

广西旅游景区翻译语料库的创建及运用

覃始龙¹, 林媛媛¹, 刘卫东²

(1. 广西国际商务职业技术学院, 广西 南宁 530007; 2. 梧州学院, 广西 梧州 534002)

摘要:该文首先回顾了广西旅游景区翻译语料库建立的背景, 继而从语料的采集、语料文本的数字化、语料的对齐、语料的标注四个方面阐述了语料库创建的具体流程。最后从语料库本身的功能出发, 介绍了广西旅游景区翻译语料库在两个方面的应用空间, 一是在教育教学方面的运用; 二是在旅游翻译及研究方面的运用。

关键词: 广西; 语料库; 教育; 翻译

中图分类号: H315 **文献标识码:** A **文章编号:** 1009-5039(2015)16-0213-02

On Establishment and Application of the Translation Corpus in Guangxi's Tourist Attraction

QIN Shi-long¹, LIN Yuan-yuan¹, LIU Wei-dong²

(1. Guangxi International Business Vocational College, Nanning 530007, China; 2. Wuzhou University, Wuzhou 534002, China)

Abstract: The thesis first will review the background of the establishment of translation corpus in Guangxi's tourist attractions. Then the detailed process of establishing corpus will be explained from four aspects, the collection of corpus, the digitization of text corpus, the alignment of corpus and the annotation of corpus. Finally, starting from the corpus function itself, the range of application of translation corpus in Guangxi's tourist attraction will be introduced from the following two aspects, that is, the application in education and teaching and application in tourist translation and research.

Key words: Guangxi; corpus; education; translation

广西地区的旅游资源(如少数民族风情、历史名人博物馆、革命纪念馆等)极具地方特色和文化多样性的特点。这些特点一方面凸显了广西旅游资源的独特性, 另一方面也加大了旅游资源翻译的难度, 如翻译过于脱离文化背景, 无法准确传递旅游资源文化意蕴; 翻译过于支离破碎, 造成国外旅客难以理解^[1]。现有的大型通用型语料库无法准确地体现广西旅游的文本特征, 此种情形下, 建立整合大量实际应用的汉英翻译语料就显得尤为迫切。实质而言, 语料库创建的本意是满足创建者实际需求及实现创建者预期目标。以广西旅游景区翻译语料库为例, 其设计目标包含三个方面。一是用于翻译教学, 即将语料库作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学习教材, 创建地方特色教育。二是实现研究价值, 包含对旅游景点翻译目的语特点和旅游用语特点等方面的研究。三是保证实用性, 即满足实际应用需求, 能够为部分旅游景区管理委员会和涉外旅游工作者提供参考价值。

1 广西旅游翻译语料库的创建

完成语料库的创建, 需要语料的采集、语料文本的数字化、语料的标注、多语种语料的对齐、语料库的调试等一系列操作。

1.1 语料的采集

按收录语料内容和属性的不同, 可将语料库大致分为同质型语料库(homogeneous corpus)、异质型语料库(heterogeneous corpus)、专用型语料库(specialized corpus)和系统型语料库(systematic corpus)四种。因此, 该语料库为同质、专用型语料库, 仅采集涉及旅游景区或景点介绍的中英文原生性文本(spontane-

ously sourced texts), 进而使该语料库的文本采集方式与异质语料库不同, 仅可采取目的性取样原则, 限制了该语料库的全面性。为实现所有采集文本与语料库设计目标相符, 在译入语和源语言方面满足要求, 该语料库在创建过程中除采用研究小组多成员把关措施之外, 还另外邀请多名资深专业人士进行文本的审核、确定, 确保采集文本的适合性。此外, 在部分介绍较多的旅游景点文本材料选择方面, 该语料库依据三角验证法(triangulation), 详细研究入库文本材料的准确性和可行性。结合前人宝贵经验, 并积极创新采集方式, 将使该语料库拥有更高的使用价值和研究价值, 更好地实现其设计目标。

1.2 语料文本的数字化

通常情况下, 语料库创建所需的源材料大多是网络上的电子文本、储存大量与旅游景区或景点介绍相关的翻译用语等信息的图片和源语言介绍及其他文字叙述的纸质印刷品等, 这些材料最终都需要以txt格式储存在语料库中, 此时就需要语料文本的数字化处理技术了。

