

西藏旅游外宣汉英平行语料库的

构建设想及其应用价值

□李伟超 宋亚军

【内容摘要】本文阐述了建设西藏旅游外宣汉英平行语料库的必要性和可行性,提出了建设西藏旅游外宣汉英平行语料库的构想,并在语料选取、语料录入与保存、语料标记与标注、语料对齐等方面提出了详细的构思,同时也展望了西藏旅游外宣汉英平行语料库的广阔应用前景。

【关键词】西藏旅游;旅游外宣;汉英平行;语料库;建设构想;应用前景

【作者简介】李伟超(1987~),男,河南开封人;西藏大学旅游与外语学院助教,硕士;研究方向:语料库语言学、应用语言学
宋亚军,西藏大学旅游与外语学院

一、构建西藏旅游外宣汉英平行语料库的必要性和可行性分析

(一)构建西藏旅游外宣汉英平行语料库的必要性。外宣翻译是要完成那些以对外宣传材料的翻译任务为基本内容的翻译实践活动的总称(张健,2013)。对旅游业出现的文本进行的翻译实践活动可称之为旅游外宣,其目的在于“外”,即向来自海外的旅游者解释、说明、推介相关旅游资源,其传播方式在于“宣”,即以各种媒介为渠道,对来自海外的旅游者宣传相关旅游资源。旅游外宣翻译的质量高低,在一定程度上决定了旅游资源的宣传效果以及海外旅游者的旅游舒适度。通过中国知网CNKI的检索发现,学者们对西藏旅游外宣翻译的研究并不多见,但也有学者做了深入的研究。比如谭益兰(2013)通过运用语篇翻译的相关理论,选取了布达拉宫、罗布林卡和拉鲁湿地三个旅游景区中的地图、旅游指南、门票、旅行社宣传资料等的公示语篇,详细分析了此类语料库汉英翻译存在的问题,并提出了提升西藏旅游景区公示语语篇汉英翻译质量的具体对策与建议。此外,弋睿仙、贾丽萍(2014)运用赖斯的翻译批评理论,对布达拉宫门票上的汉英翻译质量进行了评估,从文本类型、语言内因素和语言外因素对译文进行了全面、客观的分析,发现了译文在语义、词汇和语法等方面存在的一些不足之处。上述两个研究针对西藏旅游外宣文本的汉英翻译进行了个案研究,西藏旅游外宣文本汉英翻译现状不容乐观,存在着的大量问题亟待解决。随着语料库语言学以及语料库翻译研究的发展,不少学者提出建设旅游外宣类汉英平行语料库为解决旅游外宣文本翻译质量低下的问题提供了更为实用、方便的途径。

(二)构建西藏旅游外宣汉英平行语料库的可行性。语料库(corpus)是应用计算机技术对海量自然语言材料进行处理、存储,以供自动检索、索引以及统计分析的大型资料库(李文中,1999)。在建设语料库的过程中,首要的任务就是搜集海量的自然语言,在语料稀少的情况下,是无法建设成

具有一定规模的语料库的。西藏自治区旅游资源丰富,充斥着大量的汉英对照的旅游文本,包含公示语、景观介绍、门票信息、交通地图、酒店餐饮介绍、博物馆展品说明、特色旅游纪念品介绍等等。此外,每年西藏都有大型的节庆、博览展出等活动,比如藏历新年、赛马节、雪顿节、林卡节等一系列具有民族特色的节庆活动。可以说,随着西藏旅游走向国际化的道路,旅游外宣文本的多语对照已经形成常态,我们可以轻而易举地获得大量的语言数据。

