

基于类比语料库的英语旅游文本文体特征考察^①

李德超 香港理工大学

唐芳 广东外语外贸大学

摘要: 本文基于自建英语旅游文本类比语料库, 比较英语旅游翻译文本与英语旅游原生文本在类符型符比、词汇密度、平均句长、平均词长、高频词等参数上的区别。研究发现, 英语旅游翻译文本较之英语旅游原生文本具有以下特点: (1) 符合“简化”趋势, 主要表现为标准化类符型符比较低, 动词变化程度较低, 原形形容词使用重复率较高, 高频原形副词使用较频繁; (2) 符合“显化”趋势, 主要依据为代词密度偏低; (3) 时态的运用更为丰富, 主要体现在完成时和将来时所占的比重更高; (4) 文体更为正式、客观, 主要表征为动词名词化较多、被动语态的使用更为频繁和极低的缩略式使用频率。

主题词: 类比语料库; 旅游文本; 文体特征

中图分类号: H0-05

文献标识码: A

文章编号: 1672-9382(2015)04-0088-09

1 引言

类比语料库多用于揭示翻译文本相较于原生文本的特性。随着语料库翻译研究的不断发展, 除文学文本之外, 越来越多针对某种特定文类而建立的类比语料库应运而生。目前, 国内外已有数个针对旅游文本而创建的语料库:

国外有芬兰萨翁林纳翻译研究学院研制的英语旅游文本语料库(共67万词)(Wilkinson, 2005), 英国埃塞克斯大学研制的英语旅游文本语料库(共约3.8万词)和日本大学研制的“京都旅游语料库”(共约4万词)(参见李德超、王克非, 2010); 国内有青岛科技大学康宁及其团队研制的英语旅游文本类比语料库(共约188万词)(康宁, 2012), 无锡旅游发展研究中心崔建周研制的英语旅游文本类比语料库(共约20万词)(崔建周, 2012)和山东菏泽学院侯晋荣及其团队研制的英语旅游文本类比语料库(共约12万词)(侯晋荣、秦洪武, 2012)。以上语料库中, 除康宁团队的库容达到百万次, 其余规模都较小。由于康宁团队建库的主要目的是发现中国网站旅游文本在英语写作及英译方面的问题及不足(康宁, 2012: 106), 因此收录的翻译文本主要为介绍中国大陆地区景点的旅游网站, 且翻译文本

子库仅有约53万词, 与原生文本子库的135万词相比, 两库库容差距过大。鉴于此, 我们研制了一个库容较大、收录范围更加广泛的英语旅游文本语料库, 以期能更全面揭示源自中文的旅游英译文本的特性。

2 旅游文本语料库简介

本研究依据的新型双语旅游语料库由香港理工大学中文及双语学系研究团队研制, 也是香港以至全国第一个较大容量的英汉/汉英旅游语料库。此新型双语旅游语料库包括两个子语料库: (1) 类比语料库; (2) 平行语料库。本研究选用的是类比语料库中的文本。

该类比语料库库容超过125万词, 由原生文本和翻译文本两个子库构成。其中原生英语旅游文本选自英国、澳大利亚、爱尔兰、加拿大、美国及马耳他这六个以英语为官方语言的国家旅游文本, 共628 669词; 而其中的翻译英语旅游文本取自中国(包括大陆、香港、台湾)和新加坡的旅游文本, 共628 642词。这些文本的收录范围包括网页形式的电子文本和书面形式的文字文本(包含书籍、宣传册、旅游区区内张贴的介绍等)。所有文本均以完整篇幅收录, 用UTF-8编码形式逐个存档于txt文

作者简介: 李德超, 香港理工大学中文及双语学系副教授。研究方向: 翻译研究、口译研究。Email: dechao.li@polyu.edu.hk。唐芳, 广东外语外贸大学高级翻译学院讲师, 翻译研究中心研究员。研究方向: 翻译研究、口译研究。Email: candy.fangtang@hotmail.com。

