

新型双语旅游语料库的研制和应用*

香港理工大学 李德超 北京外国语大学 王克非

提要:在评介国内外现有单语或双语旅游专门语料库的基础上,本文着重论述香港理工大学正在研制的一种新型中英双语旅游语料库的设计理念和操作程序,包括语料文本的数字化、语料的标注、语料的对齐和语料的篇头标注等。本文最后探讨了旅游语料库在推动旅游翻译研究与教学的前景。

关键词:双语旅游语料库、设计理念、旅游翻译教学与研究

[中图分类号] H313 [文献标识码] A [文章编号] 1003-6105(2010)01-0046-09

1. 引言

过去 20 年来,基于语料库的翻译研究从无到有,发展迅速。目前,翻译研究中常用的语料库主要有译文语料库(translational corpus)、类比语料库(comparable corpora)和对应语料库(parallel corpus)三种。译文语料库收录译文,其研究旨在揭示翻译语言本身独有特征(如 Mona Baker 建立的“翻译英语语料库”,即 TEC)。类比语料库主要是同一种语言的原生文本和翻译文本(译自某种外语)构建的语料库,它们之间无翻译对应关系,但在时代、体裁、主题等方面具有可比性,可用来研究翻译语言的特点。这类研究可参看 Laviosa (2002)。

相较以上两种语料库,翻译研究中更为常用的是对应语料库,它收录原文与译文双语对照的文本,通常会按事先设定的标准(如以句或段为单位的方式),对语料进行某一语言单位上的对齐(如句或段),以方便检索。对应语料库常用于考察原文中的某些语言现象如何在译文中得到反映,最终目的在于揭示翻译活动中隐性的规律。目前对应语料库在不同的国家和地区都有广泛应用,比较知名的有挪威的“英语-挪威语对应语料库”、英国的“德语-英语文学文本对应语料库”以及我

国由北京外国语大学中国外语教育研究中心研制的“通用汉英对应语料库”,等等。

纵观目前为翻译研究而建的上述三种语料库,多数以文学文本为主,或是以收纳百科文本(包括各种文学与非文学作品)为特色,而结合地域特点的或针对某一文类而专门研制的双语专门语料库则较为少见。专门语料库指的是“关于特定主题文本的集合”,且这些文本“均由行内专家为不同的读者群所写”,这些读者群可能是同行的专家,或是缺乏相关专业知识的大众,亦可能是学生等(Kübler 2003: 29)。双语的专门语料库则收集包括原文与译文的专题文本。

迄今为止,专门为翻译研究而研制的应用型双语专门语料库并不多,且通常规模较小,从几万到几十万词。小型的双语专门语料库因其规模小、建造方便和针对性强等特点,常在术语翻译研究或特定领域(如法律翻译、财经翻译等等)的译员培训中作为辅助工具,如 Kübler (2003)为培训软件指南译员而建立的小型英法计算机语料库。

双语专门语料库无论对翻译研究还是译员培训都有其潜在的价值。Johansson 早在 1991 年就提出,双语专门语料库可以帮助我们解决不少特定的问题,因而在各种语料库中,尤其值得提倡。从语言运用角度而言,双

* 本研究得到香港理工大学项目资助,编号为 1-ZV74 及 4-ZZ75。在此亦感谢匿名审稿人的宝贵意见。

语专门语料库(尤其是非文学专门语料库)能帮助译者理解和掌握同一领域中不同语言的常用术语、惯用表达式、文章语气及典型的语篇结构,避免在专业翻译中经常出现的语言“不自然”或“形似而神不似”的现象。但时至今日,无论国内国外,较具规模的双语专门语料库还不多见,如旅游方面就仅有一些小规模语料库。

2. 现有旅游专门语料库

迄今,国内外专门收录旅游文本的单语或双语语料库均不多见。纵览相关文献,我们发现国外的旅游专门语料库主要有以下三个,分别为芬兰、英国和日本学者研制。

1) 芬兰萨翁林纳翻译研究学院(Savonlinna School of Translation Studies)英语教师 Michael Wilkinson 研制的合共 67 万词英语旅游文本语料库。该语料库主要供芬兰学生将芬兰旅游文本译至英语时参考,检验他们的译文用词是否地道。同时,学生可以利用该语料库来“确认、证实或摒弃主观的决策,或是获得搭配方面的用法”(2005: 1)。

2) 英国埃塞克斯大学(University of Essex)研究生 Carlota Alcantar 于 2007 年研制的英语旅游文本语料库。该语料库总词数为 37,795,收集的旅游文本分别来自:a)政府旅游部门的介绍;b)酒店、餐馆和旅行社的宣传资料;c)语言培训学校相关的资料。研制语料库旨在总结出英语旅游文本写作中最常用的 421 个词汇和短语,以帮助墨西哥纳亚里特州立大学旅游学院的教师有效地教授旅游常用词汇。

