

世界遗产双语语料库的研制与应用

——以福建为例

余 军,王朝晖

(厦门理工学院 外语系 福建 厦门 361024)

摘 要:目前国内外尚无以大陆地区汉译英旅游文本为对象的旅游翻译语料库,国内对世界遗产旅游翻译的研究也寥寥无几。研制以国内世界遗产旅游翻译为研究对象的大型双语旅游语料库,有助于推动基于实证语料的旅游翻译研究的发展。以福建为例,从语料库的构成、语料来源、语料标注、语料检索等方面介绍世界遗产双语语料库的研制方法,并探讨其在旅游语言及旅游翻译的特征研究、世界遗产旅游翻译的质量评估、世界遗产旅游翻译模式、计算机辅助旅游翻译教学模式等方面的应用。

关键词:世界遗产双语语料库;研制;应用

中图分类号: H315.9

文献标志码: A

文章编号: 1001-7836(2012)11-0134-03

一、国内外基于语料库的翻译研究现状及相关建议

近 20 年来,基于语料库的翻译研究日趋增多,构建的语料库主要分为译文语料库、类比语料库和双语对应语料库三类^{[1]46}。上述三种语料库,多数以文学文本为主,或是以收纳百科文本为特色,而结合地域特点的或针对某一文类而专门研制的双语专门语料库则较为少见^{[1]46}。目前,国内外较具规模的双语专门语料库很少,就旅游文本而言,国外的旅游专门语料库主要有三个,分别为芬兰、英国和日本学者研制的,其共同特点是都属于容量小于 100 万词的小规模语料库。而国内大陆地区除中科院研制的旅游咨询口语对话语料库和旅馆预定口语对话语料库外,尚无专门研发的旅游语料库,香港地区则仅有香港理工大学中文及双语学系研制的新型双语旅游语料库,是香港以至全国第一个较大容量的英汉/汉英旅游语料库,该语料库由一个双语旅游翻译对应语料库与一个双语旅游翻译类比语料库组成,语料规模为各 100 万字/词^{[1]48}。前者目前收录香港地区的英译汉和汉译英的旅游翻译文本,后者主要收录以香港地区为主的两岸三地非翻译的中文和英文旅游原生性文本。

目前国内外尚无以大陆地区汉译英旅游文本为对象的旅游翻译语料库;而国内旅游翻译研究仍较多地局限于以

旅游翻译实践诠释国外某些翻译理论,对旅游翻译理论和实践有实际指导意义的成果为数较少。旅游翻译的研究应以文化交流和传播为目的,进行基于双语语料库的实证研究,李德超博士和王克非教授在这一方面已经为研究者提供了成功的范例,先后在《中国翻译》、《外语电化教学》等期刊上发表了两篇基于语料库的旅游翻译研究论文^[2-3],其语料库的研究方法和视角令人耳目一新。

尽管如此,香港双语旅游语料库目前规模还比较小(约 200 万字/词),翻译语料来源目前仅限于香港地区。旅游翻译研究要取得新的突破,有必要研制一个规模更大的双语旅游语料库。目前,国内旅游景点众多,而最能代表中华文化和自然风光的莫过于世界文化和自然遗产,因此,研制大型双语旅游语料库以推动旅游翻译研究,首选对象是世界遗产的旅游翻译。

目前国内对世界遗产旅游翻译的研究并不多见,仅有数篇论文和一部专著。其中,张美芳教授的《文化途径看澳门——浅谈澳门世界遗产景点译名》一文^[4],考察了澳门历史城区景点名称的翻译,并讨论了导致不同译名的可能成因;刘秀芝和李红霞合著的《北京世界文化遗产人文景观介绍翻译研究》一书^[5],是首部有关世界遗产旅游翻译的专著,利用大量实例说明如何具体处理景观名称及景观介绍翻译过程中所遇到的各种现象和问题。这些研究为世界遗产

收稿日期:2012-05-02

基金项目:福建省社会科学规划项目(2011B244)及福建省教育厅 A 类人文社科研究项目(JA11246S)的阶段性成果

作者简介:余军(1973-),男,江西靖安人,讲师,硕士,从事英汉语比较与翻译、语料库翻译学的研究;王朝晖(1971-),女,吉林公主岭人,副教授,博士,从事跨文化翻译、语料库翻译学的研究。