本中,并使用TreeTagger进行词性标注。为实现不同的检索目的,所有文本均以未标注版本和已标注版本两种方式保存。

3 数据统计与分析

3.1 基本数据

基于语料库的翻译研究常用平均词长与标准化类符/型符比来考察文本特征。我们运用WordSmith5.0的wordlist工具统计了两种语料的上述参数,借以反映两者间在用词难度和词汇变化幅度等方面是否存在差异。

首先,平均词长是指特定文本中所有单词的字母总数与型符总数的比值,词长标准差表明单词长度在平均词长左右浮动的程度。以往的语料库研究显示,平均词长是反映文本文体正式程度和用词难度的重要指标之一。

表1 “政府工作报告英译”与“社会生活类同传英译”平均词长对比

	平均词长	词长标准差
政府工作报告英译	5.49	3.04
社会生活类同传英译	4.34	2.37

例如,表1比较了两种不同文体的文本在词长方面的差异,其中文体较为正式的《政府工作报告》英译文本的平均词长为5.49,标准差为3.04(陈建生等,2010:40);而文体较为口语化的同传英译语料的平均词长为4.34,标准差为2.37(李德超、王克非,2012:412),这组数据表明文体越正式、使用词汇难度越高,文本的平均词长就越长。

表2 “旅游英语原生文本”与“旅游英语翻译文本”平均词长对比

	平均词长	词长标准差
旅游英语原生文本	4.86	2.50
旅游英语翻译文本	4.84	2.54

如表2所示,本研究涉及旅游英语文本中,原生文本的平均词长为4.86,标准差为2.50;翻译文本的平均词长为4.84,标准差为2.54。从数据上看,旅游文本的文体正式程度介于《政府工作报告》与“社会生活同传语料”之间,使用的词汇难度低于《政府工作报告》,略高于“社会生活同传语料”。然而,两组旅游文本在平均词长上的差异较小,说明在旅游英语

范围内,翻译文本与原生文本用词难度相近。

一般认为,类符/型符比与文本的词汇多样性成正比关系(Baker,2000)。当各语料库的容量不一时,只有将该比率进行标准化处理后——即对每一千词的类型符/型符比进行计算——得出的平均值,即类符/型符比(STTR),才具有参考价值(Scott et al.,2006;McEnery et al.,2006)。我们对此的对比结果显示(见表3),翻译文本的标准化类符/型符比低于原生文本,表明原生文本中词汇的变化程度更高,使用的词汇更加丰富。这一结果符合翻译文本的“简化”趋势。旅游翻译文本的“简化”究竟有什么特点?在下节中,我们将按词性逐个展开分析。

表3 旅游英语原生文本与翻译文本标准化类符/型符比对比

	标准化类符/型符比
旅游英语原生文本	47.03
旅游英语翻译文本	39.25

3.2 词汇特征对比

词汇密度是用于衡量文本难度的一个常用指标。在本研究中,我们借助TreeTagger对旅游英语原生文本与翻译文本作了词性标注,然后通过AntConc 3.3.5工具中分别检索“_N”“_V”“_MD”“_J”“_RB”“_CD”“_PP”(详见TreeTagger词性赋码集)对两种语料中的名词、动词、形容词、副词、数词和代词的词汇密度进行了统计。

如表4所示,总体上看,旅游英语翻译文本的词汇密度略高于旅游英语原生文本,二者相差仅1.01%,即在相同长度的旅游文本中,翻译文本中使用的实词略高于原生文本。换言之,相同长度的旅游文本中,翻译文本使用了更少的功能词。这一数据与Baker(1993)提出的翻译共性中的“简化”趋势显然背道而驰。原因何在?具体来看,旅游英语翻译文本的词汇密度高主要是集中在名词密度和数词密度上,而动词密度、形容词密度、副词密度和代词密度均略低于旅游英语原生文本。以下我们将分词性来对此展开进一步分析。

数词密度差异主要源于中文原文好用数字,与翻译行为本身关系不大,因此本文将不在此作具体分析。排除数词后,差异最大的是名词密度,翻译文本的名词密度比原生文本高出0.75%。因翻译文本的动词密度小于原生文