3) 日本大学(Nihon University) Kiyomi 等人于 2006 年研制的“京都旅游语料库”(Kyoto Tourism Corpus)。此单语语料库收集的均为日本京都市用英文写就的关于京都旅游介绍的文本,共 885 篇,平均每篇 47

个词,收录了 2,786 个不同类符(types)和 42,025 形符(tokens)。库中所有的词语都用 CLAW7 标注软件消除屈折形式,还按词性作了标注。语料库旨在归纳出适合于日本旅游市场(尤其是京都地区)的几百个最常用的旅游英语词汇,以供在日本大学学习旅游英语课程的不同程度的学习者使用。

我国国内研制旅游专门语料库就更为鲜见。冯志伟(2001)提及,中国科学院自动化所建立了一个旅游咨询口语对话语料库和一个旅馆预定口语对话语料库,但关于这个口语语料库的情况却一直不为人知。除此之外,国内尚不见有其它专门研发的旅游语料库(无论是单语还是双语)。

综上所述,国内外旅游专门语料库的研制还很不够,现有的几个也存在以下缺点:一是规模小,除了 Wilkinson 的语料库有 67 万词外,其它两个语料库的规模均不超过 5 万词,并且形式单一,均为单语(英语)语料库。二是语料库的设计比较简单,选择语料的方法含糊不清,也没有说明其选择语料的标准,容易让人质疑其语料的代表性。三是相较其它综合语料库,这些语料库对语料的处理都很粗糙,除日本大学的语料库外,其它的均未对语料做后期整理(包括对齐、标注、加篇头等)。四是这些专门语料库的用途都较单一,都是为旅游英语(尤其是词汇)的教学而研制。至于旅游文本在结构、修辞上的特点等深层次的问题完全没有探究,也没有考虑到旅游翻译教学与研究上的问题。本文介绍的新型双语旅游语料库希望能克服以上弱点,既可直接用于指导现实中的旅游翻译(实践意义),亦可促进旅游翻译教学和研究(理论意义)。

3. 新型双语旅游语料库的设计思路和特点

本文以下着重描述香港理工大学中文及

双语学系最近研制的新型双语旅游语料库。这是香港以至全国第一个较大容量的英汉/汉英旅游语料库。我们主要介绍这个语料库的研制思路和方法、语料库的构成和特点,并讨论专门语料库的发展前景。

如 Kennedy 所言,无论何种语料库,其整体的设计和语料的汇集对于“基于该语料库的研究的信度和效度都具有十分重大的作用”(1998: 60)。它还直接影响到语料库所设计的研究目的和教学作用能不能有效实现。就本旅游语料库而言,设计目的有三:

一是研究用途。包括旅游语言本身特点的研究及旅游翻译特点的研究。前者研究旅游语言在语篇、修辞、词汇等语言层面上与普通语言不同的特点;后者则研究在翻译过程中旅游翻译与普通翻译的不同特点。研究问题包括:旅游翻译在语言特点上是否存在与普通翻译一样的共性?抑或有自己另外的特点?这些特点是与特定时期的翻译规范有关还是超越时间、时空限制?

二是翻译教学用途。具体来说,本语料库将为教授中英双语旅游翻译的教师提供真实的教学材料,并为设计数据驱动教学(data-driven learning, 简称 DDL)提供真实和多样化的资源。DDL 这里指的是“在课堂上利用电子计算机生成的索引(concordances)来帮助学生发掘目标语模式(pattern)中的规律,并根据索引结果来研发种种学习活动及设计学习练习”(Johns & King 1991: iii)。就本旅游语料库而言,就是根据语料库比较和检索的结果,开发出各种 DDL 的教学手段来辅助旅游翻译的教学。而学生亦能从语料库中揣摩职业旅游翻译译员所用的翻译策略,从比较、思考中学习他们老到的方法与技巧。开发 DDL 教学手段往往需要更大规模的语料库,本库建成后会持续扩容,以满足 DDL 需求。

