

国内旅游语料库研究述评

朱志娟,朱利勇

(江西科技师范大学,江西南昌 330013)

摘 要:文章对国内旅游语料库的研究现状及其成果进行了梳理,发现:国内旅游语料库的研究主要集中在英汉/汉英双语平行或可比旅游语料库的研制/构建和应用两个方面。研制/构建研究内容主要包括语料库的设计、语料的搜集、选取、数字化、标注、对齐和检索等。应用研究方面主要集中在教学、翻译和旅游文本特征的研究。整体而言,目前国内旅游语料库的研究已经取得了一定的成绩,但是构建的库容较小,所涉及的地域有限,应用研究也有待进一步开发。

关键词:旅游语料库;构建;应用;教学;翻译

中图分类号:H085.3

文献标识码:A

文章编号:1007-3558(2015)04-0027-05

旅游语料库是出于某些特定的研究目的而收集旅游领域的语料文本建立起来的语料库,属于专门用途语料库。就目前研究来看,国内旅游语料库的研究主要集中在平行/双语语料库的构建及应用方面。平行语料库指的是两种语料分别存储在同一语料库中,且是对应关系,一种文本是另外文本的译文。国内的旅游语料库库容都不是很大,多为地域性的旅游语料库,具体应用领域较为狭窄,多为教学、旅游文本特征方面的应用研究。构建的类型主要为双语平行语料库,最为知名的香港理工大学的李德超课题组所构建的语料库主要来源于香港地区的双语语料库,主要应用于旅游英语教学和翻译教学。其它均为市或者地区的语料库,如保定市旅游语料库、安康生态旅游语料库、康巴藏区旅游语料库、赣南红色旅游语料库等。

虽然目前国内旅游语料库的研究和应用比较有限,但也取得了一定的成绩。因此非常有必要对目前旅游语料库相关研究进行述评,肯定已有的成绩,找出不足,探索将来的研究方向。

一、国内旅游语料库的构建研究

目前国内现有旅游语料库主要有安康生态旅游资源汉英翻译平行语料库、导游词平行语料库、贺州地方性旅游景区翻译语料库、赣南红色旅游英汉平行语料库、衡阳旅游汉英平行语料库、山东旅游外宣小型语料库、陕西省旅游景区公示语翻译语料库、皖南旅游资源翻译语料库和新型双语旅游语料库等,均为地域性的旅游语料库,为当地的旅游业服务。这些语料库都是双语语料库,且都为双语平行语料库,只有李德超的新型双语旅游语料库例外。该语料库不仅具有双语平行语料库构成,还有双语可比语料库,是一个双轨语料库。而该语料库也是在所有旅游语料库中最为知名、库容最大的旅游语料库,由香港理工大学构建,是以两岸三地

收稿日期:2015-06-02

基金项目:江西省社会科学艺术规划项目(YG2014067)资助,江西科技师范大学重点科研项目(2014XJZD013)基金支持

作者简介:朱志娟(1978-),女,河南中牟人,硕士研究生,副教授,从事英语教育和翻译研究;朱利勇(1977-),男,江西高安人,硕士,讲师,从事英语教育和翻译研究。

的语料其中主要以香港地区旅游语料为主的语料库。由此可见,国内学界构建的旅游语料库均为双语语料库。这些旅游语料库都对旅游语料库的设计和语料库的构建过程进行了描述。

(一)语料库的设计

语料库的设计主要包括语料库构建的目的、语料选取的范围(包括语料选取的地域性和规模等)。语料库构建的首要目的,无一例外都具有翻译之用的目的。目前最为有名的旅游语料库,香港理工大学李德超课题组所构建的语料库,设计目的有三个。其一是研究用途,研究旅游语言本身特点和旅游翻译特点;其二是翻译教学用途,为中英双语旅游翻译教师提供真实的教学材料;其三是实用性的目的,主要为国内外从事涉外旅游的英汉语从业人员提供旅游翻译参考,提高他们的业务能力。由此可见,虽然李德超把他的语料库目的细分为三个,但总体上都和翻译相关,为翻译服务。安康生态旅游资源汉英平行语料库建设的目的也是为旅游翻译者服务的,尤其是安康地区的宣传和旅游单位,为他们提供旅游翻译参考和借鉴,也为翻译研究和语言研究提供文本资源。导游词平行语料库的目的有三,其中两个涉及到翻译:一是研究导游词特点、翻译风格和内在成因;二是为导游词从业人员翻译导游词服务。其它几个旅游语料库的设计也是如此。所查找的所有旅游语料库的设计,最少有一个目的涉及到翻译,可见为翻译服务是目前已有旅游语料库构建的主要目的之一。

