翻译英语语料库

——新型翻译研究的利器*

陈　伟

(武汉理工大学 外国语学院,湖北 武汉 430070)

摘要　翻译英语语料库(TEC)是世界上首个当代翻译英语语料库,包含有许多源语书面文本的英语译文。它是专门为研究英语译文的特征而建设的,文本除了对译文作了简单标注外,还对有关译文的译者、出版社/商和翻译过程等元数据作了详细标注。TEC文本译者都是以英语为母语的人,而且多数文本都是1983年以后翻译的,代表了当代英语译文的一般特征。在通过对TEC的全面解读的基础上,介绍了TEC的主要特征,分析了TEC作为一种利器用于新型翻译研究的优势和不足之处,并对其潜在用途作了简要概括。

关键词　翻译英语语料库;翻译研究;译语文本;标注;查询

中图分类号:H315.9　　**文献标识码**:A　　**文章编号**:1008-3456(2009)01-0100-05

Translational English Corpus

——A Powerful Tool for New Paradigm of Translation Studies

CHEN Wei

(*School of Foreign Languages*, *Wuhan University of Technology*, *Wuhan*, *Hubei*, 430070)

Abstract　TEC (Translational English Corpus) is the first and largest translational English corpus in the world, which consists of written English translations from a range of source languages. Designed specially for the purpose of studying translated texts, it has header files stored with metadata about the translators, publishers and translation process along with the light annotation of the translated texts. The translations can represent the general characteristics of contemporary English translations as they are done by native speakers of English, and most texts dated from 1983 onwards. Based on a thorough exploration, this paper introduces the major features of TEC, analyzes the advantages and limitations of TEC as a powerful tool for new paradigm of translation studies, and outlines the potential uses of TEC.

Key words　TEC (Translational English Corpus); translation studies; translated texts; annotation; query

从20世纪60年代起,语料库的发展不仅为语言研究提供了广泛的言语素材,而且使传统的语言研究方法从通过内省、自造例证或诱导询问的取样转变为调查取样,材料真实可靠。然而,在语料库发展过程中,语料库语言学家由于考虑到译语和自然"规范语言"的差异,开始似乎并没有对翻译实践和翻译研究产生多大兴趣,直到20世纪90年代中期以来,Mona Baker、Gideon Toury和Kristen

收稿日期:2008-10-27
* 国家留学基金委(CSC)留学基金项目(2003842151)。
作者简介:陈伟(1967-),男,副教授,硕士;研究方向:语料库语言学和翻译学。

Malmkjaer 等一批翻译理论家开始将语料库运用于翻译研究,对翻译的性质和特征进行描述,因为他们认为翻译要发展成为一门独立的学科,必须建立可靠的方法论和明确的研究步骤来对研究对象进行充分描述和分析,使个人和局部的研究成果能够在"同一个语料库或另一个语料库中"得到反复验证[1]81。Mona Baker 也希望探索出一种研究方法,避免主观想象、个人或局部的翻译经验的影响,探索翻译的规范和普遍性[2]389。

一、翻译英语语料库的筹建

在长期的翻译实践和翻译学术研究中,英国曼彻斯特大学翻译与跨文化研究中心的 Mona Baker 教授发现,在翻译文本中存在一些显著的特征,并且这些特征有时表现不一[3]37。她认为这些特征可能与某种语言的具体语言特征有关,于是她最先提出论断:只有使用可比语料库(comparable corpus),通过比较译文文本与非译文文本才能抓住译文自身的这些显著特征[4]176。在 Baker 教授提出使用翻译语料库和非译文的可比语料库研究翻译的观点之后,她和 Sara Laviosa 一起开始设计和编写翻译英语语料库(Translational English Corpus,简称为 TEC)。Baker 教授在 1995 年首次提出 TEC 的基本框架结构,但是其具体设计原则在 Laviosa-Braithwaite 的论文中才阐述清楚[5-6]。

