

旅游汉英双语平行语料库的建设与应用

肖庚生 陈 欣

(南华大学 外国语学院, 湖南 衡阳 421001)

摘 要: 文章在综述现有旅游语料库的基础上, 介绍了自建的衡阳旅游汉英平行语料库的建库步骤, 即语料的取样、英译、标注与对齐, 并阐述了该库在旅游文本语言特征、旅游翻译、旅游英语及旅游翻译教学研究中的应用价值。该库的建设与应用既能充分满足研究需要, 亦能促进导游、翻译等英语人才的培养, 进而推动当地旅游产业的发展及跨文化传播。

关键词: 旅游语料库; 双语平行语料库; 旅游英语

中图分类号: H030

文献标识码: A

文章编号: 1673-2219 (2012) 10-0163-03

近 30 年来, 语料库及语料库语言学从无到有、蓬勃发展。基于语料库的语言学及应用语言学研究应运而生, 语料库这一研究工具与方法业已广泛应用于词汇、句法、语义、语言对比、词典编撰、二语习得、翻译等领域的研究之中, 并取得了累累硕果。^[1]为了研究或应用的需要, 国内外先后创建了跨学科、多语言、内容迥异、库容不一的众多语料库。近年来, 双语平行语料库由于其独特的优势备受关注, 其建设与应用正如火如荼地进行。不过, 迄今为止, 国内外高校与研究机构所研制的平行语料库主要是通用型的双语平行语料库以及文学翻译类型的平行语料库, 以某一专门用途为导向的平行语料库仍较为鲜见。此外, 基于平行语料库的相关研究主要侧重于对个别语言或翻译现象的探讨。鉴于这一研究现状, 我们拟以课题研究为契机, 建设衡阳旅游汉英双语平行语料库, 在此基础上, 力图摸索其在汉语与英语的旅游文本语言特征、旅游汉英翻译、旅游英语与旅游翻译教学研究中的应用。

一 双语语料库与旅游语料库综述

双语语料库最早创建于上世纪 90 年代中后期, 迄今为

止国内外业已建立多种类型的双语平行语料库。早在 1995 年, 曼彻斯特大学研制了世界上首个翻译语料库, 主要收集从各国语言翻译成为英文的文本, 如今库容已达千万词。但是它并非严格意义上的双语平行语料库, 因为它并没有实现双语对齐, 只能称其为对比语料库。目前国际上比较知名的平行语料库有挪威的“英语-挪威语平行语料库”、英国的“德语-英语文学文本平行语料库”等等。而由北京外国语大学中国外语教育研究中心于 2004 年启动创建的“新型双语对应语料库”(含汉英、汉日两个双语平行语料库), 库容为 3000 万词次。其中, 汉英双语平行语料库的文本类型主要涵盖人文类、社科类和科技类。该语料库的建设得到了国家社科基金重大规划课题的资助, 目前正为 1 亿词次的超大库容做进一步的扩容努力。此外, 国内比较知名的双语平行语料库还包括: 李德俊以汉英词典研编为主要目的, 主持建设 2000 万词次的英汉平行语料库; 卫乃兴以研究探讨英汉对等表达为目的, 主持建设 900 万词次的英汉平行语料库; 哈尔滨工业大学所创建的库容为 40 万对的英汉双语平行语料库; 绍兴文理学院创建的中国古典文学英译双语平行语料库等等。上述双语平行语料库多数以文学文本为主, 抑或以收集百科文本(含文学与非文学作品)为特色, 而针对某一特定文类或某一地域特色而专门建构的专门用途双语平行语料库仍较为鲜见。

近年来, 随着交通与通讯技术的日新月异, 世界各地人们之间的科技、文化、商务交流、合作与往来日益频繁。更为重要的是, 各国人们纷纷走出国门, 畅游世界, 体验世界各地的异域风情, 许多国家的旅游产业获得了飞速发展。为更好地发展旅游产业、传播旅游资讯、吸引外国游客, 同时亦为了满足语言学及应用语言学研究需要, 旅游专门语料库应运而生。

收稿日期: 2012-08-26

基金项目: 衡阳市社科基金项目“顺应论视阈下衡阳旅游汉英双语平行语料库的建设与应用研究”(项目编号 2011C011); 全国教育科学规划课题“基于语料库的大学英语教学平台建设与研究”(项目编号 GPA105029) 阶段性成果。

作者简介: 肖庚生 (1980-), 男, 湖南衡阳人, 讲师, 应用语言学博士生, 主要研究方向为语料库语言学。陈欣 (1983-), 女, 湖南郴州人, 讲师, 语言学硕士, 主要研究方向为翻译理论与实践。