语料库设计的第二个主要目的是为教学服务。在贺州的旅游语料库构建的设计中,作者明确设计目的是为“翻译教学用途”之用。衡阳旅游汉英双语平行语料库的目的之一是促进衡阳高校的旅游英语和旅游英汉互译的教学所用。李德超的新型双语语料库构建三个目的虽然都是涉及到了翻译,但有两个目的地涉及到教学之用:翻译教学用途和为旅游从业者提供自学材料。

语料库设计的第三个目的在皖南旅游资源语料库中提及,就是提供双语文本资源为旅游相关领域,为研究旅游文本的文本功能和语体特征做出贡献。当然这一旅游文本并不只是汉语旅游文本,而是汉英或者英汉两种语言的旅游文本功能和语体特征。

由此可见,国内旅游语料库构建设计的三个主要目的是教学、翻译和文本特征研究。教学主要为培养翻译人才和旅游服务人才;翻译的目的主要是为了探讨旅游文本的翻译,包括公示语的翻译、导游词和景点等的旅游相关文本翻译;而文体特征的研究目的主要是为了探讨旅游文本的文本功能和语体特征。但是所有这三个目的,无论是教学、翻译亦或是文本特征的研究之用都是为了规范翻译、促进旅游业的发展、促进跨文化交流。

(二)旅游语料库的构建研究

语料库的构建包括语料的搜集、选取、数字化、标注、对齐和所应用的技术等方面的内容。

1.语料的搜集

国内所有旅游语料库的语料来源具有相同的特点。搜集的语料主要是构建该旅游语料库所涉及的地域。如赣南红色旅游双语平行语料库搜集的语料是江西赣南地区红色旅游资源的双语文本资源。李德超所搜集的语料范围主要为香港地区的旅游文本。贺州旅游语料库的语料地域范围主要是贺州地区。

语料的来源主要为景点介绍、观光册、旅游服务部门和博物馆的官方文本等。从国内现有的旅游语料库来看,语料的来源基本相似,正如李德超、王克非文中所述的自然景观、人文景观的介绍;观光小册子;酒店、旅馆、餐饮业及娱乐场所和设施的介绍;博物馆(包括展品)的说明和介绍。将来还可考虑收集与旅游业相关的其它行业的文本,如与交通、导游、票务、旅行社业务(包括旅游行程设计)、货币兑换、投诉处理、旅游安

全、旅游管理部门的宣传和公示语等相关的文本^[1]。

2. 语料的选取和数字化

初步搜集的语料并直接构建语料库,必须经过选取,甄别。旅游景点较多,宣传材料繁杂,难免出现低劣的译本,平行语料库构建组成员都经历了对语料的选取工作。主要选取有代表性的语料文本和适合的译语文本。且每个语料库都有自己选取语料的标准,如新型双语旅游语料库(李德超,王克非)的4个标准:1)依据语域标准抽样;2)错误文本不予收录;3)译文僵硬不予收录;4)请专家和翻译审阅译文以保证质量。

选取典型语料之后,将文本数字化,转化为TXT文本进行储存。在这个转化过程中,所采取的方式电子文本如网络资源只进行格式的修改和文字的校对。印刷品需要采取辅助设备,如高清扫描仪,扫描之后进行人工修正。照片的数字化依赖人工输入。

3. 语料的标注和对齐

语料标注采取人工和机器结合的方式进行。国内学者一般采用软件标注之后,再人工修正的方式。常用的英文标注软件有CLAWS,能够对词类、词性进行标注,称之为POS(part of speech tagging)标注。而汉语的标注主要采用中科院计算研究所研发的“汉语词汇分析系统”,称为ICTCLAS,进行切分标注。有些标注无法达到要求,国内学界有手动进行修辞方面的标注的例子。用RHE来表示总的修辞手法,然后用RHE加上修辞名称的开头字母表示修辞方法的标注。这有利于旅游语篇的对比分析和研究学习。也有其它建库者自创码集,对语料进行深度附码标注。其中有李德超等的新型双语语料库、黄永新等的导游词平行语料库等。

在对语料标注的基础上,对语料库的语料进行对齐。语料的对齐是指将源于语料和译语语料保存,使两个文本分别按照段与段和句与句的关系对齐。中英语言隶属不同的语系、语言特征相异之大,难以在语篇中进行词的对齐,所以旅游双语平行语料库主要是句对齐和段落对齐,而没有采取词的对齐。将汉英两种语料分别保存,使两种语料文本分别按照语段和语句的关系一一对应。但是在旅游文本的翻译中,存在着省译、扩译和摘译的现象,语句的一一对应无法全部实现。在进行对齐时,采取软件对齐和人工对齐结合的方式进行语料的对齐。主要应用ParaConc软件进行对齐,有的是先进行软件对齐,然后人工检验进行对齐。有的则正好相反,先进行人工对齐,然后再进行软件对齐、但是最后还是要应用软件进行对齐检验。

