

◎福建省社会科学规划项目（2011B244）成果

THE CONSTRUCTION AND APPLICATION OF
THE BILINGUAL FUJIAN WORLD HERITAGE CORPUS

福建世界遗产双语语料库 构建与应用

余 军 王朝晖 著



厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

THE CONSTRUCTION AND APPLICATION OF
THE BILINGUAL FUJIAN WORLD HERITAGE CORPUS

福建世界遗产双语语料库 构建与应用

余 军 王朝晖 著



厦门大学出版社 国家一级出版社
XIAMEN UNIVERSITY PRESS 全国百佳图书出版单位

图书在版编目(CIP)数据

福建世界遗产双语语料库构建与应用/余军,王朝晖著. —厦门:厦门大学出版社, 2018. 11

ISBN 978-7-5615-7184-2

I. ①福… II. ①余…②王… III. ①文化遗产—语料库—福建 IV. ①K295.7

中国版本图书馆 CIP 数据核字(2018)第 262823 号

出版人 郑文礼
责任编辑 王扬帆 高奕欢
封面设计 李嘉彬
技术编辑 朱 楷

出版发行 厦门大学出版社
社 址 厦门市软件园二期望海路 39 号
邮政编码 361008
总 编 办 0592-2182177 0592-2181406(传真)
营销中心 0592-2184458 0592-2181365
网 址 <http://www.xmupress.com>
邮 箱 xmup@xmupress.com
印 刷 虎彩印艺股份有限公司

开本 720 mm×1 000 mm 1/16
印张 14.5
插页 2
字数 261 千字
版次 2018 年 11 月第 1 版
印次 2018 年 11 月第 1 次印刷
定价 48.00 元

本书如有印装质量问题请直接寄承印厂调换



厦门大学出版社
微信二维码



厦门大学出版社
微博二维码

余军

厦门理工学院副教授，主要从事语料库翻译学研究。近年来，主持教育部规划基金项目 and 福建省社科项目各 1 项，主持市厅级项目 2 项；在《中国翻译》、《外语界》等核心期刊及其他学术期刊发表论文 10 篇，出版专著 1 部。2010 年获《中国翻译》创刊 30 周年翻译研究有奖征文“翻译研究优秀论文”二等奖。获福建省教学成果二等奖（排名第二）。

王朝晖

东北师范大学博士，厦门大学博士后。现为厦门理工学院副教授，研究方向为跨文化交际、语料库翻译学。近年来，主持教育部规划基金项目和福建省社科项目各 1 项，主持市厅级项目 3 项，在《美国研究》、《东北师大学报》（哲学社会科学版）、《中国翻译》等核心期刊及其他学术期刊发表论文 10 余篇，出版专著 1 部。2010 年获《中国翻译》创刊 30 周年翻译研究有奖征文“翻译研究优秀论文”二等奖。获福建省教学成果二等奖（排名第三）。

福建省社会科学规划项目（2011B244）成果
本书获厦门理工学院学术专著出版基金优先资助

前言

世界遗产作为文化传播的重要媒介，翻译在其中起了不可替代的作用。翻译促进了不同文化间的有效沟通，进而实现世界遗产文化符号的传播。世界遗产翻译的质量，在一定程度上影响了文化符号的传播进程，相关研究具有重要意义和价值。但目前世界遗产翻译的研究存在总量极少、视角有限、缺乏系统考察等问题，基于语料库的实证研究更是付之阙如。

有鉴于此，本书以翻译这一世界遗产文化传播的重要途径为视角，以福建世界遗产为对象，以语料库的实证语料为依据，探究世界遗产翻译存在的问题与对策。通过语料库的构建及应用，评估福建世界遗产翻译质量，探索 CCAT 平台下的世界遗产翻译模式，提高福建世界遗产翻译的质量，促进福建文化的传播和文化产业的发展。

本书将语料库与 CAT 两大翻译技术结合，将其应用于福建世界遗产翻译研究，提出了世界遗产翻译的“信、达、效”原则，阐述了福建世界遗产双语语料库的研制及其应用，并从福建申遗文本、景区文本、酒店介绍和导游词等角度评估福建世界遗产翻译质量，考察了福建世界遗产网站的本地化情况，分析了存在的问题并提出对策。这些探讨，对于世界遗产翻译质量的提高，具有一定的实用价值。

