

内蒙古生态旅游资源汉英翻译语料库的设计与应用

○耿素

(内蒙古大学 外国语学院 内蒙古 呼和浩特 010010)

[摘要] 本文以内蒙古生态旅游资源为初始元语料母本,采用线上线下的设计模式和应用模式。提出分布式智能语料搜集 APP 以及网络数据库特定搜索方式方法来实现语料线上的搜集,以及走访生态旅游景区以及相关博物院、史料馆的线下搜集。另外,对于无英译元汉语语料部分,采用翻译建立英语语料库,最后实行双库合并的方式完成汉英语料库建设。无英译元语料的翻译通过连接相关翻译专业师生团队进行。最后阐述汉英语料库“一库双模”的应用模式以及意义。

[关键词] 语料库;汉英翻译;生态旅游;语料搜集;一库双模

中图分类号:H 315.9

文献标识码:A

文章编号:1672-8610(2016)10-0025-03

一、引言

作为外语教学的一部分,翻译教学旨在传授基本翻译知识及日常的方法和技巧,通过反复实践培养学生的双语转换能力,使学生掌握必要的翻译知识与技能。但是在以教师为中心的傳統翻译教学中,最为突出的问题是教师把教学重点放在课程讲授和学生翻译结果评价而忽视了学生翻译的过程,在以“学生为中心”的思想指导下,学生在翻译过程中用什么翻译理论作为指导,利用什么工具,怎样利用这些工具,更为重要,而语料库是除了不同种类工具书更加方便、高效的翻译辅助工具。^[1]生态旅游作为一项天、地、人多元参与的综合性活动,人类在其中与外部环境保持着动态的交互联系,以实现知识库的更新,同时加深对外界的认识。语言作为人类参与此项交际活动最基本的工具,在面对新的语言事实和文化现象时,也在不断进行语言和语法的更迭,文化交流密集又频繁的旅游活动,急切要求一种活的语言更新管理体系提供区域生态资源和自然人文景观在世界文化背景下的天、地、人多元互动交流。而新信息经济时代背景下的翻译教学以及旅游翻译,也面临诸多挑战。比如学生面对翻译文本时可以深刻理解,却难以精确表达,旅游资源文化内涵和景观信息的国际化推广,缺少统一高效的翻译指导工具和参考。随着数据驱动型翻译教学的逐渐推广和旅游信息化的升级以及智慧旅游的兴起,为促进内蒙古生态旅游资源 and 生态旅游文化的广泛外宣,以及提升翻译教学改革在特殊语境中的质量和效率提供强大支撑。^[2]

二、研究意义

内蒙古的旅游资源丰富,尤以生态旅游资源突出,极具

地方民族文化特色。但是,这些特点为旅游业的对外宣传和对外翻译带来了难题,造成翻译不够规范,文化现象的翻译不够准确传神等问题。现有的大型通用语料库难以准确体现关于内蒙古生态旅游文本的文体特征和语篇功能,无法满足翻译研究与实践的检索统计需求。内蒙古生态旅游汉英翻译语料库(Translational Corpus for the Eco-Tourism of Inner Mongolia,简称TranCoETIM)是将大量高质量的实际应用汉英翻译语料经处理整合起来,是一项有意义而先期未能实现的项目。作为内蒙古首个特殊语域的汉英翻译语料库,它的成功构建与开发可以大幅度提高译者的翻译效率和翻译质量,将为内蒙古民族文化的对外传播和发展注入活力,有助于传播内蒙古“非遗”文化,推动内蒙古民族文化的国际化,提升内蒙古作为文化旅游重要地区的国际竞争力和文化软实力。具体有以下几项意义:

第一,辅助翻译。在TranCoETIM相应的web检索平台上输入关键字,即可获取包括关键字的所有英汉双语对应文本及语篇信息,从而可以在翻译实践和翻译练习中,为学生和翻译工作者等用户提供大量可靠的内蒙古生态旅游相关内容的英文表达,便于其快速、有效检索到生态旅游的专有名词和术语的英译。同时,语料库还提供了大量内蒙古生态旅游资源的英文介绍与描写,用户可进一步借用这类平行文本(parallel texts)提高英译文的质量,同时也可对这类旅游翻译文本进行对比研究,进而理解其结构及语体特征。此外,作为机器翻译的支撑数据库,TranCoETIM可丰富其后台翻译文本的储备,进而大大提高旅游翻译领域机器翻译的效率。