三是实用性。即能够为国内外从事涉外旅游的英汉语从业者提供方便的旅游翻译参考或用作自学材料,提高他们的业务水平。

为了实现上述目标,本语料库在设计时

必须严谨,确保语料库的效度和信度。语料库设计时要考虑的因素很多,大体可分为整体设计和具体操作两个层面。下面分述这两个层面。

3.1 整体层面的考虑

从性质来看,本语料库属于专用(specialized)、同质的(homogeneous)语料库(参见黄昌宁 2002),只收集与旅游景点介绍相关的中、英文原生性文本(spontaneously sourced texts)(Mason 2001: 68)及相关的中-英互译文本(在时机成熟后亦会扩展至收集与旅游相关的行业,如酒店、交通等介绍的文本),这就局限了本语料库的文本采样不能象异质语料库般采用随遇性的文本采样原则,而只能是采取目的性取样标准(purposeful sampling)。为了确保采集到的文本在内容上与语料库研制目的相关,在语言上符合标准,本语料建设中除了会实行多个专家把关制之外,还会推出同侪报告(peer debriefing)及研究成员的检视(member check)等方法来核实文本的适合性和代表性。同时,对于某些介绍比较多的旅游景点文本材料,在选取这些材料入库时我们亦会附加以三角验证法(triangulation),多渠道、多方式地反复核对入库文本(参见 Li 2004)。以上这些确保文本数据有效性(validity)的做法均为传统双语旅游语料库研制中所鲜见。

下面我们就本语料库在构成、大小、语料的代表性及语料的选择等具体做法作进一步介绍。

1) 语料库的构成和大小:从整体上看,本语料库由一个双语旅游翻译对应语料库(parallel corpus, 简称 PC)与一个双语旅游翻译类比语料库(comparable corpora, 简称 CC)组成。在研究的第一阶段,这两个语料库暂定为各 100 万字/词(为统计方便,中文部分按字数计算,英文部分按词数计算),目前各个语料库已经完成三分之二。按照我们的设计,PC 与 CC 均为动态语料库(dynamic corpora),即在将来会按每两年一次的频率,

把符合条件的文本补充至各库之中,并不断提高加工的深度。与那些只是选择特定时间范围内部分语言现象的静态语料库相比,本语料库的动态设计能够反映旅游用语及旅游翻译的时代变迁和语言风格的变化。动态设计的思路会让本旅游语料库的规模逐渐扩大,同时也为研究旅游翻译和旅游语言提供了纵向研究的可能。在现阶段,PC与CC合并起来有200万字词。语料库是否适用的主要因素并非全在于其规模的大或小,而主要在于“适用于语料库的文本是否容易获得,或是取决于其它因素,诸如文本是否需要手动标注等等”(Hunston 2002: 26),以及语料库能不能实现研制目的(如旅游翻译的教学)和解决预定的一些研究问题(如旅游语言的特点)。对于本研究而言,各100万字词的PC和CC,在语料收集和分析上较易做到,且在现阶段为达到这个目标,在财力和在人力的资源上尚可操控。

PC是一个双向的汉英/英汉旅游翻译语料库,先收录香港地区的英译汉和汉译英的旅游翻译文本,并逐步将收集的范围扩大到大陆、台湾以及全球华人地区的旅游文本。这些文本包括以网页形式的电子文本和以书面形式的文字文本(包括书籍、宣传册、旅游区内容介绍等等)。

CC主要收录以香港地区为主的两岸三地非翻译(non-translation)的中文和英文旅游原生性文本。英文旅游文本将涵盖英、美、澳等主要以英语为母语的国家(国别在语料中加以标识)。与PC一样,CC收录的文本范围同样包括来自电子、书面以及其它媒介形式的文本。在构建CC时,我们注重让中文文本库与英文文本库在以下几个主要参数上具有一致性,以确保收录在CC中的中英文文本具有最大的可比性,如:文本均全文收录、长度大致相近、文本创作的时间跨度均为过去十年、文本的作者尽可能多样等等(参考

Laviosa 1997)。这些参数的一致性至少能确保CC在设计上避免由于中英文本语料的不均衡而造成研究中出现错误的或有偏差的结论。

PC与CC相结合的设计考量在于让这两种语料库的特点互补。Johansson (1998: 6)认为,PC研究中出现的一些不容易解答的问题,需以CC为比较参照系才能得到让人信服的答案。这些问题包括:译文在多大程度上体现普通语言的运用?译文在多大程度上受到原文的影响?以及译文具有什么样的总体特征?等等。正因为PC与CC结合能在语料库翻译研究中发挥更大的增效作用,近来有不少翻译学者都提倡把PC中的一些发现运用在CC的背景框架中解释,充分发挥语料库翻译研究的定性与定量研究相结合的优势(参看Kenny 2006)。除此之外,本语料库里具有双向性的PC与CC结合还能提供更广阔、更丰富的语言对比潜能。如下图所示(C表示本旅游语料库的组成部分):