TEC 是当今翻译英语语料库,也是世界上首个这种类型的语料库,由 Mona Baker 教授负责建造和管理,该项目得到英国科学院的基金资助,开始于 1996 年,从 1999 年起用户可在线浏览免费使用,其网址为 http://www.art.man.ac.uk/SML/ctis/research/或 http://www.monabaker.com/tsresources/ 或 http://ronaldo.cs.tcd.ie/tec/。客户端处理语料库的 Java 程序可从网上免费下载,其程序由都柏林三一大学(Trinity College Dublin)的 Saturnino Luz 设计,目前他本人在负责该语料库的技术维护。

二、翻译英语语料库的主要特征

1. TEC 的技术构成

TEC 由世界上一些语种的公开发行出版物未加删改译成英语的文本组成,译文的源语包括法语、德语、西班牙语(包括西班牙的国语、南美洲和中美洲的西班牙语)、葡萄牙语(包括葡萄牙本国和巴西的葡萄牙语)、意大利语、威尔士语、波兰语、阿拉伯语、汉语、希伯来语、泰语和泰米尔语等。文本范畴有四种:传记、小说、报纸和飞机上的休闲杂志(in-flight magazines)。其中 80%以上是小说,小说和飞机上的休闲杂志这两者的内容约占 95%。至今为止,其库容为 1000 万词次。一旦获得其它的文本版权许可,扫描、编辑和简单的标注之后就可继续增加新的翻译文本[3]60。TEC 文本译者都是以英语为母语的人,译者中男女都有,而且多数文本都是 1983 年以后翻译的,代表了当代英语译文的一般特征[7]。

2. TEC 的标注

为了能进行深层次的译文特征研究,正如目前世界上的其他大型语料库一样,TEC 的建设者们也对它进行了标注,标注形式有两种:文本标注和元数据标注。对于译文文本,TEC 只是做了简单标注,其目的是确保译文文本自身的整体性,对于编辑提示和前言等非翻译部分,尽管也作了标注,但是在数据库索引程序中这种与译文本身的相关信息作了特殊的技术处理,因此它们被隐含起来,不会在索引词条或词频表中出现[8-9]。由于研究的需要,TEC 的元数据详细记录了翻译文本如超语言特征:译者的姓名、性别、国籍、职业、翻译的方向、译文的源语和出版社/商名、文本的类型和字数统计、原文作者姓名、性别、国籍、地点和年代等。所有这些是以独立的文本附加信息(header file)形式标注的,采用 XML 标码,与目前世界上通行的标注法则如 TEI 或元数据标码虽然不一致,但这些标注支持以下一些方面的翻译研究,如比较和分析实词/虚词比率(lexical density)[10]237,243、类符/形符比率(type/token ratio)[11]、句长、词语搭配规律、具体词语在男女不同译者的译文中的使用频率,同一分库中译自不同源语的译文差异和具体译者的翻译特点等[8]。

3. TEC 核心软件的特色

由于 80%以上的 TEC 文本是翻译小说和传记,它们受到版权保护,所以网上用户只能通过专用索引服务器和浏览器浏览部分文本及其扩展语境文本。一般来说,传统语料库建成大型的集约文本库,以刻录成光盘等主要形式向公众发行或让用户通过国际互联网访问语料库网站。而 TEC 采用了不同的设计理念,那就是在不同地点建成几个小型或中型的语料库归属不同单位维护,而语料查询、子库的选择和索引合并等用户与服务器的互动则在用户端

完成,而且其用户端浏览器能兼容多种平台[8]。因此TEC专用软件的设计可谓是独具匠心,别具一格。

具体说来,TEC索引浏览器有两个版本:其一是用于多数浏览器的applet插件,其次是完全用Java语言编写的程序。其核心程序兼有以下功能:索引行排序、浏览与索引行相关的元数据、激活用XML编码的文本文件和文本附加信息、提取扩展语境以及在客户端电脑磁盘上保存搜索结果。用户句法查询有三种功能:区分大小写,采用统配符*,设定查询词语在句法中的具体词序。元数据是通过HTTP协议传递的,因此主要用户端既可处理索引服务器的主打协议,还可处理用于查询元数据HTTP标准协议[8]。