在本书成稿之际，我们要感谢所有关心、支持本书的机构和个人。本书获厦门理工学院学术专著出版基金优先资助，厦门理工学院外国语学院也为本书的出版提供了资助，在此一并致谢！感谢魏志成教授在学术上给予我们的帮助和激励。

本书部分章节由王朝晖撰写，约 5 万字，其他部分由余军撰写；全书由余军统稿。

我们希望本书能引起更多人对世界遗产翻译研究的关注。由于水平有限，不当之处在所难免，诚挚地希望广大读者批评指正。

余军 王朝晖

2018年5月于厦门

内容简介

本书以福建世界遗产为对象,以语料库的实证语料为依据,探究福建世界遗产翻译存在的问题与对策。在 CCAT 理论架构的基础上,提出切合世界遗产翻译的“信、达、效”原则及标准。从福建申遗文本、景区文本、酒店介绍和导游词等方面评估福建世界遗产翻译质量;考察了福建世界遗产网站的本地化情况;提出了基于 CCAT 的世界遗产翻译模式,即机器辅助人工翻译模式和人工辅助机器翻译模式,并分别以申遗文本和酒店介绍翻译为例阐明了两种模式的具体应用。

目录

第1章 绪论	1
1.1 引言	1
1.2 福建世界遗产翻译研究现状	1
1.3 语料库与世界遗产翻译研究	2
1.4 世界遗产翻译研究的 CCAT 视角	3
1.5 本书主要内容	4
第2章 福建世界遗产双语语料库的研制	5
2.1 引言	5
2.2 设计思路	5
2.3 语料来源及建库类型	6
2.3.1 语料来源	6
2.3.2 建库类型	6
2.4 建库工具	6
2.5 语料标注	7
2.6 术语库制作	8
2.7 记忆库制作及转换	15
2.7.1 记忆库制作	15
2.7.2 记忆库转换	16
2.8 小结	17
第3章 福建世界遗产双语语料库的检索	18
3.1 引言	18
3.2 双语对应语料库的检索	18
3.2.1 贾云龙工具	18

福建世界遗产
双语语料库构建与应用

3.2.2 网络检索平台.....	35
3.3 小结.....	36

第4章 福建世界遗产翻译质量评估.....37

4.1 引言.....	37
4.2 世界遗产翻译的“信、达、效”原则.....	37
4.3 基于CCAT的世界遗产翻译质量评估模式	41
4.4 评估个案I——申遗文本.....	41
4.4.1 语言错误.....	41
4.4.2 一致性.....	44
4.4.3 类比语料库比较.....	45
4.5 评估个案II——景区文本.....	46
4.5.1 武夷山景区文本的翻译质量评估.....	46
4.5.2 问题成因与对策.....	50
4.6 评估个案III——酒店介绍.....	51
4.6.1 缤客网酒店英文介绍特点——以鼓浪屿酒店为例.....	52
4.6.2 酒店介绍翻译质量评估.....	56
4.6.3 问题分析及对策.....	64
4.7 评估个案IV——导游词.....	65
4.7.1 导游词翻译研究概述.....	65
4.7.2 语料来源.....	66
4.7.3 基于双语语料库的土楼导游词译文质量考察.....	67
4.7.4 基于类比语料库的土楼导游词译文易读性分析.....	71
4.7.5 提高易读性的对策.....	73
4.8 小结.....	86

第5章 福建世界遗产网站本地化研究.....87

5.1 引言.....	87
5.2 网站本地化研究现状.....	88
5.3 福建世界遗产网站建设现状.....	89

5.4 福建世界遗产网站本地化质量评估——武夷山及永定土楼	89
5.4.1 网站入口	90
5.4.2 网站设计	93
5.4.3 网站内容	98
5.4.4 语言特点	102
5.4.5 文本翻译	105
5.5 福建世界遗产网站本地化对策	115
5.5.1 网站设计优化	115
5.5.2 翻译质量优化	115
5.5.3 关键词优化	117
5.5.4 众包翻译	121
5.5.5 Wiki 词条编辑及维护	122
5.6 小结	123