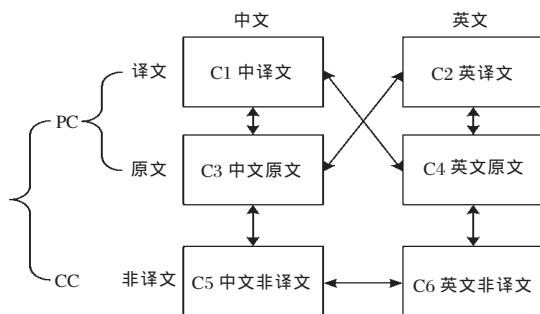


图1 旅游语料库结构图¹

从图1可见,若以库为单位,本语料库内可以作子库之间的比较,至少有以下几种可能性:一、可以是PC研究中常见的原文与译文的比较(C1:C4;C2:C3);二、可以作同一种语言之间译文与非译文的比较(C1:C3;C5:C3;C2:C4;C6:C4);三、亦可以是不同

¹ 本图部分改编自 Johansson 和 Hofland (1994: 26)。

语言之间(这里即中文与英文)非译文之间的比较(C5:C6)。同时,C5和C6还能起到参考语料库的作用,为C1:C4和C2:C3时提供尽可能多的参照系。本旅游语料库的这种设计方式为最大限度地有效利用语料库创造了条件。

PC与CC选取的旅游文本主要包括:1)自然景观、人文景观的介绍;2)观光小册子;3)酒店、旅馆、餐饮业及娱乐场所和设施的介绍;4)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。

2)语料的选取和代表性问题:香港旅游业兴旺,市场上充斥着大量与旅游相关的宣传页和小册子。数量多,质量难免良莠不齐,这就更凸现了语料选取过程的重要性。Kennedy特别指出语料的代表性问题——即所选语料能否代表这个领域中的大多数文本——是语料库建设必须考虑的一个重要问题。他认为,选择语料时必须注意的问题包括:语料的文学价值、社会上的影响、所面向的读者群体、文本地区性和时间因素等等(1998: 62-66)。由于Kennedy针对的是单语普通语料库(尤其是文学语料库)的建设,他所提及的上述要素并非全都适用于本语料库。但就本语料库而言,在面对大量有关旅游的文本时,“研究者需要考虑一系列的问题,确定一系列的标准,帮助作出选择和排除某些文本的决定”(Olohan 2004: 46)。

借用Maia(2003:45)的做法,本旅游语料库订出以下几条标准:1)据社会语域标准(如分为景点介绍、观光介绍、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结合使用,使入库语料具有一定代表性;2)所有文本,若含任何语法、用词或印刷上的错误,均不收录,或略加更正后收录;3)就翻译文本而言,任何太过贴近原文句法结

构以致僵硬的译文不予收录;4)在最后阶段,部分文本请精通中英双语的专家和翻译教师审阅其语言和用法上的可接受性。由于本旅游语料库除了用于语言研究外,还会用于教学,上述严谨选材除了尽量使语料具有代表性外,还要确保文本有一定的质量,这样才不会对旅游翻译教学造成反效果。

可能有人认为,语料若不是全部用随机方式而任意抽取,反而不利于呈现旅游文本的“最自然的状态”,因而语料库也并非所有香港旅游文本的最真实反映,而是带有很大的主观色彩。对此,我们认为:一、对话料进行适当选择是当今绝大部分语料库的通用做法,亦是一种发展趋势(如BNC就对不少语料进行了剪裁),只不过很多语料库建设者未将选材标准明确列出而已。未经挑选的大量原始语料的集合只会让语料库内容参差不齐,水平不一。由这些材料集合而成的语料库虽然最少受到研究者的“干扰”(即研究者主观性的影响),但它却不利于开展研究或教学,而后者恰恰才是语料库建设的目的。二、所有语料库都不可能做到完全客观,不受研究者主观性的影响。正如Crisafulli所言,“语料库从开始设计如就是一种(设计者)诠释的表现”(2002: 32)。不少学者亦指出,语料库的设计以及语料的收集(包括要容纳哪些主要信息和次要信息等等),都要根据研制语料库之前所设定的研究目的和研究问题而定(参看Olohan 2004: 42),只要语料库的选材能够达到预定的目的,语料库的设计和语料的组成能与研究目的相符,能让学者进行相应的研究,这就是一个好的语料选取过程。