表 4 旅游英语原生文本与翻译文本的词汇密度统计

	旅游英语原生文本	旅游英语翻译文本	差异百分比
名词	38.66%	39.41%	-0.75%
动词	11.61%	11.21%	0.40%
形容词	10.09%	9.65%	0.44%
副词	2.94%	2.88%	0.06%
数词	1.97%	3.62%	-1.65%
代词	2.49%	2.01%	0.48%
总词汇密度	67.77%	68.78%	-1.01%

本, 这可能与英语中的动词名词化现象有关。这类名词是由动词派生而来的, 它们与其他非动词派生的名词在句法功能上相同, 但在意义上常用于表示该动词含有的动作、行为、变化、状态、品质、情感等概念(王武兴, 2003)。英语动词添加后缀构成名词的方式非常多。本文在此只对几种常用的名词化后缀, 包括-tion(s)、-sion(s)、-ment(s)、-ence(s)、-ance(s)、-ing(s)、-al(s)、-ure(s)等, 借助AntConc3.3.5用“-tion(s)_N”“-sion(s)_N”“-ment(s)_N”“-ence(s)_N”“-ance(s)_N”“-ing(s)_N”“-al(s)_N”和“-ure(s)_N”进行了检索, 并通过人工逐个排除了干扰词, 如population(s)、mansion(s)、everything之类。表5列出了这八类最常用的动词名词化后缀在旅游英语原生文本与翻译文本中的使用情况。

表5数据表明, 总体上看, 翻译文本中动词名词化的出现频率更高: 翻译文本中动词名词化的使用频率为21 992次, 而原生文本中仅为15735次。此外, 八种常用的名词化后缀在

旅游英语翻译文本中的使用频率均高于旅游英语原生文本。两个语料库中出现频率较高的动词名词化的例子包括accommodation、celebration、impression、admission、development、entertainment、experience、residence、entrance、performance、shopping、parking, arrival、departure等。Halliday曾经提出, 通过名词化现象, 作者可以减少对语境的依赖, 在呈现现实世界的经验和现象时, 可以把它当做物体来看待, 从而让文章显得更客观、更正式(1985: 91)。我们认为旅游英语翻译文本更频繁地使用了动词的名词化, 是因为在此过程中动词所涉及的施事与受事可以被省略, 这就为文本铺垫了一种独立于施事者与受事者的客观氛围。与此同时, 原有动作的动态形象也转化为一种抽象的静态概念, 呈现出一种客观、正式的文体风格。例如:

例(1)

原生文本: If you want to **experience** a traditional London restaurant, try Rules on Maiden

表 5 旅游英语原生文本与翻译文本的动词名词化使用对比

	原生文本		翻译文本		差异	
	次数	百分比	次数	百分比	次数	百分比
-tion(s)	5 451	0.87%	9 529	1.52%	-4078	-0.65%
-sion(s)	393	0.06%	494	0.08%	-101	-0.02%
-ment(s)	1 060	0.17%	1 105	0.18%	-45	-0.01%
-ence(s)	934	0.15%	993	0.16%	-59	-0.01%
-ance(s)	788	0.13%	1 287	0.20%	-499	-0.08%
-ing(s)	6 705	1.07%	7 593	1.21%	-888	-0.14%
-al(s)	122	0.02%	335	0.05%	-213	-0.03%
-ure(s)	282	0.04%	656	0.10%	-374	-0.06%
总和	15 735	2.50%	21 992	3.50%	-6257	-1.00%

Lane.

翻译文本: When it comes to a serious shopping experience, Hong Kong has everything in abundance.

例(2)

原生文本: Much time and effort has been spent in recent years to ensure that travelling with a disability won't stop you enjoying all that Australia has to offer.

翻译文本: The biodiversity of Zharu Valley is a rare treasure in China that presents exciting opportunities for enjoyment and exploration.