二、西藏旅游外宣汉英平行语料库的构建设想

(一)建库目的。语料库并不是大量自然语言数据的简单集合,它是有严格的设计标准的。就语料库的应用目标而言,其收录的语言材料应该具有代表性(常宝宝、俞士汶,2009)。这就要求我们在建库前明确哪些语言材料是建库所需要的,同时应明确哪些语言材料是应剔除的。对于本研究中提出的西藏旅游外宣汉英平行语料库建设来说,要建设高质量、精品化、本地化的旅游外宣汉英平行语料库,使其能很好地服务于西藏旅游外宣汉英翻译的工作;服务于西藏区内高校旅游专业学生翻译教学的工作;服务于西藏区内旅游从业人员英语培训提高的工作。因此,这就对建库过程中的语料选取提出了更高、更优的要求。

(二)语料选取。在构建平行语料库时,其中一个基本原则就是平行文本的来源和领域必须严格界定(李文中,2010)。针对本研究中语料库的语料选取问题,首先要明确的是旅游外宣文本的含义。广义的外宣翻译涵盖所有的对外宣传方面的翻译活动,狭义的外宣翻译包括各种媒体报道、政府文件公告、政府及企事业单位的介绍、公示语、信息资料等实用文体的翻译(张健,2013)。此外,根据上述的语料库的建库应用目标,本研究提出的西藏旅游外宣汉英平行语料库侧重于旅游文本在对外宣传过程中的汉英翻译,而非外宣文本的汉英翻译。本研究中的旅游外宣文本属于外宣文本中的一个类别,它包含了所有在旅游活动中出现的,起到对外宣传作用的文本。因此,在语料选取过程中,主要

搜集以下几类汉英对照的旅游文本材料:景观介绍;观光指南;景区公示语;餐饮住宿介绍;博物馆陈列说明。通过购买印刷材料、收集网站资料、实地考察旅游景区等方式获取相关语料。确定了语料来源之后,另一个需要解决的问题就是语料的代表性问题。采取根据社会语域的标准,利用分层抽样、均衡抽样和比例抽样三者结合的方法,使得西藏本地旅游业中具有代表性的汉英翻译文本成为入库语料。由于本语料库是应用型语料库,主要用于为西藏旅游外宣翻译提供可靠的参考,以及西藏本地高校旅游英语翻译教学,那么必须要建设成为高质量、精品化、本地化的旅游外宣汉英平行语料库,也就需要对收集到的文本中的语言材料的质量进行评估,以免质量不高的语言材料入库,对以后的研究和英语产生负面效果。因此,在语料入库时要注意以下几个方面:一是含有语法、搭配、拼写错误的文本,略加更正后入库;二是出现译文表达不准确甚至错误的文本不予入库;三是组织翻译专家对入库语料进行二次审核把关,剔除或修改质量不高的译文。

(三) 语料录入与保存。由于语料来自于印刷材料、网络电子文本、拍摄下来的图片资料,因此需要大量的人力物力将这些语料进行数字化的转换。对于纸质的印刷材料,可以采取先扫描成 PDF 格式的文件,再转换成 TXT 格式文本的方法;对于网络文本,其转换方法很简单,需要注意的就是剔除嘈杂信息;然后拍摄的图片资料则需要人工输入。所有入库语料经过扫描、转换、剔除网络嘈杂信息等过程后,经过多次的文字校对和格式整理,全部保存为 TXT 格式的文档备用。

(四) 语料标记与标注。对于语料库的标注的定义,李文中(2012)认为各家观点不同,但分歧不大,因此他主张使用“标记”和“标注”两个术语,语料库标记表示添加与文本结构正字特征的表达和有关文本识别特征的过程,语料库标注则表示词性标注、语法标注、话语标注以及错误赋码、语用特征赋码等。关于语料库的标注,李文中(2012)指出语料库学界对于语料库标注支持和反对的两种截然不同的态度,不仅反映了语料库研究中两种既个性又相辅相成的学术思想(即基于语料库的研究与语料库驱动研究),也反映了研究者对待语料库中语言数据态度和使用路径的不同,但二者在研究实践中并不是互相排斥的。他指出对语料库设计标注或应用标注时,要立足于自己的研究实际。对于本语料库的应用来说,在设计标注时只考虑对文本的元信息进行标记,以及对中英文进行词性标注。文本的元信息指的是关于文本的信息,在本语料库中,在进行元信息标记的时候,采用多数语料库常用的 XML 语言进行标记,主要注明语料来源、景区类别、景区所在地、译者信息等。为方便进一步的检索和研究,对英文文本采用免费的 CLAWS7 软件进行计算机自动词性标注。然而机器自动标注的准确率只能达到 96% 左右,因此,在完成自动标注后需要对标注后的文本进行人工修正。