3.2 具体操作层面的考虑

语料库具体操作层面的考虑因素包括语料文本的数字化、语料的标注、语料的对齐、语料的篇头标注等实际问题。

1)语料文本的数字化。目前,语料库所收录的文本绝大部分来源于以下三种途径:网上的电子文本、以纸质形式出现的印刷品、书籍或宣传页和拍摄下来的以建筑物形式树

立的有关旅游景区的介绍。所有的这些材料都需转为 txt 文档备用。

不少印刷品都带有插图、照片、标志等 Nord 称之为“非言语因素”(non-verbal elements) (2005: 118) 以及用不同字体大小、形式、下划线等来模仿语调、语气和停顿功能的“超音段特征”(suprasegmental features), 它们都构成了旅游文宣中不可缺少的一部分。但在转换成语料库电子文本后, 无论是“非言语因素”和“超音段特征”都得去除, 以方便标注和以后的检索, 因此, 我们对于具有这些特征的文本就在其电子文本的“篇头”上注明, 方便研究者和学习者如有需要时去打开相应的扫描文档来查找, 同时亦会在文本中出现这些“超音段特征”的地方以赋码 <SUS> 做人工标注, 让使用者能够通过计算机快速查找。以相片出现的旅游介绍则需要手工输入, 之后亦存为 txt 文档备用。

2) 语料的标注。在本文, 标注指对语料成份(通常是词性、词类)的划分, 以方便进一步检索和研究, 这种标注方式称之为 POS 标注法(part-of-speech tagging)。对于英文语料, 一般的 POS 标注法就是参考 BNC 对语料词性和词类的划分和缩写方法(如 NN1 表示单数名词、NN2 表示复数名词、NP0 表示专有名词等, 参看 BNC 网站), 先用软件(如 CLAWS 标注系统)进行机器标注。由于机器标注都不能达到百分百的标注准确率(最好的 CLAWS 标注系统也只能达到 96% 左右, 见 Kennedy 1998: 218), 因此机器标注完成后还要进行人工修正。中文亦有类似标注词性和词类的系统, 但其准确性不如 CLAWS, 且中文词性的区分不如英文的清晰, 因而标注中文语料人工后期修正的时间要花费更多。

本语料库除了按照传统的 POS 标注外, 还根据旅游文本中的一个突出的修辞特点, 在语料上采用相应的修辞手段标注。研究员在研读文本时, 分别会根据文本在语音、词汇

和语法上的修辞特点, 在语料中标出所使用的修辞或比喻手法。这种标注方法由于没有现成的软件可以做到, 一切只能人工进行, 所以费时甚巨。但一旦标注完成, 它将为揭示英汉旅游文本在修辞用法上的不同, 甚至为文本背后两种民族的审美观念的不同提供有力的佐证, 这比较修辞学、比较美学研究等都大有裨益。

针对旅游文本及旅游翻译的特点, 本语料库亦会使用一些 CLAWS7 的码集中没有的自主创设码来对语料进行标注。这些自主创设码主要有两类: 一类是用于标注上述提及的“超音段特征”, 以创设码 <SUS> 表示; 另一类是各种类型的修辞手段(rhetorical devices), 在创设码上全以 <RHE> 开头, 后面会加一个或二个字母来表明语料中所用的具体修辞手法, 如隐喻以 <RHEM> 表示, 头韵以 <RHEA> 表示, 排比以 <RHEP> 表示等等。因为 CLAWS7 软件只能标注词类, 上述标注因而不能采用机器进行, 只能手工赋码。之所以创作这些码集, 主要是因为它们所代表的语言特点在旅游文本体裁中非常显著, 且作用重大, 往往决定旅游文本的呼唤功能(vocative function)是否能得以实现, 因而有必要在语料库中将这特点明示。

另外, 标注还涉及 CLAWS7 的码集与自主创设码的兼容问题。因为作者并非 CLAWS7 的开发人员, 亦无权将自创的码集加入至 CLAWS7 的码集之中, 从而达到以 CLAWS7 软件统一以计算机附码的目的。事实上, 即便作者有权将自主创设码加入 CLAWS7 的码集, 但对于作者自主创设码的标注(如修辞手段的标注), 主要是基于对整个句子的认知理解, 以目前人工智能的发展程度来看, 计算机亦不能对此附码。为了与 CLAWS7 的码集相兼容, 我们采取的手法是: 在所有的语料用 CLAWS7 的码集标注后, 再用手工形式对处理过的语料加注自主创设码。手工赋码费时, 但准确性高。我们力求在手工赋码时做到标注格式与 CLAWS7

