

# 海丝本土双语旅游语料库建设构想<sup>①</sup>

——以泉州为例

黄娟娟

(华侨大学外国语学院, 福建 泉州 362000)

[摘要]国内外专门旅游语料库的研制已经取得一定的成果,但基于种种原因,这些语料库到目前为止并未对外开放。要提高地方旅游翻译质量,推动福建地方文化发展与建设,培养地方旅游翻译人才,推动地方旅游翻译相关教学与研究,建设海丝本土旅游语料库是一种必然。文章以泉州为例,在探讨泉州本土双语旅游语料库建设意义的基础上介绍该语料库的建设步骤、语料库研制过程中遇到的问题,并展望该语料库的发展前景。

[关键词]旅游语料库; 双语; 本土; 泉州

[中图分类号]H0 [文献标识码]A [文章编号]1672-4720(2016)04-0249-05

## 1 引言

20世纪60年代,美国布朗大学建立的BROWN英语语料库标志着计算机语料库的诞生。而世界上第一个翻译语料库(TEC)则于1995年由英国曼彻斯特大学建立。

目前翻译研究中常用的三种语料库,即译文语料库、类比语料库及对应语料库多以文学文本或百科文本为主(王克非2012),结合地域特点而专门研制的双语专门语料库较为少见。旅游语料库作为一种专门用途语料库更不多见。据已有的文献,到目前为止,国内外最知名的旅游语料库有四:一、芬兰萨翁林纳翻译研究学院所建的67万词英语旅游文本语料库;二、英国埃塞克斯大学研究生Carlota Alcantar 2007年所建的37,795词的英语旅游文本语料库;三、日本大学Kiyomi Chujo等所建单语语料库“京都旅游语料库”,收集京都旅游

介绍相关文本885篇;四、香港理工大学中文及双语学系所建的200万字新型双语平行旅游语料库(李德超2010)。CNKI数据显示,自香港理工大学李德超教授2010年2月在《现代外语》发表文章《新型双语旅游语料库的研制和应用》后,国内期刊陆续刊登了9篇关于地方旅游语料库建设或者旅游语料的某一方面,如旅游公示语语料库建设的文章。

综上所述,国内外专门旅游语料库的研制已经取得一定的成果,但这些文献中提及的语料库到目前为止并未对外开放,而且即使开放了,由于地域文化的巨大区别,可借鉴的内容亦有限。在“一路一带”战略下,打造海丝旅游枢纽区成为福建的工作重点之一,这一目标的实现,翻译的作用不可小觑。要提高地方旅游翻译质量,推动地方文化发展与建设,培养地方旅游人才,建设海丝地方旅游语料库是一种必要。从小型、单个的地方语料库建设入手,

最后完成一个大的综合海丝双语旅游语料库的建设是可行的。本文的论述以泉州本土双语旅游语料库的建设为例。

## 2 建设海丝本土双语旅游语料库的意义

作为古代“海上丝绸之路”的起点,“21世纪海上丝绸之路先行区”泉州文化积淀深厚,旅游资源丰富,自然、人文景观独特。建设本土双语旅游语料库意义重大。

### 2.1 提升城市形象,促进地方文化有效传播

泉州市“十二五”旅游产业发展规划确定“把旅游业建成泉州国民经济的战略性支柱产业”,旅游目的地目标定位为“海峡西岸文化旅游中心”、“全国著名的旅游目的地”和“东亚地区知名旅游目的地”。

泉州旅游业正在蓬勃发展,前景光明,泉州文化正在不断的走出去,这其中翻译这座桥梁不可或缺。普通游客了解泉州主要有几个途径:网络、城市文化、旅游宣传手册、城市人文、自然景点介绍、城市各类博物馆、展览馆介绍等的英文版本。笔者对这几个渠道的旅游资料略作梳理后发现其中普遍存在以下几类问题。1、低级拼写、标点、语法等错误:如泉州海外交通史博物馆门口的指示牌居然出现拼写错误“Quanzhou History Museum of Marine Transportation”,正确拼写为 transportation;“古伊斯兰墓盖石”下面译文为 The cover of sarcophagus,正确拼写应为 sarcophagus;2. 译名不统一:如南音,有的翻译作 Nanyin Music,有的则译为 Nanyin Opera;茶都、香都、瓷都的“都”有 center, metropolis, capital 等多种译法;3. 粗暴直译:在一段介绍传统习俗的文字中,文题“清明祭扫”被译为: Mourning Ancestors on the Fes-

