Article/201111/2700.html>.
- [3] Mason I. Translator behavior and language usage: Some constraints on contrastive studies [J]. *Herrvees*, 2001, (6): 66.
- [4] Li Defeng. Trustworthiness of think-aloud protocols in the study of translation process [J]. *International Journal of Applied Linguistics*, 2004, (3): 301.
- [5] 谢家成. 小型英汉平行语料库的建立与运用[J]. 解放军外国语学院学报, 2004, (3): 46.
- [6] 崔刚,盛永梅. 语料库中语料的标注[J]. 清华大学学报(哲学社会科学版), 2000, (1): 90.
- [7] 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, (7): 103.
- [8] 胡开宝. 语料库翻译学概论[M]. 上海:上海交通大学出版社, 2011: 109.

## Creation and Application of Regional Scenic Spot Translation Corpus: A Case Study in Hezhou

TAN Xing, SHI Jieyu

(School of Foreign Languages, Hezhou University, Hezhou Guangxi 542800, China)

**Abstract:** By conducting case study of Hezhou, a detailed discussion about the design idea and creation process of the regional scenic spot translation corpus is provided. The creation process includes the language data collection, digitalization, alignment and annotation. And then the application prospect of this corpus on teaching and analysis of tourism translation is explored.

**Key words:** regional; scenic spot; translation; corpus; application prospect

(责任编辑:张楠 责任校对:石木)