的码集格式完全一致,以便在检索语料时用软件统一提取。

3)语料的对齐。PC里的中、英文语料都先用 Paraconc 软件以句为单位粗略对齐,然后再人工检查,以方便用 concordancer 检索。中文语料还会用中科院计算所开发的“汉语词汇分析系统”(ICTCLAS)进行切分。旅游文本的翻译很多时候并非完全遵守句对句的原则,摘译、扩译或省译情况还比较多见。软件在某些地方找不到一对一的对应,就容易发生错配。我们设想将来在此新型双语旅游语料库的基础上,根据词频来自主开发中、英文旅游文本中常用的双语词汇(bilingual lexicon),这样就方便在语料库扩充和更新时,把这些双语词汇用作语料句与句对齐的另一附加条件,进而提高对齐的准确率。但无论如何,机器对齐之后,均需要人工检查作进一步的确定。对于摘译、扩译或是省译的现象,运用 concordancer 检索相关的句子时,可以在检索时设定 concordancer 同时出现相对应的句子的相邻的几个句子,提供更为详尽的语境,这样就能较为准确的知道是那一部分原文受到改动。语料对齐后,会将文本转换成 XML 格式,这样就能让语料库以多种语言格式编码,方便传播和检索。

CC 中汉语旅游语料的词汇切分与赋码亦是采用的是由中科院计算所的 ICTCLAS 软件。该系统分词精度高,分析速度佳,唯一不足就是对词性的标注仍不够精准。因此,经过 ICTCLAS 分词后的语料仍需大量的手工校对,以提高中文语料的词性标注正确率。

4)语料的篇头标注。篇头提供所收语料的进一步信息。根据不同研制目的,不同语料库纳入篇头的内容亦不同。本语料库里,篇头内容包括篇名、作者名、作者背景、译者名、译者背景、年代、出版信息(书籍、电子或图片)、文本字数、有无“非言语因素”及“超音段特征”等。

从以上介绍可以看出,与传统的旅游语

料库相比,本语料库的“新”主要体现在以下几点:(一)从语料库的结构成份来看,本语料库一反传统旅游语料库均由 PC 单轨组成的特点,改为由 PC 与 CC 联合双轨组成。正如图一所示,这种组合形式能够更为丰富的各子库之间比较的可能性,使研究者能更全面地考察旅游文本的体裁惯例,更深入地了解旅游翻译与其它类型的翻译(如文学翻译、法律翻译等)在词汇、句法和修辞等多个语言层面上呈现出来的不同特点;(二)从语料库的词容来看,本语料库中的 PC 与 CC 加起来共 200 万字,这亦是目前世界上最大的双语旅游语料库。更为重要的是,本语料库建成之后,每二年将更新一次语料库的内容,使语料库能反映旅游语言的最新变化。可以设想,在较长的一段时间之后,语料库这种动态的、增进式(incremental)的特点将会使研究者能够对旅游语言和旅游翻译的特点进行历时或共时的研究,从而丰富目前双语旅游语料库研究中较为单一的研究范式;(三)从语料的收集上,本语料库的语料选材严谨,特别是针对双语旅游语料良莠不齐的普遍现象,提出双语语料均需由专家认定后才能入库。这就能避免目前某些双语旅游语料库中出现的部分收录译文错误百出的毛病,从而在保证本语料库的语料具有代表性的同时,亦不会误导用户;(四)从语料库的运用前景上,传统旅游语料库作用相当单一,几乎都是为旅游英语教学,尤其是词汇教学而设;而本语料库的用途则广泛的多:它除了可以采用 DDL 形式,开展多层次、多类型的旅游英语教学与中英-英中旅游翻译教学之外,亦可以用于语言研究及供旅游从业人员使用,这样就使双语旅游语料库的价值得到更大的体现,亦能给其它专门语料库的研制和发展提供宝贵的经验。

4. 研究前景

目前,本旅游语料库只是初步建成,以后

将会逐渐由主要收集香港地区的语料扩展为收集两地三岸的中英文旅游语料,并考虑将语料库上网或制成可供检索的光盘。同时,我们还会不断完善现有的标注体系,根据研究和教学目的,进行新的语义和句法标注(参看王克非 2004: 75),最大限度地发挥语料库的潜能。在教学方面,将根据语料特点,开发出一系列可供教师在课堂上使用以及学生随时使用的教学材料(拟另文叙述),为开展DDL旅游翻译教学提供直观和真实的语料基础。另外,我们亦考虑建立学习者旅游翻译语料库,收集大学生旅游翻译作业的电子文本,把它与其它两个语料库相对照,找出学生常犯的错误,寻求产生这些错误的文本内及文本外的原因。