tival of Qingming ( Clear Brightness), 更佳的译文为 Mourning Ancestor on the Festival of Qingming ( Tome - Sweeping Day); 4. 错译: 休憩亭 pavilion 或 gazebo 被译为 Rest Pavilion, 承天寺祖师塔 Tombs of Ancient Monks 被译为 The founder of Chengtian Temple's Pagoda; 5. 汉英差异引起的段落翻译中式化表达法。

作为对外宣传、输出泉州文化的窗口与手段,旅游语料翻译中的种种失误会使泉州的城市形象,文化素质大打折扣。双语旅游语料库为地方旅游语料的翻译提供参考范式,同时在术语库中载入地方文化中一些专有名词、常用术语的翻译,有助于译名的统一,避免混乱。

### 2.2 提供范式制定,提高旅游翻译人才素质

作为“21世纪海上丝绸之路先行区”,仅2014年,泉州举办、承办或参与的与海丝文化相关的大型活动就有12场。这其中涉及的翻译量巨大,但翻译的质量无法保证。笔者有幸参加了由泉州旅游局、泉州文化局、泉州宣传部等组织编辑的各类旅游文化宣传手册的翻译工作。在翻译过程中,笔者发现各类材料虽然有所不同,但是有相当大一部分内容重叠或者相似。如果运用语料库进行翻译,可以节省很多的人力、物力和财力。

除了前期大量的笔译工作外,各类活动的成功举行也离不开翻译志愿者。翻译志愿者除了现场的翻译工作外还肩负传播地方文化的责任。这些翻译志愿者主要来源于各高校英语专业或者外语优秀的大学生。笔者走访了一些学生志愿者,发现他们在活动前虽然有受过培训,但是培训的内容简单,时间极短,有学生反应他们在与外籍人士交流时根本没想到,或者不知道如何向对方介绍泉州的地方文

化。语料库的建设可以为翻译志愿者提供现成的文化素材,通过志愿者更好的传播地方文化。对于语料库对翻译志愿者的培养笔者将另外撰文讨论。

大型活动结束后必然有许多外宾逗留泉州参观游览,此时的翻译主要由导游担任。据泉州旅游发展服务中心提供的数据,全市目前只有89人持有英文导游证。且这些持证导游参加的口试只有中文材料,英文翻译必须由参考人员自行翻译,翻译质量以及评委评定标准不可考。笔者走访市区主要的旅游公司,如中通,发现这些公司并无专门的英文导游,也无持证导游挂靠,有外国团的时候他们一般联系会外语的教师或者其他人员临时带团。这样的导游团队翻译质量堪忧,更不用谈文化传播意识。旅游语料库可以为导游考试提供多样的考试材料、相对标准可靠的译文以及相应的参照标准,同时也为英文导游提供自我学习、自我提升的重要渠道和窗口。

### 2.3 推动 DDL 教学,促进产学研的有效结合

语料库的使用在改良翻译教学及翻译教材编写的同时还为翻译研究提供新的切入点。基于语料库的翻译教学形式以其“文本电子化、学习直观化、分析多元化、结论科学化、成果有机化”等特点和优势,促进学生发现式学习,推动数据驱动式学习(DDL)翻译教学模式,有效克服传统翻译教学中普遍存在的一些问题,如“以教师为中心,缺少师生互动;轻技能培养而重知识传授;翻译教学方法单一;翻译教材不统一,内容陈旧且编写混乱”(黎士旺 2007)。”

语料库的使用使得学生在踏入社会前就熟悉甚至能够熟练使用翻译软件,这样可以帮他们更快的融入社会。因为目前会使用翻译

软件已经成为许多翻译公司的招聘条件之一。而且语料库的建设工程浩大繁杂,教师通过发动学生搜集语料,展示语料的整理、入库等让学生在实践中学习,从而有效促进产学研结合。旅游语料库为旅游领域的研究提供更丰富翔实的例证,以定性定量的优势为旅游翻译领域的研究提供新的切入点,如旅游翻译语言特征、译者风格、规范研究等。学习者旅游翻译语料库还可应用于旅游翻译“翻译教学大纲和教学教材的设计、语料库驱动翻译学习和翻译测试等领域的研究”(胡开宝 2007)。”