(三)旅游语料库构建技术

旅游语料库的构建主要涉及到语料的搜集、选取、数字化、标注、对齐。其构建技术就是围绕上述几个步骤进行。在语料的搜集、选取和数字化过程中所运用的技术相对容易操作。语料的搜集和选取采取人工方式进行。应用高清扫描仪数字化文本。语料的标注是指对语料成分的划分,便于检索和研究,称为POS标注,全称为part-of-speech tagging。用软件CLAWS标注系统进行机器标注。由于机器标注不能达到完全准确,仍需要人工检验进行修正。语料标注使用的CLAWS无法满足所有的标注要求,有的专家会自主创设码对语料进行标注,但是这些创设码必须能够与CLAWS兼容。(李德超、王克非,2010)。也有语料库应用TreeTagger进行英文标注。该软件是德国斯图加特大学计算语言学开发的自动词性赋码工具,准确率较高,可达到90%以上。(黄永新,张黎黎,2011)中文语料的标注主要采用中国科学院计算技术研究所研发的ICTCLAS软件,也有语料库用SVMTool进行中文的标注^[2]。

语料的对齐主要应用ParaConc和人工结合的形式进行。先用ParaConc粗略对齐,然后进行人工检验,校对。或者先人工进行段落对齐,然后用ParaConc对齐,再进行人工检验。山东旅游外宣资料小型语料库的对齐采取的是塔多思(TRADOS)中的WinAlign进行文本对齐。WinAlign通过翻译的文档创建翻译记忆数

据,将文档导入 TRADOS 翻译工作平台。然后 WinAlign 确定原文和译文,将其对齐排列。

二、国内旅游语料库的应用

就目前研究而言,旅游语料库主要运用于教学、旅游文本的翻译及文本比较方面。

(一)教学

旅游语料库应用于教学方面的研究,主要在旅游英语教学和翻译教学两个方面。利用旅游语料库能够提供真实、鲜活丰富的语料这一特点,学习者对语料库进行检索和考察,既能够把握微观层面,也能够把握宏观层面的知识,促进学习者对旅游文本的理解。同时通过对双语平行语料库对应语料的检索,学习者能够自己归纳翻译策略,教师稍加引导就可以。相比传统的教学,运用旅游语料库更加有利于学习者翻译策略的掌握,更有利于激发学生的学习主动性和学习兴趣,也能使课堂教和学更加有效。

李德超(2011)利用旅游语料库在翻译课堂中进行了 DDL 教学的尝试,把旅游语料库应用于翻译教学。改变了传统的教学流程,把传统课堂上教师讲授的内容改为把学生分组、下达任务,让学生运用旅游语料库考虑问题,从微观到宏观上把握翻译的文本,然后动手翻译^[3]。实践证明,学生由主动变为被动,积极性大为提高,并具有较多独立、原创的见解,实现了教学相长。

(二)旅游文本翻译研究

基于旅游语料库的旅游文本翻译研究目前已在景点名称、景区介绍、公示语、酒店宣传和文化负载词方面得到了体现。旅游平行语料库的语料涉及到双语语料的微观和宏观两个方面。微观方面可以从大量真实语料中发现并归纳出旅游文本翻译中的词法和句法的转化规律,如词类的转换和句式的转换等方面。汉语是属于动词性的语言,而英语属于名词性的语言,在旅游英语的平行语料中是否也存在词类的转换,是如何转换的,有什么样的特征。基于这种活的语料的归纳,更具有可见性和说服力。在宏观方面利用语料库研究旅游文本翻译的文体、风格和共性方面的翻译。旅游文本语料隶属于专门用途语言,具有自己独特的文体特征,对于旅游文本英汉/汉英的翻译,遵循什么规律,有什么特点,运用什么方法,通过旅游双语平行语料库的应用、检索、归纳、整理,可以快速掌握这种文体的翻译。杜渤、张静、张燕(2013)把旅游语料库应用于保定市旅游文本中文化负载词的翻译研究,得出了文化负载词的翻译特点^[4]。李德超、王克非(2009)基于汉英旅游平行语料库,归纳了酒店宣传材料在文本翻译中所呈现的特点,提出了实际可行的酒店宣传材料的翻译方法和策略,提高译文和目的语对齐文本中的文内连贯^[5]。

(三)旅游文本特征的研究

目前开发的旅游语料库除了应用于教学和旅游翻译研究之外,还用来研究旅游文本的特征。词汇方面,用来比较英汉旅游语篇的词汇色彩和特征,是正式的,还是口语化的亦或是其它。还通过词频统计特定场景中的常用词汇,还可以考察同一汉语词汇对应的译文。句子方面,用来研究旅游语篇中的句子结构和复杂性,或者研究句式选择的原因。或者考查语句在不同文化场合对读者产生的语用效果。语篇层面,用来比较英汉旅游语篇的衔接差异,归纳语步结构和题材结构的不同,以考察其社会认知根源异同。修辞方面,用来比较两种语篇常用的修辞方法,了解读者的反应,把握修辞方法在英文中的表达处理。侯晋荣(2011)研究了基于语料库的旅游文本语言特征及语篇分析^[6]。康宁(2012)基于类比语料库研究了中国网站英语旅游文本的语言,比较了英语国家旅游文本和中国网站英语旅游文本^[7]。