在以上两例中,原生文本的作者将“experience”和“enjoy”用作动词,列明了动作的施事“you”和受事“a traditional London restaurant”以及“all that Australia has to offer”,通过“experience”和“enjoy”这两个动作串联起主体与客体之间的互动,增强了读者的参与感。而翻译文本中,作者将“experience”和“enjoy”用作名词,整句中并未点明动作的参与者,仿佛仅仅是在陈述一件与读者并无关系的客观事实,不能有效地感染和调动读者。根据Reiss的翻译类型学,文本的主要功能科分为三种:信息、表情以及感染。旅游文本的功能更倾向于信息和感染,而“感染性文本”主要是为“直接引发所期待的反应”(Reiss, 1989: 109)。因此,按照目的论,翻译旅游文本的一大重点是要引发读者萌生到所描述之处旅游的冲动。若译者将这两句话译为“If you want to experience

a serious shopping, visit Hong Kong”和“The biodiversity of Zharu Valley is a rare treasure in China that presents exciting opportunities for you to enjoy and to explore”,将读者置于主动位置,相信能更好地激发读者的参与热情,实现旅游文本“感染”读者的功能。

在动词密度方面,总体上看,翻译文本中的动词密度略低于原生文本,但差异较小,仅0.4%。以上发现的旅游翻译文本动词名词化现象更频繁的趋势部分解答了这个现象。除此之外,旅游文本的译文中动词使用是否还有其他特点?众所周知,“be”“have”“take”“come”“can”和“become”是英语中的超高频动词,而旅游文本由于涉及大量地理位置因素,“locate”和“situate”也应该为常用动词。我们在AntConc3.3.5中检索了两种文本中这八类动词及其变体(包括它们的原形、第三人称单数、进行式、过去式及过去分词的变体形式)的使用频率,发现翻译文本中这些词出现的频率远远超过原生文本。在排除这八类动词后,翻译文本的动词数量低于原生文本(相差1.75%),证明翻译文本中使用的动词数少于原生文本。基于以上分析,我们不难得出两类文本中动词密度虽总体差异不大,但翻译文本动词的使用仍呈现“简化”趋势,其特点表现为:(1)翻译文本中动词名词化现象更频繁;(2)翻译文本中高频动词的使用频率更高(表6)。

从形容词和副词密度的角度来看,翻译文本在这两方面的密度均低于原生文本,其中形容词密度低0.44%,副词密度低0.06%(见表4),这说明翻译文本中形容词和副词的使用频率都较低。通过在AntConc 3.3.5中检索“_

表6 旅游英语原生文本与翻译文本中高频动词使用情况对比

	原生文本		翻译文本		差值	
	词数	百分比	词数	百分比	词数	百分比
动词总数	68 639	10.92%	66 313	10.55%	2 326	0.37%
be及其变体	14 054	2.24%	20 054	3.19%	-6 000	-0.95%
have及其变体	2 314	0.37%	3 548	0.56%	-1 234	-0.20%
take及其变体	1 383	0.22%	1 814	0.29%	-431	-0.07%
come及其变体	483	0.08%	467	0.07%	16	0.01%
can及其变体	2 113	0.34%	2 131	0.34%	-18	0.00%
become及其变体	239	0.04%	372	0.06%	-133	-0.02%
locate及其变体	530	0.08%	1256	0.20%	-726	-0.12%
situate及其变体	64	0.01%	203	0.03%	-139	-0.02%
高频动词总和	21 180	3.37%	29 845	4.75%	-8 665	-1.38%

JJS”“_JJR”“_RBS”“_RBR”统计出形容词和副词的比较级和最高级形式的使用情况(见表7):

表8的数据显示翻译文本中原形副词出现的频率更高,这似乎又违背了“简化”的趋势。用AntConc 3.3.5对两组文本中的高频副词(出现频率均超过100次的副词)的统计结果显示翻译文本中高频副词的使用频率总体高于原生文本,且在15个分项中翻译文本也基本呈现较高的频率。这反映出翻译文本中的原形副词虽然总量高于原生文本,但其使用的副词重复率更高,因此翻译文本中副词的变化程度仍为相对较低。