## 3 泉州本土双语旅游语料库建设构想

笔者构建的泉州本土双语旅游语料库是一个开放的地方性汉英平行专门语料库,初步预设为海丝双语平行语料库的一个子库,本库在初期基本板块建设成功的基础上将不断收集添加新的语料。

### 3.1 语料搜集

语料库构建的基本步骤是语料搜集。泉州旅游语料来源为有汉英对译的网络资源、已出版书籍、政府各部门出版发行的旅游与文化相关书刊与宣传册、各大博物馆及展览馆介绍说明、景区景点介绍、各类旅游公示牌。语料搜集的过程中按照预设的语料库板块存入相应的文件夹,非遗展示、景区景点、民俗节庆、饮食文化、交通运输、娱乐休闲、购物、酒店、教育等,文件结尾出标明语料出处、作者、时间等信息,并细化到子类,如景区景点下设如下子类:宗教、闽南建筑、名人史迹、滨海、生态、乡村、博物馆。而在宗教这一子类下再分设若干子类。

### 3.2 语料整理与清洁

由于语料库的语料来源众多,所搜集的语料的格式多样,网络资源可用离线浏览下载软

件将文本下载到电脑后整理。DPF与CAJ格式文档需转换成纯文本,印刷文本需扫描后再用文字识别软件识别与转换。所有的文档处理后保存为纯文本文档。常用的微软WORD文档并非纯文档,一般多用UNICODE格式,但由于本语料库为双语平行语料库,中英混排,故使用UTF-8编码。

本语料库搜集的语料多为官方双语报告、宣传材料或正式出版的书刊,为避免因同一译者或同一出版社影响过大,使语料库失去参考意义(麻丽莉 2009),我们先对不同来源的语料进行抽样,以保持不同译者和出处的作品在语料库中的平衡。此外,如前文所述,研究者发现这些语料或多或少的存在这样或那样的翻译问题。因而我们组织翻译专家团队,对译文进行再核对校审并修改,严把语言质量与翻译质量。

这些经过审核修改的语料因未经过任何语料标注,仍为生文本。这样的文本存在各种不规范的符号、格式,如果不加以清理,会导致自动词性赋码时出错,因此我们需要对文本进行清洁。我们可以通过一款网上可以直接搜索下载的名为“软件处理器”进行文本清洁。该软件可针对多个文本进行批处理,并且可以通过设置的整理方案一次执行多个替换、转换功能(梁茂成 2010)。但必须指出的是,后续,在本语料库建成的基础上,我们会建设旅游翻译学习者语料库,这样的库对于学习者的文本们应该尽可能保持原貌,包括所有的格式与语用错误。

### 3.3 语料对齐与标注

语料对齐是语料库建构的核心步骤。本语料库对齐运用TRADOS2015版中的WinAlign程序进行文本对齐。首先将原文和

译文创建成相关的翻译记忆库项目。在SDL Trados 2015主页工具栏中,点击按钮“对齐已翻译文档”,进入Winalign界面,创建新项目,设置原语言为中文,目标语为英语,在中文栏和英文栏中分别添加原文及译文并点击按钮“Align File Name”,将两个文件之间相连。在该选项卡中,可以添加多组原文及译文,并将它们之间相互连接。添加文件完成后,把文件导出成翻译记忆库交互文件tmx后导入翻译记忆库中。接着选择需要匹配的文件对,WinAlign后自动匹配的句段以虚线连接。接下来是进行检查的工作,有错误的地方可人工修改。SDL Trados 2015是一款需购买付费的专业翻译软件,且经过Trados 2015对齐存储的文件只能在该软件窗口打开。目前可以获得的免费自动分句软件有许家金、贾云龙于2010年开发的英文自动分词工具BFSU Sentence Segmenter 1.0,Transmate免费语料对齐工具,以及雪人翻译软件提供的可免费使用的双语对齐的插件对齐。

