

## 旅游翻译类比语料库建设的可行性研究

## ——以江门旅游外宣翻译语料库为例

冯晓菲

(江门职业技术学院 外语系, 广东 江门 529000)

**摘 要:** 本文以江门旅游外宣翻译语料库为例, 讨论地方旅游翻译类比语料库建设的可行性。研究主要围绕语料库设计思路的三个步骤展开: 语料采集、语料整理与分词、语料词性还原与赋码, 并从文体特征和文化因素两个角度考究其应用价值, 以期自建语料库研究者提供参考。

**关键词:** 语料库; 旅游外宣翻译; 文体特征; 文化因素

**中图分类号:** H315

**文献标识码:** A

**文章编号:** 1008-9640 (2018) 06-0061-02

### 一、研究背景

我国综合国力的增强、“一带一路”倡议进入战略推进关键期、经济纵向深入推进, 以及越来越多的外国游客和客商到中国旅游、商务谈判等, 推动了经济发展的另一引擎——旅游业。而要借助旅游业发展契机打造中国名片、传播中国声音, 旅游对外宣传(简称外宣)资料的建设成为首个突破口, 肩负重任。因此, 在讲好中国故事、推动本土旅游文化走向世界的过程中, 旅游外宣翻译隐藏的跨文化传播能力不容忽略。然而根据笔者搜集的著名侨乡江门市4A涉外景点406份对外旅游宣传资料发现, 存在大量诸如本土化词汇缺失、文化错位、译者欠缺跨文化意识等问题, 大大削弱了外宣资料的宣传效果。

语料库作为新兴网络工具, 通过对大量真实语言文本的整合, 达到全面客观分析文本规律的目的。目前, 基于语料库进行翻译资料建设的先例为数不少, 如青岛科技大学的英语旅游文本类比语料库及无锡发展研究中心的英语旅游文本类比语料库, 这些语料库涵盖面广, 但缺乏地方文化针对性。鉴于此, 笔者以江门五邑地区4A景点外宣英语翻译文本为基础, 研制了一个收录内容更具针对性的江门旅游外宣翻译语料库, 以期更全面地揭示地方旅游外宣翻译文本的特性。

### 二、语料库的设计

#### (一) 语料采集

当今研究所采用的语料库类型主要包括“多语料库”和“单语料库”, 前者包括平行语料库和类比语料库, 后者则包括同一语言的原创文本和

翻译文本。江门旅游外宣翻译语料库是类比语料库, 属于多语语料库的一种, 此类语料库中的双语或多语语料虽然并不需要翻译对等, 但需要在语域、年代等方面保持一致。根据主题内容, 将江门旅游外宣翻译语料库建成温泉、海岛及碉楼三个旅游英语语料子库, 每个主题均由英译中文语料库和旅游外宣资料的英文语料库两个类比库组成。英译中文生文来自江门市五邑地区的旅游英译文本, 而外国旅游宣传生文则选自英国、澳大利亚、日本及美国的官方旅游文本。

语料库生文本的收录范围包括网页形式的电子文本及书面形式的文本(包括宣传手册、景区内的标示语及简介等)。两个类比库的取样框架趋于一致, 取样为完整篇章样本, 单个样本大小为200~800个英文单词。碉楼在国外暂时没有同类目的旅游主题, 而根据碉楼的建筑风格接近欧式城堡风格, 所以在做类比库时采用了欧洲一些重要城堡的外宣资料文本作为对比。目前, 温泉、海岛及碉楼三个主题库的规模分别达到11万词、9万词及5万词。

#### (二) 语料整理与分词

生文本收集后, 整理为电子文本, 并利用PowerGrep软件对生文本进行去噪处理, 删除不必要的信息、符号等, 得出清洁文本。同时, 利用分词程序Tokenizer对清洁文本进行英文文本的分词处理, 也就是将一连串的字符转换成互相分离、容易识别的形符。

#### (三) 语料词性还原与赋码

在文本中, 一些实词存在不少的屈折形式, 如

收稿日期: 2018-06-02

基金项目: 2016年江门市科技局基础与理论科学研究类科技计划项目。

作者简介: 冯晓菲(1980—), 女, 广东江门人, 江门职业技术学院讲师。研究方向: 应用语言学。

build 就有 build、builds、building 和 built 共 4 种不同的屈折变化形式,检测主题词时,如果把这 4 种形式作为 4 个词来看待,明显有所偏颇。因此,需要把同一个词的不同屈折形式归并到一起,视为同一个词。笔者使用 AntConc3.2.1 把这些屈折形式都用原形进行替代。