此外,表7表明翻译文本中形容词和副词的比较级和最高级形式均低于原生文本;而翻译文本中的形容词原形形式远低于原生文本,

差值达到4 574词,这说明翻译文本中形容词和副词的“简化趋势”不仅表现在两者的原形上,翻译文本也较少使用形容词的比较级和最高级形式。

通过观察wordsmith5.0生成的词频表,我们可以得出两组文本中词频高于300次的形容词,共有66个词。比较这些形容词的出现频次,表9中列出了两组文本中上述形容词出现频次差值大于200的词。符合条件的词共有16个,较之66个高频词而言,说明旅游翻译文本所使用的高频形容词有较高的重合率。这其中两个文本都使用超过500次且频次差值小于200的词包括:“many”“national”“international”“first”和“beautiful”。此外,表9还显

表7 旅游英语原生文本与翻译文本形容词及副词使用对比

	原生文本		翻译文本		差值	
	词数	百分比	词数	百分比	词数	百分比
形容词总数	63 454	10.09%	60 688	9.65%	2 766	0.44%
形容词原形	60 060	9.55%	55 486	8.83%	4 574	0.73%
形容词比较级	1 845	0.29%	1 265	0.20%	580	0.09%
形容词最高级	3 016	0.48%	2 665	0.42%	351	0.06%
副词总数	18 488	2.94%	18 088	2.88%	400	0.06%
副词原形	16 189	2.58%	16 372	2.60%	-183	-0.03%
副词比较级	669	0.11%	475	0.08%	194	0.03%
副词最高级	950	0.15%	737	0.12%	213	0.03%

表8 旅游英语原生文本与翻译文本高频原形副词使用对比

	原生文本		翻译文本		差值	
	词数	百分比	词数	百分比	词数	百分比
also	1 141	0.18%	1 401	0.22%	-260	-0.04%
here	886	0.14%	695	0.11%	191	0.03%
about	783	0.12%	1 029	0.16%	-246	-0.04%
just	737	0.12%	201	0.03%	536	0.09%
not (& n't)	540	0.09%	832	0.13%	-292	-0.05%
only	404	0.06%	514	0.08%	-110	-0.02%
well	381	0.06%	514	0.08%	-133	-0.02%
even	301	0.05%	186	0.03%	115	0.02%
still	251	0.04%	204	0.03%	47	0.01%
now	251	0.04%	304	0.05%	-53	-0.01%
once	244	0.04%	147	0.02%	97	0.02%
very	198	0.03%	323	0.05%	-125	-0.02%
often	146	0.02%	104	0.02%	42	0.01%
around	129	0.02%	111	0.02%	18	0.00%
there	104	0.02%	300	0.05%	-196	-0.03%
总和	6 496	1.03%	6 865	1.09%	-369	-0.06%

表9 旅游英语原生文本与翻译文本高频原形形容词使用对比

	原生文本	翻译文本	差值
local	1 062	587	475
outdoor	323	52	271
traditional	246	455	-209
cold	109	323	-214
other	557	790	-233
every	398	649	-251
various	52	308	-256
related	18	275	-257
famous	468	727	-259
geographic	18	360	-342
such	406	755	-349
average	140	570	-430
annual	157	604	-447
cultural	316	804	-488
ancient	279	924	-645
scenic	399	1 389	-990
总和	4 948	9 572	-4624

示这16个词中的大部分(除“local”“outdoor”和“traditional”外)都更频繁地出现在翻译文本中,这证明了翻译文本对高频形容词的使用量更大,即重复率更高,这也从另一个侧面反映了翻译文本词汇变化度较小的现象。

在代词的使用方面,如表10所示,翻译文本中代词出现的频次明显低于原生文本,这说明在谈及动作的主客体时,翻译文本更倾向于使用名词而非代词,即翻译文本倾向于将指代关系明晰化。

例如,在例(3)中,原生文本用代词“**They**”替代了前句中出现的“the Aborigines”,而在翻译文本中,虽然第二句仍在谈论北京,但却未用代词,而是沿用了“Beijing”这一名词。虽然用“Beijing”使得指代更清晰,但若将第二句改为“**It features with fairly distinctive seasons, namely...**”或“**its seasons...**”,不仅不会出现指代不明(因旅游文本是书面文本,读者可以迅速从上句中找到所指),也是一种增强文内衔接

的有效方式。

例(3)

原生文本: About 40, 000 years ago, the Aborigines were the first to settle. **They** lived as hunters and gatherers for this entire time, living with a close link to nature...