根据该语料库的文本来源的不同, 可将语料文本数字化处理技术大致分为扫描识别、文本转化、手工录入等几大类型^[2]。如对于以图片形式储存的语料文本, 可借助扫描仪制成高清图片, 并将其转化成PDF文件, 再利用文本转换器工具, 将图片上的文字转换成文本进行采集, 必要时可对模糊不清的图片进行修复; 对于已经出版的涉及旅游景区或景点介绍的翻译用语的书籍资料, 可通过扫描识别等先进技术将其录入语料库。对于具有使用价值的音频或视频文件, 可采用人工听取并录入的方式将其采集至语料库。对于网页上出现的具有参考价值的双

收稿日期: 2015-06-07 修回日期: 2015-06-20

基金项目: 本文系 2015 年广西高校科学技术研究项目“基于文化适应理论广西景点英语翻译与民族文化对接研究(项目编号: KY2015YB438)”的系列研究成果之一

作者简介: 覃始龙(1975—), 男(壮族), 广西河池人, 广西国际商务职业技术学院讲师, 硕士, 研究方向为专门用途英语; 林媛媛(1984—), 女, 广西横县人, 广西国际商务职业技术学院讲师, 硕士, 研究方向为英语教育; 刘卫东(1967—), 男, 广东郁南人, 广西梧州学院教授, 博士, 研究方向为英美文学翻译。

本栏目责任编辑: 谢媛媛

语言学研究 213

语翻译文本,若网页禁止复制,则可利用一些文字抓取软件将其抓取并转化为文本格式,再储存在语料库中。

1.3 语料的对齐

若将语料库实际应用,则应当使语料库具备信息检索和分析功能,这就需要采用ParaConc或Antconc等软件来完成,但前提是必须要对语料库的语料文本进行双语或多语的对齐处理。将源语语料与译语语料区分并单独储存,并使两种语料按照段落或单句格式相互对应的处理方法就是语料对齐,例如文言文的翻译以及部分翻译软件的工作原理等。目前,语料文本中句式对齐是创建语料库的难点。因为欧美语系具有相似性,西方已出现较为先进的应用工具,能够对语料文本进行高精度的句式对齐处理,譬如Multiconcord。但由于英语和汉语差异较大,在标点、句式等方面存在诸多不同,特别是在旅游景区或景点的翻译方面,翻译并未采取句式相对标准,扩译、省译或摘译等现象时有发生。所以,利用软件进行语料文本自动对齐处理并不能满足实际需求,语料对齐还需人工处理,必要时可利用一些office办公软件的特定功能减轻工作量。

1.4 语料的标注

作为原始语料机读化的重要环节,语料的标注对建成后的语料库的检索应用功能影响较大。其包括语料标注的理念、模式及类型。尽可能的方便标注者和使用者,是语料标注一贯的理念。语料标注的模式多种多样,该语料库应用的是TEI标注模式(Text Encoding Initiative)。

语料标注的类型包括语料文本信息、词性、语义、文本结构和句法分析等。其中,文件名即能反映出语料文本的部分信息,如bh_eng表示是与北海景区英文翻译有关的文件。但仅凭文件名提供的信息不足以满足实际需求。完整的文本信息还应包括标题、作者信息(如年龄、性别、国籍等)、语料来源(如出版物、标识语等)、内容格式和类型等方面。该语料库的语料标注应用TEI标注模式,注重语篇信息编码能力。本文仅以词性标注为例,今后如有其他方面(如语义、文本结构等)需求再另行补充。词性标注是将词的类别标注出来,如名词、动词等,其目的是令读者明晰单词的词性类别。作为语料标注最基础的词性标注,对于语料库中文本的句法分析和文本分析方面具有重要价值。该语料库的中文语料采用中科院中文分词程序ICTCLAS对文本词性进行标注;英文语料则应用德国斯图加特大学计算语言学研究所Helmut Schmidt开发的Tree Tagger工具进行词性标注。

2 广西旅游景区翻译语料库的运用

2.1 旅游景区翻译语料库在外语教学和外语学习中的应用

自上世纪八十年代初,一些教学法在语料库建设和相关学术研究发展潮流中应声而出,与凭借频率信息进行的词汇描述相辅相成,如我们常见的“词汇大纲”(Lexical Syllabus)、“词汇中心教学法”(Lexical Approach)和与上述两者存在紧密关联的“数据驱动学习”(Data-Driven Learning DDL)。

“数据驱动学习”以提高学生自主学习能力作为原则,与“以学生为中心”的理念遥相呼应。教师作为主导人员,将预先准备的教材发放到学生手中,并积极引导学生自主观察学习,培养学生养成假设、推断、讨论和总结的一系列自主学习习惯。这符合实际语言教学中的教学理念,与发现问题、解决问题的教学方针相得益彰。特别是对于旅游管理专业的学生,在双语教学或旅游英语翻译教学等方面均可利用该语料库,实现数据驱动学习中最经典的“单词项、多语境”学习的目标。