旅游翻译进行了有益的探索,但都限于某一地域,难以反映国内世界遗产旅游翻译的全貌。

笔者认为,收集覆盖国内 40 个世界遗产的旅游翻译文本及英语国家世界遗产的原生性英文文本,研制以国内世界遗产旅游翻译为研究对象的大型双语旅游语料库,以世界遗产旅游翻译为研究对象,以英语国家世界遗产原生性英文文本为“平行文本”,充分利用语料库存储数据大、计算机运行速度快、语料库工具能对大量数据进行分析的诸多优势,采用比较翻译的研究方法,从定性与定量的角度研究中国世界遗产旅游翻译,在翻译技巧运用、语言选择与文化内容处理等方面系统地在国内 40 个世界遗产的旅游翻译进行比较分析,全面、准确地描述国内各个世界遗产旅游翻译的长短优劣与异同得失之处,揭示世界遗产旅游翻译中的隐性规律,建立基于实证语料的世界遗产旅游翻译理论及应用体系,为世界遗产旅游翻译评价提供一个可检验的标准,为我国的遗产旅游翻译研究和教学实践提供参考,将极大地推动基于实证语料的旅游翻译研究的发展。目前,笔者正以福建土楼、武夷山为研究对象,进行世界遗产双语语料库的研制,并拟在此基础上,将该语料库扩容至包括国内 40 个世界遗产的大型双语语料库。下面简要介绍该语料库的研制方式及应用。

二、世界遗产语料库的研制与应用

(一) 世界遗产语料库的研制

目前笔者研制的世界遗产双语语料库仅包含福建土楼和武夷山两个世界遗产,库容在 100 万字/词左右,扩容至包括国内 40 个世界遗产后,规模应在 2 000 万字/词左右,而且将保持动态更新。

1. 建库类型。参照语料库,即英语国家(如美国、英国等)世界遗产原生性英文语料库,与双语对应语料文本进行“平行文本”比较,研制汉英对应语料库,即世界遗产汉英对应语料库。

2. 语料来源。全部语料限定在与世界遗产相关的范围,如景观介绍、景点酒店介绍、景点中英文招牌、民俗风情、历史文化、饮食文化、物产资源、人名地名等;语料来源主要包括正式出版物(图书、音像制品等)、非出版物(网页、各类旅游宣传材料,如酒店介绍、旅行社传单、景点有关部门免费发放的旅游小册子等)及景点标牌上的介绍文字和公示语等。

3. 语料库标注。对全部语料进行文本头及词性标注;抽取部分典型汉英对应语料进行错误标注。制定标注规范,包括错误标注分类、标注赋码、标注流程,并实施严格的标注质量监控。

4. 建库路线图。语料搜集(实地考察、图书资料购买、网络资源等)→语料加工(文字识别、校对)→建库(编制文本头各项信息填写要求;利用 WinAlign 制作双语对齐语料)→语料库标注(ICTCLAS2008 及 Claws 分词软件自动标注;设计错误标注赋码,人工标注)。

(二) 世界遗产语料库的检索

世界遗产语料库的参照语料库可使用 Wordsmith、

Antconc 及 PowerGrep 等软件检索,双语对应语料库具备两种形式,一种是单文本形式,文本中每行都包括原文及译文,原文及译文之间用 tab 符分隔。单文本形式的双语语料库可用 PowerGrep 及贾云龙的检索工具检索。检索示例如图 1 和图 2:

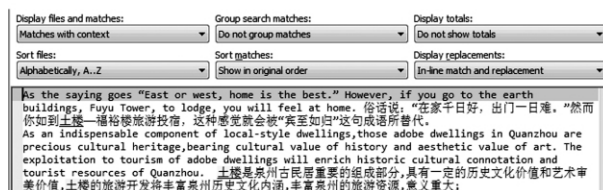


图 1 PowerGrep 检索“土楼”

No	Text	Line	File
1	The tasty dishes of Southeast Asian countries are fully available in the buffet supper in our cafe. 馨香可口的东南亚各国食品一切尽在武夷山悦华酒店咖啡厅自助晚餐, 会令您的胃口大开。	1	CE-0215.txt
2	The Wuyishan city has the glorious tea history. The tea industry also is the local prop industry, it has 1/5 local population steps in the tea industry. 武夷山市具有悠久的茶历史, 茶业也是当地的支柱产业, 有 1/5 的当地人口涉足茶业。	2	CE-0215.txt
3	Tourists climb to the top of Wuyi Mountain in Fujian, China. 在中国福建, 游客们正在向武夷山顶进发。	3	CE-0215.txt

图 2 贾云龙检索工具检索“武夷山”

另一种是双文本形式,即原文及译文各一个文本,两者的总行数一样,原文及译文在各自文本中的行数一致,从而形成平行对应关系,可使用著名的双语语料库检索软件 Paraconc 检索。双文本形式的双语对应语料库采用中科院的 ICTCLAS 2008 对汉语原文进行分词,并人工校对,分词后方可在 Paraconc 中检索。

除可使用上述软件对语料库进行本地检索之外,还可编制网页检索程序,对世界遗产语料库进行在线检索,检索程序可以实现多种检索方式,如关键词检索,正则表达式检索等。

(三) 世界遗产语料库的应用

目前国内有关世界遗产旅游翻译的研究很少,语料库技术在世界遗产旅游翻译研究中的应用则更是寥寥无几。研制大型世界遗产语料库,有助于进行如下几个方面的研究:

1. 旅游语言及旅游翻译特点的研究。在比较分析世界遗产汉英对应语料库及参照语料库的基础上,研究旅游翻译在词汇、修辞、语篇等语言层面的特点,以及旅游翻译与普通翻译的不同之处。