翻译文本: Located in the semi-moist area of warm temperate zone, Beijing has a sub-humid continental monsoon climate with an annual precipitation of 571.8 millimeters. Seasons in **Beijing** are fairly distinctive, namely, a windy and sandy spring, a hot and rainy summer, a sunny and dry autumn and a chill and fierce winter.

3.3 句法特征对比

3.3.1 时态特征对比

利用AntConc 3.3.5,通过正则表达式搜索各种时态(包括一般现在时、一般过去式、

表10 旅游英语原生文本与翻译文本代词使用对比

代词	原生文本		翻译文本		差值	
数量	15 684	2.49%	12 667	2.01%	3017	0.48%

进行时、将来时、完成时)的使用频次,结果如表11所示:

由图1、图2可以看出,旅游文本中一般现在时是最常用的时态,其次是一般过去式。对比两种文本,可以发现翻译文本对时态的运用更丰富。这具体表现在更多地使用了将来时和完成时。其中差异最大的在于完成时的使用上,而由于原生文本子库中出现的过去完成时有82次,翻译子库中有151次,可见完成时的差异主要在于现在完成时的使用上。现在完成时和一般过去时一样,都是对过去发生的动作的描述。不同的是,现在完成时侧重表现动作的持续(或重复)以及对现状产生影响,而一般过去时只体现动作在过去某一时间点曾发生,不涉及现状。翻译文本中一般过去时较少,现在完成时较多,表明在对过去动作的描述上,旅游翻译文本更关注动作的持续性及对现状的影响,如下面的例子:

例(4)

原生文本: Four blocks east, the gem-like Rodin Museum **was** a gift to the city from Philadelphia-born film mogul Jules Mastbaum.

翻译文本: For over 50 years, Mandai Orchid Garden **has been** a beautiful sanctuary where

a diverse array of spectacular hybrids bloom on a gentle hillside.

例(5)

原生文本: Melbourne **became** the main financial centre in all Australia.

翻译文本: Denyin Academy **has become** a popular tourist destination because many students and exam takers come here to pray for luck.

在例(4)中,与翻译文本中选取的例句一样,原生文本中陈述的情况仍适用于现在(罗丹博物馆至今仍为电影大亨Jules Matbaum送给他出生城市费城的礼物),但原文并没有使用现在完成时,不强调现状。而翻译文本采用现在完成时,突显了该花园从过往到如今一直是迷人的状态。同样,例(5)中,墨尔本至今仍是澳大利亚主要的金融中心,但原文中只用一般过去时,不涉及墨尔本现在的金融地位。但在翻译文本例句中用现在完成时,强调登瀛书院不仅在过去乃至现在一直是一个受欢迎的旅游目的地。

3.3.2 语态特征对比

在汉英的对比研究中,一般认为英语中的被动句使用量大于汉语。但旅游类翻译文本及原

表 11 旅游英语原生文本与翻译文本时态对比

	原生文本	翻译文本	差值
一般现在时	18 558	15 855	2 703
一般过去时	3 808	3 330	478
进行时	501	415	86
将来时	983	1 164	-181
完成时	860	1 390	-530