(五) 语料对齐。相较于单语语料库,平行语料库需要进行双语语料的对齐这一必不可少的步骤。对于语料对齐的问题,目前大部分平行语料库能普遍能做到的是段落对齐和句子对齐,目前来说,还没有对齐率达到百分百准确性的对齐软件,因此,在进行语料对齐时要做到机器对齐和人工对

齐相结合的办法。Paracon 软件可以对语料库中的中英文文本进行粗略的句子对齐,由于旅游文本的翻译会出现摘译、扩译或省译的情况,所以机器自动对齐软件在某些句子无法实现一对一的对应,因此会造成对应混乱,这就要求进行人工核查,完成进一步的人工对齐。在人工对齐的时候,一句中文原文有可能译成一句英文,因此实现一句中文对应一句英文;一句中文原文有可能译成多句英文,因此实现一句中文对应多句英文;多句中文原文有可能译成一句英文,因此实现多句中文对应一句英文。

三、西藏旅游外宣汉英平行语料库的应用价值

一是研究者可以从旅游文本的汉、英语料的对比研究入手,从文体和功能方面探究二者的差异和共性,发现不同类型的西藏旅游文本及其英译文本的语言特点,进一步结合翻译学理论,探讨西藏旅游文本外宣翻译的翻译策略、翻译方法及其内在成因,为今后旅游文本的翻译提供可靠的理论和实践基础。二是由于在建库时严格控制入库语料,所以建设成的西藏旅游外宣汉英平行语料库可以为译者提供可靠的翻译参考,能够有效地提高旅游文本翻译的质量,达到良好的外宣效果。三是西藏旅游外宣汉英平行语料库可以应用与西藏本地高校的旅游翻译教学。基于该语料库可以开发数据驱动式(data-driven learning,简称 DDL)旅游翻译教学手段。DDL 旅游翻译教学内容丰富多彩,教学形式也更加多样化,能够有效地培养学生主动学习、独立思考的学习习惯,进一步提高旅游翻译教学的效果,为培养服务于西藏旅游业的专业人才提供更好的教学方法。

四、结语

在建设西藏旅游外宣汉英平行语料库的时候,需要大量的人力物力财力,需要旅游专业人士、翻译专业人士以及计算机专业人士的共同协作和参与。尽管西藏旅游外宣汉英平行语料库的建设并非一蹴而就,但可以肯定的是,建成后的语料库其应用前景广泛,必将对西藏旅游翻译和旅游翻译教学产生积极的影响。

【参考文献】

- [1]常宝,俞士汶.语料库技术及其应用[J].外语研究,2009,5:43~51
- [2]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
- [3]李文中.语料库、学习者语料库与外语教学[J].外语界,1999,1:51~55
- [4]李文中.平行语料库设计及对应单位识别[J].当代外语研究,2010,9:22~27
- [5]李文中.语料库标记与标注:以中国英语语料库为例[J].外语教学与研究,2012,3:336~345
- [6]谭益兰.从语篇分析角度研究西藏旅游景点公示语汉英语篇翻译[J].西藏大学学报(社会科学版),2013,4:57~64
- [7]弋睿仙,贾莉萍.赖斯翻译批评在旅游文本翻译中的应用——以布达拉宫门票汉英翻译为例[J].牡丹江大学学报,2014,3:125~127
- [8]张健.全球化语境下的外宣翻译“变通”策略刍议[J].外国语言文学,2013,1:19~27