在研究过程中,我们基于本语料库已经做了一些研究。例如,研究英汉酒店文宣的文本比较模式,从而得出在语篇结构、语言惯例上更为适切的中译英酒店文宣译文(李德超、王克非 2009),研究基于旅游语料库而开发各种DDL教学方法。本语料库将来的研究前景相当广阔。比如,在词汇层面,我们可以比较汉英旅游语篇的词汇色彩和风格(如是正式、幽默、低调、口语化还是其它),或是通过词频来统计在特定旅游场景中最常用的词汇,作为编纂不同程度的旅游教材的基础,或是考察特定的形容词(如“美丽的”)在译文中对应的词汇,从而挖掘旅游翻译过程中的审美因素,等等。在句子层面,我们可以考察汉英旅游语篇在句子结构和句子复杂程度的偏好,或是研究句子选择主动或被动表达式的深层原因,或是了解句子在不同的文化场合所传达的不同语用效果,所能激发起读者不同的文化联想,等等。在语篇层面,我们可以比较汉英旅游语篇在衔接手段上的差别,或是归纳出它们在语步结构和体裁结构潜能上的不同特点,从而考察它们在社会认知根源上的异同。而在修辞层面,我们亦可以比较汉英旅游语篇最常用的修辞手法,考察读者对这些手法的接受和反应,或是考察中文旅游

语篇中用到的特定修辞手法在英文中的习惯处理方法,并了解它们在译文读者大脑中构成的或主观、或客观、或夸张的印象,等等。除此之外,我们亦可从翻译的共性入手,探讨旅游翻译是否具有“共性”,而目前翻译研究中发现的一些所谓“共识”(common ground)是否又适用于旅游翻译,等等。当然,这些研究前景都需在进一步提升本语料库研制程度之后进行。

参考文献

- Alcantar, C. 2007. Construction of a special corpus to extract the most frequent words in a particular field [J]. *Essex Graduate Student Papers in Language & Linguistics* 9: 1-17.
- Barnbrook, G., P. Danielsson & M. Mahlberg. 2006. *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora* [M]. London and New York: Continuum.
- Chujo, K., M. Utiyama & K. Oghigian. 2006. *Selecting Level-Specific Kyoto Tourism Vocabulary Using Statistical Measures* [P]. Paper presented at Conference of 2006 International Teaching and Learning: New Aspect of English Language Learning in Taiwan. Available online at URL <http://www2.nict.go.jp/x/x161/members/mutiyama/pdf/Kyoto-Article-Chujo-Utiyama-Oghigian.pdf>
- Crisafulli, E. 2002. The quest for an eclectic methodology of translation description [A]. In T. Hermans (ed.). *Crosscultural Transgressions: Research Models in Translation Studies II* [C]. Manchester: St. Jerome.
- Hermans, T. (ed.). 2002. *Crosscultural Transgressions: Research Models in Translation Studies II: Historical and Ideological Issues* [C]. Manchester: St. Jerome.
- Hunston, S. 2002. *Corpora in Applied Linguistics* [M]. Cambridge: Cambridge University Press.
- Johansson, S. 1998. On the role of corpora in cross-linguistic research [A]. In S. Johansson & S. Oksefjell (eds). *Corpora and Cross-Linguistic*