到目前为止,国内外最知名的旅游语料库有如下四库:一是芬兰萨翁林纳翻译研究学院语言教师 Michael Wilkinson 所创建的库容为 67 万词的英语旅游文本语料库,旨在为芬兰学生在将芬兰旅游文本英译时提供参考标准;二是英国埃塞克斯大学研究生 Carlota Alcantar 于 2007 年创建的库容为 37,795 词的英语旅游文本语料库,其内容包括旅游行政机构的介绍、酒店及旅行社信息等,旨在为旅游英语教学总结旅游写作常用的词汇与短语;三是日本大学 Kiyomi 等人于 2006 年所构建的“京都旅游语料库”。该库为英语单语语料库,收集的材料均为用英文撰写的有关京都旅游介绍的文本,共 885 篇,建库目的在于总结适合于日本旅游市场(尤其是京都地区)常用旅游英语词汇,为旅游英语课程提供教学参考;四是香港理工大学中文及双语学系所创建的新型双语平行旅游语料库,是中国第一个较大库容的英汉/汉英旅游语料库,它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,库容为各 100 万字/词。旨在促进香港旅游翻译研究以及旅游英语教学。

尽管世界各地的旅游语料库建设工作取得了丰硕的成果,但鉴于各地有着各自独特的自然与人文景观,各地旅游资源迥异,因此旅游语料库的本土化至关重要。为了对衡阳旅游资源进行跨文化传播,并满足语言学与应用语言学研究需要,促进在衡高校旅游英语及旅游英汉互译教学,我们自建了衡阳旅游汉英双语平行语料库(Hengyang Tourism Chinese & English Bilingual Parallel Corpus, HTCEBPC)。

二 HTCEBPC 的创建

HTCEBPC 是一个开放式的语料库,也就是说,其库容将会随着旅游宣传或研究需要不断得到扩充。HTCEBPC 的内容主要涵盖以下几个方面:衡阳自然景观、人文景观的介绍;介绍衡阳旅游资源的观光小册子;衡阳本土酒店、旅馆、餐饮业及娱乐场所与设施的介绍;与衡阳旅游业相关的其它行业的文本,如交通运输、导游服务、旅行社业务、货币兑换、投诉处理、景区公示语及旅游管理部门的宣传等相关的文本。HTCEBPC 的建库工程主要包括以下三个步骤:

首先是对汉语语料的取样。语料的选择主要借鉴 Maia^[2] 的标准,依据社会语域(如分为景点介绍、观光介绍、景点公示语、旅游配套设施介绍等若干类),将分层抽样、均衡抽样和比例抽样有机结合使用,以确保入库语料的代表性。语料来源既包括正式出版物(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衡阳市旅游局与衡阳市旅游协会的官方网站-衡阳旅游网(www.hyly.com)以及南岳衡山官方网站(www.nanyue.net.cn/)等权威网站的网页内容。此外,我们亦深入 5A 级景点南岳衡山以及回雁峰、石鼓书院、王船山故居、蔡侯祠等著名景点进行实地考察,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。同样,我们亦通过利用网络资源与实地调研相结合的方式获取有关衡阳林隐假日酒店、神龙大酒店等旅游配套设施的相关语料。实现衡阳旅游汉语文本的数字化,建成衡阳旅游汉语语料库(Hengyang Tourism Chinese Corpus, HTCC)。

接着我们在完成汉语语料的取样基础上对衡阳旅游汉语文本进行英译。笔者曾于 2009 年参加了南岳衡山申遗活动中的口、笔译工作,发现当前南岳景区建设中明显滞后的一个环节就是其英译工作,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。其他景区的旅游英译则更是一块未曾开垦的处女地。我们以 Verschuere^[3] 所提出的顺应论为翻译实践的理论指导,将旅游文本英译的全过程视为一个不断顺应和选择的过程。我们在理解原文的阶段,运用顺应理论,可以通过认知语境进行推理和探究,找出原语中的心理因素和文化特质,从而加深对旅游文本交际意图的理解。与此同时,在用译语进行重新编码时,运用动态顺应的原理,在众多的语言选择中,找出最能让译语读者接受的语项,从而顺利地满足译语读者需求的翻译任务。尽管对旅游文本的汉译英是一项艰巨浩大的工程,但庆幸的是,笔者所在的课题组其组员普遍英语技能较高、翻译能力较强、理论功底扎实,他们发扬精诚合作的精神,顺利完成相关语料的汉译英工作,建成衡阳旅游英语语料库(Hengyang Tourism English Corpus, HTEC)。

最后则实现语料的标注、对齐。语料标注分为词性标注与修辞标注两类。HTCC 的词性标注将通过运用 SVMTool 得以实现,SVMTool 在中文词性标注中具有较高的准确率。而 HTEC 的标注则以 CLAWS7 来得以实现,它的标注准确率已达到 96% 以上。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。HTCEBPC 中、英文语料都首先启用 Paraconc 软件以句为单位进行粗略对齐,辅之以人工校对,以备 concordancer 检索所用。而中文语料还通过利用中科院计算所研制的“汉语词汇分析系统”(ICTCLAS)进行切分。不过需指出的是,我们在顺应论的指导下,对旅游文本的翻译并非完全按照句对句的原则,我们灵活采纳摘译、扩译或省译等翻译策略。故软件在某些地方难以实现一对一的对应。我们便在运用 concordancer 检索相关的句子时,设定同时出现相对应的句子的相邻的几个句子,通过提供详尽的语境,能较为准确地了解是哪一不奋斗原文被改动。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。