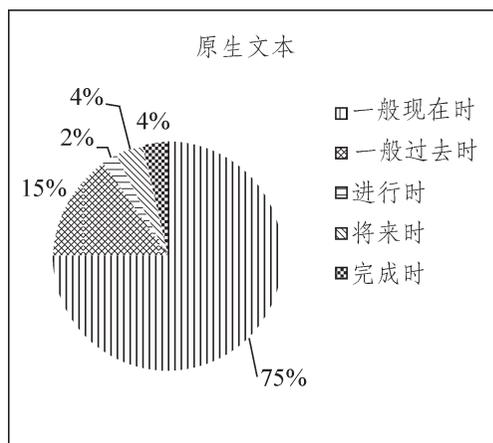


图 1 旅游英语原生文本时态分布图

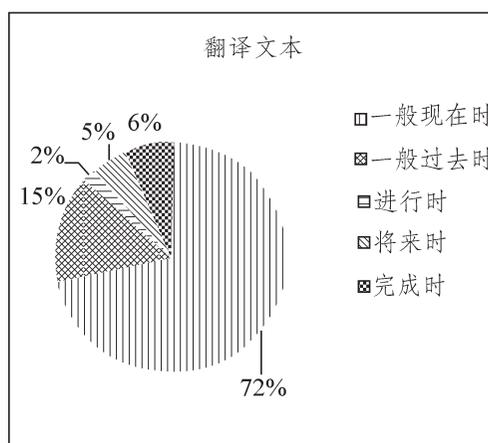


图 2 旅游英语翻译文本时态分布

生文本被动句的使用情况却鲜有探讨。我们用正则表达式 (\S+_VB\w*\s (\S+_ [RX]\w+\s) *\S+_VwN\s) 在AntConc 3.3.5中搜索两种文本中被动语态的使用情况, 结果如表12所示, 翻译文本被动语态的出现频率大大高于原生文本, 在每万词中的出现频次差值达到467次。这

说明在旅游文本中译英的过程中, 大量的主动句被译成了被动形式。由于被动语态更多出现在正式语体中 (Biber et al., 2000), 且可用于省略动作的施事与受事, 这与上节证明的翻译文本中动词名词化现象更频繁相互佐证, 再次说明旅游的翻译文本较原生文本更为客观和正式。

表 12 旅游英语原生文本与翻译文本被动语态对比

	被动语态	子库总词数	被动式使用频次(每万词)
原生文本	3 570	628 669	568
翻译文本	6 483	628 641	1 031

例 (6)

原生文本: You probably **know** about the read monolith in Australia's Red Centre.

翻译文本: Hangzhou **is known** for its heavenly appeal.

在例 (6) 中, 两个句子的中心意思均是表达某个景点因某个特点而广为人知。原生文本用了主动语态来表达, 并用第二人称“you”做主语, 缩短了读者与作者之间的距离; 翻译文本运用被动语态, 未提及动作的发出者, 表述方式也略显刻板, 给读者呈现的是客观和不带任何感情色彩的信息。由于“旅游文宣多属于感染文本” (李德超、王克非, 2009: 54), 译者似乎可以适当为突显景点的吸引力作出一些努力。在本例中, 若将翻译文本改为“You probably know

about the heavenly appeal of the Hangzhou City”, 似乎更能引起读者的共鸣和想象。

3.4 缩略式使用对比

除此之外, 另一个常用来分析文体正式程度的参数是“缩略形式的使用数量”。我们利用AntConc 3.3.5检索了几种常用的缩略表达式 (包括“s”“ll”“t”“re”“ve”) 在两个语料库中的出现频率。

结果如表13所示, 翻译文本中的缩略式出现频率不仅总体上远远低于原生文本, 相差高达2211次, 而且这种相对较低的频率适用于每一类缩略式。由于正式文本中一般不允许出现缩略式, 因此缩略式出现频次越高, 文本的正式程度便越低。基于此, 表13的数据也再次证明了翻译文本较原生文本而言正式程度更高。

表 13 旅游英语原生文本与翻译文本缩略式对比

	's	'll	't	're	've	总和
原生文本	927	780	598	362	108	2 775
翻译文本	304	58	140	40	22	564

4 结论

通过以上基于语料库的分析与对比, 我们发现英语旅游翻译文本较之英语旅游原生文本具有以下特点:

(1) 符合翻译的“简化”趋势, 主要表现为标准化类符型符比较低, 动词变化程度较低 (更集中地使用了“take”, “come”等高频动词), 原形形容词使用重复率较高, 高频原形副词使用较频繁。在康宁 (2012) 的研究中, 中国旅游网站的英语文本在形容词和副词的使用上远远多于英国国家旅游网站。康宁将较高

的形容词和副词词频归因于受汉语原文“重整合, 重中和”而英语“重理性, 重写实, 忌华而不实”的影响 (ibid. 108)。但康宁的研究只考察了类比语料库, 并未回归源文本, 所以其解释只能被视为一种合理猜想。本研究归纳的第一个特点并不只是推断, 而是基于对两个子库中高频副词和形容词频次的比较, 因而更具说服力。