语料库除了通过帮助译员重复利用以前翻译过的资源提高翻译速度并保持术语的一致性以减少翻译项目的花费,还服务于翻译学习、教学与研究。后者三者均离不开对语言的对比分析,因此我们需要记录文本的各种信息,包括语料的来源、文本的内部结构、语法信息等。这种信息记录被称为语料标注。语料标注是一个重要的检索依据,方便我们日后通过各种语料库分析软件提取信息,为我们分析和研究语言提供大量的实证数据。语料库的标注取决于语料库的研究目的,XML是一种通用的标注语言,结构灵活,注重对信息意义的描述。我们还可以通过梁茂成教授等开放的Lemmatizer软件进行词形还原,用德国斯图加

特大学 Helmut Schmidt 设计的全免费用具 Tree Tagger 进行词性赋码。

#### 4 本土双语旅游语料库研制的现状与前景展望

在泉州本土双语旅游语料库的研制过程中,我们遇到了一些困难。首先,为确保收入语料的效度,需要大量的校正工作。其次,句子对齐工作是语料库的核心,但由于汉英语言的差异,目前正在使用的软件只能实现局部自动对齐,我们仍然需要使用人工逐条逐句进行核对对齐,这样人工消耗巨大。再次,语料库需要不断更新资料,一些官方资料我们需要经过多种渠道和程序才可获得。最后,通过大型商业翻译软件如 Trados 建立的翻译数据库能有效节省翻译工作量,便于翻译项目管理,但它服务于翻译人员,对于翻译教学与研究帮助有限,且耗资巨大。上述种种均需要投入大量的人力、物力与财力。

正如胡开宝教授指出的“语料库的建设需要不同专业人才组成的团队、长期稳定的资金支持、不同地区或国家间的合作和明确的研究计划等”(2007),本土双语旅游语料库的建设最佳的方案就是由政府提供资金,指定部门,如旅游局牵头,依托协会,委托高校进行研制。旅游局负责各类相关双语文件文书的搜集,委托本地翻译协会组织专家团队,对译文进行审校,高校负责语料的最终处理。后期,语料库建设完成后,借助已有旅游语料库在翻译教学中的运用,可以建设旅游翻译学习语料库,可以协调福州、漳州等地,建立福建海丝双语旅游语料库,最后扩展到的海丝综合旅游语料库。

#### 5 结语

上述只是语料库研制的一部分,后续还需

不断扩充语料,完善标注体系,根据不同的目的进行新的语义及句法标注,最大限度发挥语料库的潜力。作为一种新兴技术,翻译语料库具有广阔的应用前景。以泉州本土双语旅游语料库为先,逐步发展起来的海丝旅游语料库的建设并应用于地方旅游翻译人才培养和旅游翻译研究既是顺应当前国内外翻译教学与研究的新趋势,也是对传统翻译研究和教学方法的有益补充,该语料库在服务译员,提高译员素质与翻译质量的同时推动海丝地方文化更有效的传播,具有不断改进发展的空间,对服务地方文化建设与培养地方旅游翻译人才有着特殊的优势。

注释:

- ① 本文为本人主持的“2015 年福建省中青年教师教育科研项目《旅游翻译学习者语料库与旅游翻译人才培养》项目批准号: JAS150094”的相关成果。

#### 参考文献

- [1] 胡开宝, 吴勇, 陶庆. 语料库与译学研究: 趋势与问题[J]. 外国语, 2007, (5): 64-69.
- [2] 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: 21-55.
- [3] 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, 2010, (1): 46-54.
- [4] 黎士旺. 语料库与翻译教学[J]. 中国科技翻译, 2007, (3): 46-49.
- [5] 麻丽莉, 王祥兵. 军事平行语料库的建立及其在军事翻译方面的应用[J]. 国防科技, 2009, 30(1): 38-41.
- [6] 王克非. 语料库翻译学探索[M]. 上海: 上海交通大学出版社, 2012: 209-210.

收稿时间: 2016-1-27

作者简介: 黄娟娟(1977-), 女, 福建泉州人, 硕士, 讲师, 研究方向: 翻译理论与实践