第 6 章 CCAT 平台下的世界遗产翻译模式 124

6.1 引言	124
6.2 CCAT 平台	124
6.3 世界遗产翻译模式	124
6.4 应用个案 I —— 申遗文本的机器辅助人工翻译	125
6.4.1 武夷山申遗文本翻译	125
6.4.2 土楼申遗文本翻译	129
6.5 应用个案 II —— 酒店介绍的人工辅助机器翻译	136
6.5.1 机器翻译在旅游翻译中的应用——以 TripAdvisor 为例	136
6.5.2 酒店介绍文本来源	139
6.5.3 人工辅助机器翻译的操作流程	139
6.6 小结	144

第 7 章 福建世界遗产双语语料库的译文修订 145

7.1 引言	145
--------	-----

7.2 申遗文本译文的精细修订样例.....	145
7.2.1 武夷山申遗文本.....	145
7.2.2 土楼申遗文本.....	151
7.3 入库译文的修订样例.....	155
7.3.1 武夷山申遗文本.....	155
7.3.2 土楼申遗文本.....	159
7.4 小结.....	164
参考文献.....	166
附录 1 福建世界遗产网站本地化长尾关键词样例.....	172
附录 2 福建世界遗产双语语料库之术语库样例.....	186
附录 3 福建世界遗产申遗文本句对样例.....	191

第1章 绪论

1.1 引言

福建文化源远流长,闽南文化、客家文化、朱子文化、海丝文化,各具特色,在海内外颇具影响。在文化“走出去”战略及“一带一路”背景下,福建文化占据天时地利,其对外传播,可谓充满机遇,但仍不乏挑战。

以世界遗产为例,申报者众,一旦入选世遗名录,不仅扬名国际,而且游客将纷至沓来^①,文化产业潜力无穷。竞争之激烈,自不难想象,但福建凭其深厚的文化底蕴和丰富的申遗经验,在拥有武夷山、土楼和泰宁丹霞3个世界遗产之后,2017年厦门鼓浪屿又在历经数年的申遗之路上修成正果——入选2017年世界文化遗产名录。

然而,福建世界遗产作为福建文化对外传播的重要媒介^②,相关研究极少,不利于福建世界遗产核心文化符号的输出。

有鉴于此,本研究以世界遗产翻译这一福建文化传播的重要途径为视角,以语料库的实证语料为依据,探究其存在的问题与对策。

1.2 福建世界遗产翻译研究现状

目前有关福建世界遗产的翻译研究并不多见,论文仅有10余篇,涉及武夷山、土楼及鼓浪屿等3个世界遗产。这些研究可概括为以下几个方面:

(1) 公示语翻译研究:针对武夷山景区标识用语翻译若干问题及其原因进行简要分析(陈孝静,2008);实地采集武夷山市双语公示语资料及相关采访,发现其英译存在表达不当、中式英语、过度翻译等12类问题并予以剖析(何兰芳,张美君,张素芳,林丽端,2016);调查留学生对武夷山市公示语英译的满意度,提出在改进现有公示语英译准确度的同时,增加重要公共场所双语向

① 例如,我国的黄山申报世界遗产成功之后,旅游收入从每年几百万元猛增到2亿元,地处云南的边陲小城丽江,仅2000年便接待了海内外游客258万人,旅游收入达到13.44亿元(刘红婴,王健民,2003:230)。

② 陈先元(2004:67)认为,从现代传播学的媒介理论来看,世界遗产是一种媒介,它的信息特征是历史阶段性和恒久性的统一,地域局限性和共享性的统一,文化汇聚性和延展性的统一。

导系统的设置及提供更多的诸如地图、各类宣传册及实时网络平台等双语信息媒介(何兰芳, 2018)。

(2) 外宣/旅游文本翻译研究: 以福建土楼世遗申报报告为语料, 以接受美学的视野融合为视角, 探析对外宣传文本的翻译策略(韦忠生, 2011); 提出“以信息为中心”的福建土楼旅游文本英译策略(潘涓涓, 2010); 以目的论为指导, 指出鼓浪屿旅游文本中的主要英译错误, 提出译者在翻译实践活动中可采用的翻译策略(汪懿婷, 2014); 从框架理论的视角, 分析客家土楼摄影文本的英译问题, 从语言框架的词汇、单句、语篇及文化框架中的文化缺省和文化冲突等方面对误译进行阐释并加以纠正(肖晓玲, 2015); 基于翻译适应选择论和多维转换原则, 审视福建土楼旅游外宣英译, 引入翻译生态环境、适应选择、适者生存等生态学概念, 并以此诠释和指导翻译实践(谢爱玲, 2016); 以功能加忠诚的翻译理论分析漳州土楼旅游文本的汉英翻译, 探讨有利于旅游文本翻译与本土文化对接的翻译策略和方法(李静雯, 2017)。