- Research* [C]. Amsterdam; Rodopi.
- Johansson, S. & K. Hofland. 1994. Towards an English-Norwegian parallel corpus [A]. In U. Fries, G. Tottie & P. Schneider (eds). *Creating and Using English Language Corpora* [C]. Amsterdam; Rodopi.
- Johansson, S. 2003. Reflections on corpora and their uses in cross-linguistic research [A]. In F. Zanettin, S. Bernardini & D. Stewart (eds). *Corpora in Translator Education* [C]. Manchester; St. Jerome.
- Johansson, S. 1991. Times change and so do corpora [A]. In K. Aijmer & B. Altenberg (eds). *English Corpus Linguistics: Studies in Honour of Jan Svartvik* [C]. London and New York; Longman.
- Johns, T. & P. King. 1991. Classroom concordancing [J]. *ELR Journal, New Series* 4.
- Kennedy, G. D. 1998. *An Introduction to Corpus Linguistics* [M]. London / New York; Longman.
- Kenny, D. 2005. Parallel corpora and translation studies: Old questions, new perspectives? Reporting that in GEPCOLT: A case study [A]. In G. Barnbrook, P. Danielsson & M. Mahlberg (eds.). *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora* [C]. London; Continuum.
- Kubler, N. 2003. Corpora and LSP translation [A]. In F. Zanettin, S. Bernardini & D. Stewart (eds.). *Corpora in Translator Education* [C]. Manchester; St Jerome.
- Laviosa, S. 1997. How comparable can “comparable corpora” be? [J]. *Target* 9; 289-319.
- Laviosa, S. 2002. *Corpus-based Translation Studies: Theory, Findings and Applications* [M]. Amsterdam; Rodopi.
- Li, Defeng. 2004. Trustworthiness of think-aloud protocols in the study of translation process [J]. *International Journal of Applied Linguistics* 3; 301-313.
- Maia, B. 2003. “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.
- Mason, I. 2001. Translator behavior and language usage: Some constraints on contrastive studies [J]. *Hermes* 26; 65-80.
- Nord, C. 2005. *Text Analysis in Translation: Theory, Methodology, and Didactic Application of a Model for Translation-oriented Text Analysis* (2nd edition) [M]. Amsterdam/Atlanta; Rodopi.
- Olohan, M. 2004. *Introducing Corpora in Translation Studies* [M]. London/New York; Routledge.
- Wilkinson, M. 2005. Discovering translation equivalents in a tourism corpus by means of fuzzy searching [J]. *Translation Journal* 9; 1-10.
- Wilkinson, M. 2009. *Compiling a Specialized Corpus to Be Used as a Translation Aid* [R]. Available online at URL <http://www.joensuu.fi/hallinto/jopke/dokumentit/Wilkinson.doc>
- Zanettin, F., S. Bernardini & D. Stewart (eds). 2003. *Corpora in Translator Education* [C]. Manchester; St. Jerome.
- 冯志伟, 2001, 中国语料库研究的历史与现状 [A]. 国际中文计算机会议 ICCCL2001 论文集(C). 新加坡。
- 黄昌宁, 2002, 语料库语言学 [M]. 北京: 商务印书馆。
- 李德超、王克非, 2009, 平行文本比较模式与旅游文本的英译 [J]. 中国翻译 (4): 54-58。
- 王克非, 2004, 新型双语对应语料库的设计与构建 [J]. 中国翻译 (6): 73-75。
- 收稿日期: 2009-07-09;
作者修改稿, 2009-07-28;
本刊修订, 2009-12-28
- 通讯地址: 北京外国语大学中国外语教育研究中心
<kfwang126@126.com> (王)
香港理工大学中文及双语学系
<ctdechao@polyu.edu.hk> (李)

contrasting.

On the semantic construal of [shèng] and [bài] from the Construction Approach, by Li Guohong, p.31

The Chinese antonyms [shèng] and [bài] sometimes are synonymous with each other. To address this strange phenomenon, the paper analyzes their semantic properties from the cognitive perspective, with the findings that (1) [shèng] bears a definite semantic direction, while [bài] is opaque, and (2) when acting as a complex predicate juxtaposed with other words in a two-argument construction, [shèng] and [bài] take the following sequence in deciding their semantic direction: direction of the directional words > direction of the imperfective > inherent direction of the perfective > direction of the atemporal > prototypical direction of the perfective; when acting as a one-argument predicate, [shèng] and [bài] are semantically directed to the agent.

A study of Kenneth Burke's philosophy of language from his concept of "terministic screen", by Ju Yumei, p. 39

This paper explores Kenneth Burke's philosophy of language. It begins by analyzing his concept of "terministic screen". And based on the analysis, it studies Burke's philosophy of language from the following four aspects, i.e., human beings cannot live without language; language performs the function of being attitudinal and persuasive; language does not simply reflect reality; language use is full of motives. Through the study, the paper attempts to argue that rhetoric, especially Western rhetoric, is interdisciplinary in nature and hopes that it might shed enlightenment on the development of modern Chinese rhetoric.

Development and application of bilingual corpora of tourism texts: A new approach, by Li Dechao and Wang Kefei, p.46

Based on a critical review of the existing monolingual and bilingual corpora of tourism texts both at home and abroad, this paper introduces the design rationale and the practical consideration for Bilingual Corpora of Tourism Texts (the Corpora), which are being developed in the Hong Kong Polytechnic University. The practical consideration of the Corpora includes the digitization, the tagging, the alignment and the design of the header for the texts to be included in the Corpora. The paper concludes by pointing out the potentials for the teaching and research of tourism translation based on the Corpora.

The representation of ambiguous words in Chinese EFL learners' mental lexicon: A developmental model, by Zhao Chen, p.55

This paper reports on an experiment exploring Chinese EFL learners' mental representation of three types of ambiguous words: homonymy, metaphoric polysemy and metonymic polysemy. The results show that the representation of the three types of ambiguous words in Chinese EFL learners' mental lexicon is a developmental model: the sense relatedness of homonymy and metaphoric polysemy becomes ever lower as the learners' language proficiency improves. On the contrary, the sense relatedness of metonymic polysemy is ever higher when the learners achieve a higher proficiency. The model is consistent with the usage-based approach for second language learning.

An empirical study of the psychological reality of the gap in Chinese topicalization: Evidence from filler priming, by Cai Rendong and Dong Yanping, p.64