第二,对比研究。用户可以通过对此语料库中英译文本

[基金项目] 大学生创新创业训练计划项目 201509107。

[作者简介] 耿素,内蒙古大学外国语学院本科生。

的学习和研究,参考权威、正确的英文译本进行翻译质量的自我检测与评估,从而进一步完善翻译作品,学习翻译技巧,不断增强翻译能力。同时,通过对比分析旅游领域中英文本在风格、结构、功能上的异同,将自身心得体会与现代翻译理论相结合,用户可研究归纳旅游文本翻译的基本方法和原则,尤其是旅游文本翻译中的归化和异化现象。此语料库也为内蒙古历史文化词语的翻译研究提供了客观的数据和案例支持,为词典编纂和内蒙古文化英文书籍编写提供重要的资料来源。

第三,翻译教学。“一库两模”的架构为翻译教学提供了详细周全且以教学为导向的文本标注模式。教师根据对语料库信息的对比分析和辅助的标注信息,可进一步明确教学目标、优化课程设计,协助教学评估,实现教学理念的现代化。语料库的应用可以增加学生的参与程度,拓宽学生的视野;可以提高教师对学生翻译实践的指导水平,提高教学质量;翻译教学设计和教学内容更加符合内蒙古地区历史文化传播对于翻译人才的具体要求。另外,该汉英翻译语料库积累了大量内蒙古生态旅游文化英文表达词汇和例句,便于教师编制专门翻译教材和教辅资料。

第四,文化传播。英语版的旅游网站或主页对于推介我国的旅游资源,扩大信息在网络上的传播,吸引更多的境外游客到中国旅游起着不可忽视的作用。此语料库可为网络旅游资料文本创作和翻译提供参考和辅助,还可以为内蒙古生态旅游翻译者服务,为内蒙古外宣和旅游相关单位提供英文翻译参考依据,它的应用是内蒙古民族文化对外传播的创新性发展。与此语料库紧密结合的翻译教学培养的翻译人才将是传播内蒙古文化的重要力量。

三、研究设计

(一)设计特点^[4]

表格 1设计特点

语库规模	2 万词次
语料范围	内蒙古地区旅游部门的调查报告、官方宣传材料(如联合国教科文组织官网、英文报道等)、公开出版的书籍(汉英互译教材、国外出版的蒙古族历史文化著作)、申报报告(申报英文项目书、支撑材料等)。
语料质量	选取语言质量较高的原本,特别注意译本质量,避免太过死板的直译和太过活泛的意译。
语库架构	一库两模,同一的元语料文本,转换编辑成两种显示模式,翻译教学模式以及通用模式,翻译教学模式显示包含通用模式可以用于旅游翻译者的翻译参考。
语库性质	双语库、平行库、笔译和口译、专业与通用。
检索类型	B/S 模式(“浏览器/服务器”模式,在线、离线检索分析,APP 兼容)
语料类型	书面文本以及官方网络文本
语料分类	以生态旅游资源类型分类,见表 2。
翻译语向	汉英双向
双语比例	汉语原创文本、英语译本、英语原创文本、汉语译本

(二)设计思路

1.语料搜集。从各种渠道收集合格的、经过筛选的语料,包括从互联网上下载、人工录入和电子扫描,汉英双语匹配,即双语文本间具有翻译关系。前期人工搜集为主,机器搜集为辅;后期机器搜集为主(APP或者Web)人工搜集为辅。机器收集利用语料库社交型APP(主要功能为巧妙利用分布式客户端从游客手机相册里建议性提取在生态景区拍摄的有关生态旅游资源的照片,作为形成实际应用文本的第一手资料)。