(2) 符合翻译的“显化”趋势, 主要依据为代词密度偏低。这与崔建周 (2012) 考察的江苏省旅游景区英文网页特点一致, 说明受汉语“形合”特点的影响, 旅游英语翻译文本对语篇衔接方面的考虑仍显不足。

(3) 时态的运用更为丰富,主要体现在完成时和将来时使用频率更高。依照汉语无明显时态标记而英语时态更加多样化的语言差异,我们原本预期翻译英语中的时态运用应该会更加单调,然而数据显示的结果与这一预期恰好相悖。也许正是由于汉语原文对时态的标注较为模糊,而英语中对动作时间概念的表述又倾向于明晰化处理,才使得翻译文本呈现出更为细化的时态选择。

(4) 文体更为正式、客观,主要表征为动词名词化较多、被动语态及缩略式的使用更为频繁。这一特点与侯晋荣和秦洪武(2012)的结论一致,再次验证了旅游英语翻译文本更为正式的文体风格与更为客观的叙事态度。□

注释

① 本研究得到香港理工大学编号为A-PL73的项目资助,特此鸣谢。

参考文献

- [1] Baker, M. Corpus linguistics and translation studies: implication and application [A]. In M. Baker et al. (eds). *Text and Technology: In honor of John Sinclair* [C]. Amsterdam and Philadelphia: John Benjamins Publishing Co, 1993: 233-250.
- [2] Baker, M. Towards a Methodology for Investigating the Style of a Literary Translator [J]. *Target*, 2000(2): 241-166.
- [3] Biber, D., S. Johansson, G. Leech, S. Conrad, E. Finegan. *Longman Grammar of Spoken and Written English* [M]. Beijing: Foreign Language Teaching and Research Press, 2000.
- [4] Halliday, M. *Spoken and Written Language* [M]. Victoria: Denkin University Press, 1985.
- [5] McEnery, M. Tono, Y. Xiao, Z. *Corpus Based Language Studies* [M]. London: Routledge, 2006.
- [6] Reiss, K. Text types, translation types and translation assessment [A]. In A. Chesterman (ed). *Readings in Translation Theory* [C]. Helsinki: Oy Finn Lectura Ab, 1989: 105-115.
- [7] Scott, M. Tribble, C. *Textual Patterns: Keyword and Corpus Analysis in Language Education* [M]. Amsterdam & Philadelphia: John Benjamins Publishing Co., 2006.
- [8] Wilkinson, M. Discovering translation equivalents

in a tourism corpus by means of fuzzy searching [J]. *Translation Journal*, 2005(9): 1-10.

- [9] 陈建生, 崔亚妮. 基于语料库的中国《政府工作报告》英译本词汇特征研究 [J]. 当代外语研究, 2010 (6): 39-43.
- [10] 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.
- [11] 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(1): 73-83.
- [12] 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.
- [13] 李德超, 王克非. 汉英同传中词汇模式的语料库考察 [J]. 现代外语, 2012 (4): 409-415.
- [14] 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.
- [15] 李德超, 王克非. 新型双语旅游语料库的研制和应用 [J]. 现代外语, 2010 (1): 46-54.
- [16] 王武兴. 英汉语言对比与翻译 [M]. 北京: 北京大学出版社. 2003.

A Comparable Corpus-Based Study on Stylistic Features in Translated English Tourism Texts

Abstract: Based on a comparable corpus of English tourism texts, this paper investigates the stylistic features of translated English tourism texts by comparing their lexical variety (STTR), the average word length, the lexical density and high frequency words with those of spontaneously-written English tourism texts. The findings (1) support the translation universal hypothesis of “simplification”, which is represented by a lower STTR and a lower lexical variety of verbs, adjectives and adverbs, (2) support the translation universal hypothesis of “explicitation”, which is represented by a lower density of pronouns, (3) reveal more diverse use of tenses (especially the use of perfect tense and future tense), and (4) show more formal and objective style in translated English tourism texts with a higher frequency of the use of nominalization, passive voice and a lower frequency of the use of contracted forms.

Key Words: comparable corpus; tourism texts; stylistic features