(3) 景点名称翻译研究: 探析武夷山景点名英译中存在的问题, 并分析其缘由, 讨论景点名梯级翻译标准, 论述与景点名翻译对应的5大类翻译方法(陈孝静, 唐有胜, 2011); 探讨中国世界遗产名称的英译策略, 提出译音与译意及两者并用的翻译策略(曾咪, 2014b)。

此外, 还有武夷山景区门票英译问题研究(刘金水, 吴婧, 2013)以及土楼申遗文本的中国文化遗产特色词汇翻译研究(曾咪, 2014a), 等等。

以上研究就福建世界遗产翻译做出了一定的探索, 但主要仍集中于公示语翻译、外宣/旅游文本英译、景点名称翻译等传统话题, 对诸如世界遗产网站的本地化、申遗文本英译的质量评估、导游词的英译等问题则均未涉及。此外, 在翻译研究的技术转向背景下, 相关研究未与语料库及CAT等翻译技术结合, 未见基于语料库的系统考察, 也未从CAT的角度探讨翻译质量的提升之道。福建世界遗产翻译的研究存在总量较少、视角传统、缺乏基于翻译技术的系统研究等问题。

1.3 语料库与世界遗产翻译研究

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3类。3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(李德超, 王克非, 2010)。

世界遗产翻译涉及的文本以旅游文本为主,此外还包括申遗文本及其他类型文本。就旅游文本而言,国外的旅游专门语料库主要有3个,分别为芬兰、英国和日本学者所研制,其共同特点是都属于容量小于100万词的小规模语料库。而我国内地除中科院研制的旅游咨询口语对话语料库和旅馆预订口语对话语料库外,尚无专门研发的旅游语料库,香港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(李德超,王克非,2010)。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,语料规模均为100万字/词。前者目前收录香港的英译汉和汉译英的旅游翻译文本,后者主要收录以香港为主的海峡两岸暨香港、澳门非翻译的中文和英文旅游原生性文本。该语料库可能包含一些与世界遗产有关的语料,但就整个语料库而言,并非以世界遗产为主,而且仅限于旅游类文本,没有包括申遗文本。

综上所述,目前国内外尚无以世界遗产为对象构建的专门用途双语语料库,而世界遗产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。

1.4 世界遗产翻译研究的 CCAT 视角

CCAT是将语料库与CAT技术相融合的翻译研究范式(王朝晖,余军,2016)。所谓CCAT,指的是Corpus and computer-assisted translation(语料库及计算机辅助的翻译),其中计算机辅助部分指CAT软件,而Corpus主要指双语语料库,以专门用途双语语料库为主,也包括网络语料库。

CCAT平台由参与人员、CAT软件、语料及在线平台构成,是一个完整的系统。

第一,参与人员。包括译者、译文审查者、译评人、出版商、翻译研究者等等。

第二,CAT软件。在国内外多款CAT软件中,雪人CAT是最适合CCAT平台的,因为其句子对齐功能强大,软件易用性突出,功能符合CCAT需要,且性价比高、支持定制。

第三,语料。包括单语语料库、双语语料库,以及CAT软件的翻译记忆库及术语库,统称语料。语料库研究领域的语料库多用于翻译研究,CAT领域的翻译记忆库和术语库则侧重于翻译应用。之前学界很少注意到两者的关系,其实两者是相辅相成的,代表同一事物的理论和应用两个方面。双语语料库与翻译记忆库可以轻易便捷地相互转化,CAT的术语库也可以从双语语料库中提取。

第四, 在线平台。在线平台指语料库及翻译记忆库、术语库的网络平台, 可以是共享性质的, 也可以是收费性质的, 视将来的发展而定, 可能两种都会存在, 只是语料规模不一样。这一在线平台和 CAT 软件是联系 CCAT 系统中译者、译文审查者、译评人、出版商、翻译研究者的纽带。