表格 2内蒙古生态旅游资源分类

内蒙古生态旅游资源分类		
生态资源类型	生态旅游资源分布	生态旅游景区
草原资源	呼伦贝尔草原(呼伦贝尔市)	呼伦贝尔草原旅游区
	科尔沁草原(通辽市)	科尔沁草原湿地自然保护区、珠日河草原旅游区、大青沟国家级自然保护区、罕山国家级自然保护区
	锡林郭勒草原(锡林浩特市)	锡林郭勒草原自然保护区、锡林九曲、乌珠穆沁草原
	乌兰察布草原(乌兰察布市)	乌兰察布大草原、辉腾锡勒黄花沟景区、格根塔拉草原旅游区
沙漠资源	鄂尔多斯草原(鄂尔多斯市)	鄂尔多斯草原旅游区
	巴丹吉林沙漠(阿拉善盟)	巴丹吉林沙漠地质公园
	腾格里沙漠(阿拉善左旗)	沙坡头景区
	乌兰布和沙漠(巴彦淖尔市、阿拉善盟)	乌兰布和沙漠旅游区
	库布其沙漠(鄂尔多斯市)	响沙湾、七星湖沙漠生态旅游区、恩格贝生态旅游区
	毛乌素沙地(鄂尔多斯市)	毛乌素沙地自然风景区
	浑善达克沙地(赤峰市)	浑善达克自然风景区
森林资源	科尔沁沙地(通辽市)	科尔沁沙地自然风景区
	呼伦贝尔沙地(呼伦贝尔市)	呼伦贝尔沙地自然风景区
水域资源	大兴安岭—线原始森林	阿尔山国家森林公园
	大兴安岭南段次生林	大兴安岭南段次生林自然风景区
	黄河	老牛湾
	毕拉河(大兴安岭地区)	毕拉河自然保护区
	额尔古纳河(呼伦贝尔地区)	额尔古纳河
	额济纳河(额济纳旗)	胡杨林自然保护区
民俗资源	呼伦湖(呼伦贝尔市)	呼伦湖国家级自然保护区
	贝尔湖(呼伦贝尔市)	贝尔湖风景区
	(衣)蒙古袍、蒙古靴、帽等	此类民俗资源存在于上述生态旅游景区内
	(食)肉食品、奶食品、奶茶、奶酒等	
	(住)蒙古包、“搏罗子”等	
	(行)马、勒勒车、骆驼等	
	(生产方式)畜牧、狩猎等	
	(文化娱乐)那达慕、萨满舞等	

2.语料加工。元语料处理:将搜集到的元语料(汉英对照的文本)输入或编辑整理为word文档电子版。

对齐方式:主要应用Paraonc 软件进行对齐,先进行软件对齐,然后人工检验进行对齐,分两步进行:首先做到英汉双语在句子层级上对齐,可链接检索,以方便各种研究上的需要;其次是初步的分词处理和词性的标注;并预留今后在句法、语义等层面上深加工的余地。

标注^[5]:

(1)文本属性信息,即标题、作者、文体、语体、领域和创作时代等属性信息。

(2)文本结构信息,即标题、子标题、段落、句子以及词的边界等信息。

(3)双语对齐信息,即句子一级的对齐互译关系。

存储格式:汉英平行语料库采用XML 语言进行统一编码。

四、数据库建立

(一)语料收集

1.从正式出版物中提取元语料。通过在知网或词典中搜索关于内蒙古生态旅游以及相关文章,从中英文摘要中摘取目标元语料,以句或词为单位,建立一一对应的语料word文档。并注明来源及整理时间。例:

呼伦贝尔草原
Hulunbuir Pasture Land
英汉双向旅游词典
对呼伦贝尔草原 4 个时期的遥感监测表明:呼伦贝尔草原经历了“发展—加速发展—逆转”的沙漠化过程。 The results of remote sensing monitoring to Hulun Buir grassland during 4 periods show that the grassland has undergone a process of "developing to accelerated developing to reversing".
呼伦贝尔草原沙漠化土地动态变化过程研究 中国沙漠 第 29 卷 第 3 期 2009 年 5 月
《辽阔的草原》、《橘色的雄鹰》、《秀丽的海骝马》、《乌和尔图灰嘴》等巴乌虎尔乌尔丁歌,以其悠扬的旋律、丰富的诺古拉,给人们留下了深刻的印象。这些广为流传的巴乌虎尔民歌在内蒙古呼伦贝尔市陈巴尔虎旗和新巴尔虎旗两地的演唱风格却存在着细微的差别。 Vast Grasslands, Brown Male Hawk, Pretty Horses, Wuhertu Huiteng etc. In these songs, Barghu Urtindu greatly impresses people with its long melody and plentiful nogola. Barghu folk songs spread far and wide among people, yet its singing style exists certain differences between old Barghu and new Barghu in HuLunBuir city.
红梅. 从区域文化特点看呼伦贝尔巴乌虎尔民歌[J]. 中央民族大学, 2006, 5/8:2016