另外,需对语料进行词性赋码,意指根据文本中的上下文信息,自动标注文本中所有词的词性,以方便检索和语言处理。本研究使用德国斯图加特大学计算语言学研究所设计的 TreeTagger 进行自动赋码。

三、语料库的应用测试

(一) 文体特征考察

语料库可用于考察文体特征。笔者通过使用 AntConc3.2.1 对英译中文语料库和英文语料库的主动和被动语态进行对比,利用正则表达式 “\ S+\_VB \ w+ \ s ( \ S+\_ [RX] \ w+ \ s ) \* \ S+\_V \ wN” 在两个语料库中检索出被动结构。在以上表达式中,“\ S+\_VB \ w+” 为 BE 动词和对应的赋码,而 “\ s” 表示空格,“( \ S+\_ [RX] \ w+ \ s ) \*” 则表示 0 或者多个副词或否定词 (not, n’ t),而 “\ S+\_V \ wN” 则代表动词的过去分词。

经统计 (如表 1),英译中文语料库中的被动句式占句子总数的 34.2%,英文语料库中的被动句式只占 19.2%,远低于英译中文语料库。这与通用文本刚好相反,正常情况下,英语文本体现了西方人注重逻辑性的特征,更多从受事者的角度客观陈述,多采用被动句式;中国人看重悟性思维,倾向以施事者角度描述事情,多采用主动句式。然而,英文语料库中的外宣文本更多地以游客为中心表述,加强与读者的交流,以诱导性话语为主,如 “you”,缩短与游客距离;而英译中文语料库则以旅游资源的社会发展和历史渊源为中心进行客观描述,受众角度的描述较少。

表 1 被动语态使用的统计

语料库	句子总数	被动句式	比例
英译中文语料库	5 782	1 733	34.2%
外国语料库	5 531	1 027	19.2%

(二) 文化因素考察

利用语料库可以进行文本的关键词搜索。文本的关键词是指通过语料库中两个单词列表的比较得到的出现频率较高的词。通过分析关键词,能为语篇层面词语的分布及主题的关系提供实证依据,从而了解文本的视角。笔者基于海岛、温泉及碉楼三个主题的类比语料库,检索出前五名的关键词。

1. 海岛

海岛外宣文本中,本土关键词以海滩的餐饮配

套为主,包括海鲜 (Seafood) 和饭馆 (Restaurants),这也与我国 “民以食为天” 的文化传统不谋而合。而外国外宣重点在关于安静 (Quiet) 和隐私 (Privacy) 等词上,这与 Hofstede (2005) 提出的欧美文化崇尚个人主义密切相关,而个人主义的核心价值正是个体独立和自由。

2. 温泉

据统计,本地侧重于温泉的疗养功效宣传,泉 (Spring)、矿物 (Mineral) 及温泉池 (Pool) 被频繁提及。日本温泉的宣传文本高频使用酒店 (Hotels)、房间 (Rooms)、和服 (Kimonos) 及饮食 (Cuisine),竭力让旅游者在独特的温泉文化体验中流连忘返,以满足游客精神层次上的需求。

3. 碉楼

本土碉楼的宣传重点在于华侨 (Overseas Chinese) 文化及本地传统 (Tradition) 的介绍。外国的城堡以世纪 (century) 和朝代 (Dynasty) 为宣传的关键词,如欧洲外宣文本重点对城堡的家族史进行专题研究,以折射出当时的社会文化。

四、结论

通过语料库检索研究,建议在今后的相关外宣翻译工作中,首先应提炼外地游客的关注点,侧重介绍旅游资源中游客关注的部分,如对海岛介绍时可强化静谧及阳光充沛的特点、温泉宣传可适当引入具有当地特色的文化元素、碉楼外宣时可以更深入地介绍碉楼背后的家庭文化。另外,本地文本表述方面应多采用主动句式,多采用第一和第二人称,以拉近与外国游客的距离,增强宣传文本的亲合力。

参考文献

[1] 崔建周. 基于语料库的旅游景区英文网特文本文体特征分析 [J]. 海外英语, 2012, (12): 234-236.

[2] 胡志毅. 国外遗产旅游 “内生矛盾论” 研究述评 [J]. 旅游学刊, 2011, (9): 90-96.

[3] 侯晋荣, 秦洪武. 利用关键词分析旅游文本的文体特征: 一项基于语料库的实证研究 [J]. 外国语文研究, 2012, (8): 73-84.

[4] 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.

[5] 李文中. 基于中国英语学习者语料库的主题词研究 [J]. 现代外语, 2003, (3): 284-293.

[6] 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, 2012, (4): 105-109.

[7] Geert Hofstede. Culture’s Consequences: International Differences in Work-Related Values [M]. London: Sage Publications, 2005.

(责任编辑: 李汴红)