2. 世界遗产旅游翻译质量评估。通过研制世界遗产双语语料库,获取充分的世界遗产双语对应实证语料,利用检索软件的检索统计功能,以英语国家世界遗产原生性英文文本为“平行文本”,对国内现有的世界遗产的翻译质量进行评估,揭示国内世界遗产旅游翻译现状及目前存在的问题(如译名不统一的问题,世界自然遗产武夷山在国内就有“Wuyishan Mountains”、“Wuyi Mountain”、“Wuyi Mountains”、“Wuyishan Mountain”、“Wuyi Mount”、“Mount Wuyi”等多种译名,极其混乱),在比较分析大量真实语料的基础上构建世界遗产旅游翻译质量评价体系。

3. 构建基于语料库及计算机辅助翻译技术的世界遗产旅游翻译模式。对所研制的汉英对应语料库进行审校,一是

审核汉语原文及译文,对存在错误的译文进行修订,将经审核修订后正确无误的双语语料转换为计算机辅助翻译软件记忆库;二是从大量语料中提取出旅游术语,如景点名称、地名、人名等,并统一译名,建立计算机辅助翻译术语库。在国内世界遗产旅游翻译记忆库及术语库的基础上,利用计算机辅助翻译软件(如SDL Trados、Déjà vu x 和国产的雪人免费翻译软件),进行计算机辅助旅游翻译实践的探索,构建基于语料库及计算机辅助翻译技术的世界遗产旅游翻译模式。

4. 探讨语料驱动(corpus-driven)的计算机辅助旅游翻译教学模式。利用世界遗产双语语料库,建立语料驱动的计算机辅助旅游翻译教学模式,研究语料库与计算机辅助翻译技术在微观翻译教学环节的具体应用。有关语料库技术在翻译微观教学环节的应用,笔者曾在《中国翻译》上发表论文对此予以阐述^[6]。

5. 探讨翻译研究如何服务于中国文化产业发展和民族文化传播。在上述研究的基础上,探讨如何促进中国文化产业发展和华夏文化传播。例如,通过研制世界遗产双语语料库,为研究者和翻译工作者提供参考语料;构建计算机辅助旅游翻译术语库,推动译名统一。

三、结语

通过研制大型双语语料库的方式呈现中国世界遗产旅游翻译的真实全貌,以标注的方式对双语语料库进行深加工,采用比较翻译的研究方法,通过构建以功能对等^[7]、目的论^[8]、平行文本比较为基础的分析模型,研究世界遗产翻译,考察目前旅游翻译的现状,揭示旅游文本及旅游翻译的隐性

规律;在深入加工的语料的基础上,制作计算机辅助翻译记忆库及术语库,开展基于语料库及计算机辅助翻译技术的旅游翻译教学,促进旅游翻译教学与研究;通过大型双语语料库、翻译记忆库及术语库的研制与应用,改变旅游翻译现状。该大型世界遗产语料库的研制及推广应用,将有利于促进旅游文化产业发展和民族文化传播。

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 2010 (1): 46-54.
- [2]李德超,王克非. 平行文本比较模式与旅游文本的英译[J]. 中国翻译 2009 (4): 54-58.
- [3]李德超,王克非. 基于双语旅游语料库的 DDL 翻译教学[J]. 外语电化教学 2011 (1): 20-26.
- [4]张美芳. 文化途径看澳门——浅谈澳门世界遗产景点译名[J]. 上海翻译 2009 (2): 29-33.
- [5]刘秀芝,李红霞. 北京世界文化遗产人文景观介绍翻译研究[M]. 北京: 光明日报出版社 2008.
- [6]余军,王朝晖. 基于比较翻译教学法的教学型语料库构建与应用[J]. 中国翻译 2010 (5): 57-62.
- [7]Nida Eugene A. Taber C. R. The theory and Practice of Translation[M]. Leiden: E. J. Brill, 1969.
- [8]Nord Christiane. Translating as a Purposeful Activity: Functional Approaches Explained [M]. Shanghai: Shanghai Foreign Language Education Press 2001.

The Construction and Application of the Parallel Corpus of China's World Heritages

YU Jun, WANG Zhao-hui

(Faculty of International Languages, Xiamen University of Technology, Xiamen 361024, China)

Abstract: At present there is no Chinese-English parallel corpus of tourism texts of the Chinese mainland, and the study on tourism translation of China's world heritages is rare. The construction of a large parallel corpus of China's world heritages helps to promote corpus-based tourism translation studies in China. This thesis, taking world heritages in Fujian as an example, discusses the construction of the Parallel Corpus of China's World Heritages, including the corpus structure, text sources, corpus annotation and search. The corpus can be applied to fields such as the study of linguistic features of tourism texts, original and translated, the quality assessment of and the model for tourism translation of China's world heritages, and computer aided teaching modes of tourism translation.

Key words: Parallel Corpus of China's World Heritages; construction; application