2. 从非出版物或者现有语料库中提取元语料。以谷歌翻译为例说明在线翻译的原理。Google翻译在生成译文时,会在经过人工翻译的文档中检测各种范例,进行合理的猜测,然后得出适当的翻译。这种在大量文本中查找各种范例的过程称为“统计机器翻译”。因此,可以在线翻译的文段,其中包含了人工翻译后的很多参考资料,从而利用微软必应翻译、谷歌翻译等对国外英文网站上关于内蒙古生态资源的介绍进行翻译,可以找出具有汉英翻译关系的对应语句和词汇,相当于从网络翻译利用的网络语料库里寻找目标语料。例:

对网页 <http://www.chinahighlights.com/hulunbuir/> 的翻译



3.线上与线下结合。基于前期的调查和搜集发现,发现当前内蒙古生态旅游景区建设中明显滞后的一个环节就是英译工作。景区中许多景点介绍、标识牌等缺乏中英对照文本,即使有,也是错误百出。故而分两步进行,先完成内蒙古生态旅游资源汉语语料库,后对其进行英译,建立内蒙古生态旅游资源英语语料库,最后完成内蒙古生态旅游资源汉英双语平行语料库的构建。

首先是对汉语语料的取样。语料来源按照表2中划分,既包括正式出版物(图书、论文、音像制品等)、非出版物(网页、各类旅游宣传材料如酒店介绍、旅行社介绍、景点旅游宣传册等)。我们在建库过程中充分利用内蒙古旅游局(<http://www.nmgtour.gov.cn/>)和各盟市旅游局以及相关协会的官方网站等权威网站的网页内容。此外,我们深入5A级景点呼伦贝尔大草原、鄂尔多斯响沙湾、成吉思汗陵旅游景区等进行实地考察,获取关于景点介绍、景区标识语等信息的第一手语料,并进行录入,实现内蒙古生态资源汉语文本的数字化。

我们在完成汉语语料的取样基础上对内蒙古生态资源汉语文本进行英译,以权威统一的翻译理论作为翻译实践的
指导,由英语技能较高、翻译能力较强、理论功底扎实的课题

组承担并完成满足译语读者需求的翻译任务,最后建成并扩充内蒙古生态旅游汉英翻译语料库(TranCoETIM)。

(二)语料标注对齐

所有电子化的汉 & 英语料文本 ,最后实现语料的标注、对齐。语料标注分为词性标注与修辞标注两类。语料标注可以借助标注软件(比如SVMTool或者CLAWS7),中文语料还通过中科院计算所研制的汉语词汇分析系统(ICTCLAS)进行切分得以实现 ,而修辞标注则无现成的电脑软件进行辅助 ,完全需要人工进行。语料对齐首先启用语料对齐软件(比如Paraconc)以句为单位进行粗略对齐 ,辅之以人工校对 ,以备检索系统(比如concordancer)检索。在语料实现对齐后 ,文本转换为XML 格式 ,这样就能让语料库以多种语言格式编码 ,以便传播与检索、语料平行对齐语料库维护更新。^[6]

(三)语料库的维护和更新

设置管理员权限分级,对语料库进行不同层级的维护和更新。一级进行自动搜集或者人工搜集后的整理、审核、编辑、入库。二级对本文主张的“一库二模”(即教学模式和通用模式)的编辑、修改、完善。

五、结语

综上所述,我们通过语料取样、文本英译、语料标注、文本对齐等步骤初步创制内蒙古生态旅游资源汉英双语平行语料库,并将其应用于旅游翻译、教学研究,有着重要的实践应用意义和研究意义。此语料库填补了内蒙古生态旅游领域小型语料库的空白,有助于推动培养兼具本土化与国际化特点的英语人才,进而促进当地生态旅游文化产业发展和跨文化传播。基于语料库快速发展的现状,此后还将扩展到内蒙古其他特殊语域,结合不断更新发展的计算机技术,将为翻译教学、语言学及其他领域的研究贡献更大的力量。

【参考文献】

- [1]王克非,秦洪武.论平行语料库在翻译教学中的应用[J].外语教学与研究 2015(05).
- [2]徐肖嵘,江婧婧.皖南旅游资源翻译语料库的构建及其应用研究[J].海外英语 2011(09).
- [3]杨凤.国内近十年旅游文本翻译研究综述[J].海外英语 2012(01).
- [4]王克非.中国英汉平行语料库的设计与研制[J].中国外语, 2012(06).
- [5]常宝宝,俞士汶.语料库技术及其应用[J].外语研究 2009(05).
- [6]肖庚生,陈欣.旅游汉英双语平行语料库的建设与应用[J].湖南科技学院学报 2012(10).