三、旅游语料库的未来研究方向

国内旅游语料库的研究已经取得了一定成绩,构建了一定数量的旅游语料库,并把旅游语料库进行了实际的应用,但是还有待进一步深化。目前国内旅游语料库库容较小、地域也有限,多为地域性或者区域性的,没有全国性的旅游语料库,也没有规模较大的红色旅游语料库,所以国内旅游语料库的扩容刻不容缓。另外旅游资源各地不同,把其它地区的旅游语料库应用于当地的导游培养、旅游翻译和旅游教学的话,不合时宜,效果不佳。由此一来,国内学界必须根据实际情况,构建各地的自己的旅游语料库,以此促进当地旅游的发展。

旅游语料库的应用范围也有待深化和拓宽。目前的旅游语料库主要应用于教学、旅游文本的翻译和旅游文本特征的研究。语料库能够提供真实的语料,且能够存储云端,供师生检索使用,非常适合开展翻转课堂的教学。把旅游语料库和翻转课堂结合起来,开发信息时代的新型教学模式,无疑会大大提升教学效果,也会对话料库的应用做进一步的升华和深化。语料库作为一种语言的研究方法范畴、其应用绝不止于此,可以开发基于语料库的机器翻译、实现人机对话、节省人工成本。

另外旅游语料库可以用于开发智能双语导游。旅游资源越来越丰富,游客也在日渐增多。旅游旺季,人工导游无法满足实际需要。英语导游数量较少,无法满足国外游客的需要,因此基于旅游双语平行语料库,各地开发智能双语导游也迫在眉睫。一来可以促进当地旅游的发展,也可以促进中国旅游的对外宣传,使中国文化远播海外。

参考文献:

- [1]李德超,王克非.新型旅游语料库的研制和应用[J].现代外语,2010(1):46-54.
- [2]黄永新,张黎黎.导游词平行语料库的构建及应用[J].吉首大学学报(社会科学版),2011(2):144-146.
- [3]李德超,王克非.基于双语旅游语料库的DDL翻译教学[J].外语电话教学,2011(1):20-26.
- [4]杜渤,张静,张燕.基于语料库的保定市旅游文本中文化负载词的翻译研究[J].劳动保障,2013(9):248.
- [5]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.
- [6]侯晋荣.基于语料库的旅游文本语言特征及语篇分析[J].菏泽学院学报,2011(12):124-128.
- [7]康宁.基于类比语料库的中国网站英语旅游文本语言分析[J].青岛科技大学学报(社会科学版),2012(4):105-109.

(责任编辑:张祥卉)

(下转第19页)

参考文献:

- [1]许荣裕.汉泰语指人词缀、类词缀的对比研究[D].厦门大学硕士论文,2008.
[2]赵艳平.汉语词缀研究[D].河北大学博士学位论文,2004.
[3]กำชัย ทองหล่อ. 2545. **หลักภาษาไทย**. กรุงเทพมหานคร : โรงพิมพ์อมรการพิมพ์ .

(责任编辑:张祥卉)

A Comparative Analysis of Affixes between Chinese and Thai

Teeraparp Predeepoch

(Xiamen University, Fujian 363105, P.R. China)

Abstract: the paper aims to study the similarities and differences of affixes between Chinese and Thai by analysis and comparative study, based on the 5 characteristics of Chinese affixes (affixes categories, syllable, word class transformation, semantics, and grammaticalization) included in the 6th edition of Modern Chinese Dictionary, 2012. It tries to help the Chinese and Thai learners to have a better lexicon acquisition. Moreover, it can also be a reference for teaching Chinese or Thai as a foreign language.

Key words: Chinese and Thai; affixes; comparison

~~~~~

(上接第 31 页)

## Literature Review on the Tourism Corpus in China

ZHU Zhi-juan, ZHU Li-yong

(Jiangxi Science and Technology Normal University, Nanchang 330013, P.R. China)

**Abstract:** after a careful analysis of the research in tourism corpus in China, we finds that there are two main research fields: the structure and the application of the corpus. The construction mainly includes the design of the corpus, data collection, digitalization, tagging, alignment, and so on. The application mainly focuses on teaching, translation and tourism textual research. As a whole, tourism corpus research has achieved a lot, but there still remain some problems, such as their small capacity, the limited region coverage, and the limited application fields.

**Key words:** tourism corpus; structure; application; teaching; translation