由于语料库对语境、短语和语体差异保持着格外的关注,并对语料文本的搭配、语义倾向、类联接等具有特殊的分析思路均能应用在英语教学或翻译教学方面,如词汇分析、翻译点评、写作指导和习题的制作等^[9]。因此,可依据语料特性编写一套甚至多套的学习教材,为教师和学生提供知识获取渠道,使数据驱动教学模式更加有利于在实际旅游翻译教学中的应用。此外,还可建立学习者语料库,语料文本来源于学生翻译

及写作作业,将其与该语料库中英语和汉语部分语料库进行对比,既有助于避免学生抄袭现象发生,又有利于发现学生存在的经常性错误,并对其进行分析和探索解决方案,从而提升教学效率。

然而,上述三种教学方法并不能单独应用于教学实际中。但词汇中心教学法在词汇教学、大纲词汇选取等方面具有显著优势。此外,在全面性教学中,也可适当选用语料库方法作为辅助用以提高教学质量。

2.2 旅游景区翻译语料库与研究应用

翻译语料库除了借鉴其他学科优势外,如传统翻译学、普通语言学等,还积极创新语料库采集方式。翻译语料库既弥补了传统翻译学的缺点,又将翻译学的优势放大,并拓宽了翻译学的研究领域,提升了翻译学的内涵。因此,旅游景区翻译语料库对于其他类型语料库创建以及其他学科具有重要的借鉴意义、研究价值和良好的应用前景。

旅游翻译语言特征方面的应用。如根据翻译的特点,从词汇、句式等角度出发详细讨论旅游翻译的省译、扩译和择译问题。同时,还可依据语言类型,探讨译语的句式特征、词汇特征以及搭配特点等。在句式特征研究中,可按句式总结构和典型句法特征不同进行研究;在词汇研究方面,可采用对比法,充分利用词汇密度和词汇频率等数据进行研究;对译语的搭配特点研究而言,第一步是选取节点词(node)和搭配词(collocater),第二步是选定跨距(span)的长度,第三步则是按照词的搭配特点进行数据采集和记录。上述应用实现前提,是进一步提升该语料库的研制程度。

译者风格方面的研究应用。对于旅游翻译而言,译者风格研究包含两个方面,一是对笔译的风格研究,二是对口译的风格研究。因受部分因素影响,旅游翻译对口译人员的信息采集力度尚未满足要求,但毋庸置疑的是,旅游翻译口译人员的信息采集将是该语料库日后研究重点之一。一般而言,对译者的风格研究从两个角度出发,即翻译原则与方式运用的研究和翻译语言的特点研究。后者在前文已经叙述。前者则是对不同译者在翻译同一对象时的原则与方式比较,该研究可分为两部分:利用语料库工具,从语料文本的词汇出现频率、句式结构或标点使用次数方面进行分析和记录,同时选用原文件作为参照进行对比分析;利用翻译学、文学等学科知识,针对译者风格的特点和形成进行详细探究和说明。

旅游翻译规范应用。翻译规范是指在翻译过程中需遵守的翻译标准,从深层次而言,翻译规范反映了译者与客观世界和物质世界的关联。但实际上,在翻译过程中,译者必不可少的要面对自己的主观世界,若译者无法做到妥善处理客观世界与主观世界占自身的比重,那么翻译作品质量就不能得到保证。对于旅游翻译语料库的翻译规范研究而言,其流程大致分为:从文本词汇、句式、语法等方面入手对翻译文本进行探讨研究;充分利用各种文本信息,如翻译时间、译者文化程度和年龄等,运用相关理论进行翻译规范总结。

3 结束语

基于语料库的翻译和语言实践日益受到了国内外研究者的重视,而完成高质量的语料库创建则是翻译和语言实践的前提条件。本文简要分析了语料库创建的流程,仍有部分问题有待完善,如双语文本句级对齐的完善、语义标准技术革新等。但在可预计的情形下,随着语言研究及旅游业的不断进步,语料库必将发挥更大的旅游辅助优势,而其在语言学领域也将开辟新的研究天地。

参考文献:

- [1] 覃剑. 广西文化旅游思考[J]. 法制与经济(下旬), 2011(8): 77-78.
- [2] 梁晓鹏, 康宁. 旅游翻译文本研究的语料库途径[J]. 青岛科技大学学报: 社会科学版, 2014(12): 12-14.
- [3] 冯占峰, 路仙伟. 翻译语料库下的翻译教学研究[J]. 大家, 2011(5): 102-104.