三 HTCEBPC 的应用

目前,本旅游语料库只是初步建成,今后将会逐步增加语料扩充库容,并考虑将语料库上网或制作为可供检索的光盘。同时,我们还将不断完善对语料库的标注,依据教学与研究需要,进行语义与句法方面的标注,最大限度地挖掘语

料库的应用潜能。我们业已或将要在以下几个方面开展基于HTCEBPC的研究,以实现其应用价值。

首先,旅游文本语言特征研究。旅游双语平行语料库的构建为我们提供了鲜活、真实、海量的有关景点介绍、景区公示语、酒店文宣等内容的语料。对这些旅游文本的分析有助于揭示汉英两种语言旅游文本在词汇、句法、语篇以及文体上的语言特征的共性和差异。借助于语料库定量数据分析,我们可以得到旅游文本标准化类符/型符比,平均词长、平均句长、词类的出现频次与频率、词块的出现频次与频率等定量数据信息。我们已初步发现,英语旅游文本通常具有用词平易、丰富、通俗易懂,注重语篇衔接,强调呼唤功能等特点,而汉语旅游文本则表现为数量表达使用过于频繁,惯于堆砌华丽辞藻,常用隐性语篇衔接手段。这种基于语料库的实证研究,可以避免传统语言研究中思辨法或内省法过于经验主义的缺陷,具有较高的信度。

其次,旅游翻译研究。基于旅游平行语料库的翻译研究涵盖微观与宏观两个层面。从微观层面来讲,主要侧重于旅游文本翻译过程中词法和句法层面的转化规律与机制,前者如词类转换情况,后者如各种句式的转换情况。从宏观层面来讲,通过实证与量化的方式对旅游文本翻译的文体、风格、翻译共性进行研究,依据海量语料对译者个人偏好的语言表达形式加以分析,从中发现更有说服力的翻译文体/风格表征及其表现手段。这种基于双语平行语料库的研究方法能直观地揭示汉英两种语言旅游文本从词汇到语篇各个语言层面的共性与差异,有助于提炼旅游文本独特的汉英互译策略。李德超、王克非^[4]曾基于汉英平行语料库,分析与归纳中英文酒店文宣材料在文本惯例上所呈现的不同特点,并提出切实可行的酒店文宣翻译策略,旨在实现译文与目标语平行文本间的“文本内连贯”。基于平行语料库的旅游文本翻译研究其成果有助于解决当前许多地方景点名称、景点介绍、景区公示语、酒店文宣等文本翻译中存在的语言不规范、译名混乱、译文可接受性差等诸多问题,为衡阳本地乃至全国旅游翻译实践提供富于价值的参考。^[5]

最后,旅游英语以及旅游翻译教学研究。该库的建设为旅游英语以及旅游翻译教学提供了非常真实、鲜活、直观而且丰富的教学素材,通过引导学习者检索与考察语料库内容,促进学习者对汉英旅游文本的微观层面(字、词)到宏观层面(句子间联系、语篇结构以及修辞特点)的理解,鼓励他们阅读与讨论汉英平行语料间语言特征的异同并归纳切实可行的翻译策略。^[6]基于汉英双语平行语料库的旅游英

语以及旅游翻译教学其形式由传统的单纯由教师讲授转变为学生主动发现、学习与总结辅之以教师点评这种多样化形式,这一新型教学模式有助于引导学习者开展“归纳式学习”

(inductive learning),促进学生在自主学习中积极挖掘与批判性思考,能有效活跃课堂气氛、充分激发学生的学习兴趣。^[7]此外,值得一提的是,衡阳旅游汉英双语平行语料库具有鲜明的本土特色,其语料主要来源于衡阳地区景点介绍、景区公示语、酒店文宣及其他旅游配套相关文本,因此,在当地高校旅游英语或旅游翻译教学中,这些教学素材的应用无疑有助于提升学习者对本土历史、地理、文化的了解,亦有助于增强学习者跨文化交际能力。因此,该库的建设与应用对培养立足衡阳、面向世界的导游、翻译人才具有极大的推动作用。

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衡阳旅游汉英双语平行语料库,并将其应用于旅游文本语言特征、旅游翻译、旅游英语及旅游翻译教学研究之中。该库的建设与应用既能充分满足语言学及应用语言学研究需要,而且被初步证明有助于推动培养兼具本土化与国际化特点的英语人才,进而促进当地旅游文化产业发展和跨文化传播。

参考文献:

- [1]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.
- [2]Maia,B.“Some languages are more equal than others”: Training translators in terminology and information retrieval using comparable and parallel corpora[A].In F.Zanettin,S. Bernardini&D.Stewart(eds.).*Corpora in Translator Education*[C].Manchester:St.Jerome,2003.
- [3]Verschueren,J.*Understanding Pragmatics*[M].London: Edward Arnold,1999.
- [4]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.
- [5]乌永志.文化遗产类旅游景点名称汉英翻译规范研究[J].外语教学,2012,(2):93-97.
- [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]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.

(责任编辑:王晚霞)