CCAT 这一理论包括应用和研究两个层次的内涵。应用方面, 不论是文学翻译, 还是非文学翻译, 如果在语料库及 CAT 的共同辅助下进行, 在翻译准确性、翻译效率等方面都将有可观的改善。研究方面, 不论是语料库翻译学研究, 还是 CAT 研究, 要深入下去, 突破瓶颈, 唯有二者融合, 优势互补, 方能开辟一方新的天地。

以上有关 CCAT 的阐述, 详见《基于 CAT 及语料库技术的电子商务翻译研究》(王朝晖, 余军, 2016) 一书。我们认为, CCAT 同样适用于世界遗产翻译研究。因此, 本研究旨在构建以福建世界遗产翻译为研究对象的专门用途双语语料库, 探讨双语语料库的构建及其与 CAT 技术融合为 CCAT 的途径, 基于 CCAT 平台, 评估福建世界遗产翻译质量, 探索 CCAT 平台下的世界遗产翻译模式, 提高福建世界遗产翻译的质量, 促进福建文化的传播和文化产业的发展。

1.5 本书主要内容

全书共分 7 章。第 1 章是绪论, 简要评述了福建世界遗产翻译的研究现状, 在此基础上说明本研究的目标和意义。第 2 章阐述了福建世界遗产双语语料库的研制, 包括设计思路、语料来源及建库类型、建库工具、语料标注、术语库制作、记忆库制作等方面的内容。第 3 章介绍了福建世界遗产翻译双语语料库的检索方式。第 4 章探讨了基于 CCAT 的世界遗产翻译的质量评估标准和模式, 并以申遗文本、景区文本、酒店介绍及导游词等为个案, 对福建世界遗产翻译质量进行了评估。第 5 章考察了福建世界遗产网站的本地化情况, 分析了存在的问题并提出对策。第 6 章提出了 CCAT 平台下的世界遗产翻译模式, 并以福建申遗文本翻译和酒店介绍翻译为例, 论述其具体应用。第 7 章以申遗文本为例, 分别说明了福建世界遗产双语语料库的精细修订与入库修订。

第2章 福建世界遗产双语语料库的研制

2.1 引言

世界遗产翻译 CCAT 系统构建的重要支柱是语料库，其研制是构建 CCAT 系统的第一步。福建世界遗产双语语料库以武夷山、泰宁、土楼、鼓浪屿的申遗文本、景点介绍、导游词、世界遗产本地化网站及相关图书等为主要语料来源，将 CAT 技术与语料库技术相结合，基于 CCAT 理念构建，属于一种新型的专门用途双语语料库。

本章介绍福建世界遗产双语语料库的研制，分为两部分内容。第一部分介绍福建世界遗产双语语料库的研制方案，包括该语料库的设计思路、语料来源及建库类型、建库工具、语料标注等；第二部分介绍术语库制作、记忆库制作及转换等。

第一部分是福建世界遗产双语语料库的构建个案，对其他类型的专门用途双语语料库也具有参考价值；第二部分是技术基础，适用于任何一种新型专门用途双语语料库的术语库及记忆库构建。

2.2 设计思路

基于 CCAT 系统的构建原则，我们认为，福建世界遗产双语语料库的研制应不仅包含传统的单语类比语料库及双语对应语料库，还应从语料库与 CAT 相互融合的理念以及 CCAT 的应用角度考虑，加入术语库和记忆库，以促进福建世界遗产翻译的术语统一，提高翻译效率和质量。因此，在福建世界遗产双语语料库的构建上，我们沿用了之前构建电子商务翻译语料库时设计的一种新型的专门用途双语语料库研制模式，与传统的双语语料库构建模式有很大的不同。

以往任何一种双语语料库一般都以一种文件形式存储（多为 TXT 或者 XML 格式），属于单一的形态，可用专门的检索工具检索分析，但由于文件格式不兼容，不能被计算机辅助翻译工具所利用（王朝晖，余军，2016：185）。而福建世界遗产双语语料库不仅包含兼容于语料库检索工具的 TXT 文件，也包含雪人 CAT 软件可以加载的术语库和记忆库文件。

这一新型的双语语料库，提供了与传统的双语语料库和当代 CAT 技术的双