4.TEC的功能和使用特点

TEC的功能包括两大方面:语料查询和语料处理。查询分为两种:一般查询和特色查询。语料处理则包括设定左右±6的跨距(span)来考察与节点词(node)的搭配情况和索引行的排序。由于篇幅限制,这里主要介绍一下TEC的查询方法。TEC一般查询句法较为简单,其句法公式为:word_1[+[[no_of_intervening]]word_2...],Word_1,word_2分别表示要查询的第1个词和第2个词,no of intervening表示要查询的两个搭配词之间相隔的跨距数。这里重点提示一下使用统配符*和确定词序来进行模糊查询的方法。例如,输入test*,就可查询到词头含有test的各种词项,如test,tests,testament等。定义词序是指设定关键词或统配符*在查询句法序列中的位置以及该序列中插入词的最大长度,如输入seen + before就能找到包括...never seen before...在内的所有事例(instances);如果输入seen+[1]before就能找出包括...seen her before...,...seen ie before...在内的seen和before之间只有一个词的所有事例;如果结合统配符*输入know + before * 就能查到...know before...,...know beforehand,...等事例。

除此之外,我们还可通过TEC索引浏览器轻而易举地进行特色查询,选择和搜索可以基于原文作者名、源语的类型、译者的姓名、籍贯、性别和性别取向等参数,方式多样。这主要得力于在语料库里以文本附加信息方式在每一个文本上进行了标注,这种有关译者和翻译过程的语外信息是通过问卷调查形式从出版社/商和译者本人那里得到的,比较真实可靠[3]60。这些以XML标注的元数据被储存在SQL数据库中,因此被选到的子库也能识别标准的SQL句法[9]。特色查询需要使用以下一些代码:authName表示作者姓名,aGender表示作者性别,aNation表示作者的国籍,aSexO表示作者的性别取向,transName表示译者姓名,tGender表示译者性别,tNation表示译者国籍,tSexO表示译者性别取向,language表示源语语种,filename表示文本名。如输入以下查询语句:aGender ="female" AND tGender ="female" AND (tNation ="British" OR tNation="Australian"),就可以找出TEC中某位国籍为英国或澳大利亚的女译者翻译的女作家子库的所有文本。如果不习惯使用代码查询,还可以通过在浏览器界面主菜单中用鼠标点击上述参数,然后输入查询,接下来的步骤同一般查询方法类似。

三、对翻译英语语料库的评价

基于语料库的翻译研究具有很多优点,基于翻译英语语料库(TEC)的翻译研究也是如此。首先,它不仅有助于综合语言学和文化学的方法来研究翻译,而且还有助于了解意识影响翻译的程度,更重要的是,它能解决使用或借鉴现代科学技术来为翻译服务的方法论问题[12]657。因此,这种翻译研究具有方法论上的优势。其次,采用基于TEC的译文语料库可以研究译文的普遍性特征,这种系统的研究不仅可以采用描述式方法来辨认出译文的典型性特定句式,而且还可以提高译者的翻译水平。另一方面,我们可以进行纵深研究,从纯粹的定量分析深入到更为繁杂的句法层次的纵聚合研究(colligational patterning)[3]108。其三,采用像TEC这样的译文语料库和其他单语可比语料库来比较和研究译文普遍性特征的研究方法,综合了定量和定性的方法的优势。例如,Baker教授就通过基于TEC和BNC的研究指出了翻译文本和翻译过程中的"翻译总特征"——明朗化(explicitation),简易化(simplification)、规范化(normalization)和中性化(levelling out)[4]176-177。在通常情况之下,如果只凭个人的直觉是难以轻而易举、清楚明了地得出这些结论的。其四,像TEC这样用户界面友好、免费的自动处理文本软件和网络语料库的开通会使译者根据各自不同的需要对单个文本或分库来从事语言、文体和篇章等方面的研究,对于那些兼有研究人员素质和职

业翻译技能的翻译者来说,这会使他们结束那种长期以来学术研究和职业翻译之间不能很好结合的局面[13]。

尽管如此,TEC 无论是在技术构成、功能设计还是在基于 TEC 的翻译研究方面还存在一些局限性,有待日后改进。由于种种原因的限制,TEC 收集的译文源语数量还是不够多。众所周知,目前世界上有不少较大的语种,特别是那些与英语差异较大的语种如汉语、俄语、日语等更有研究价值,而 TEC 收集的来自汉语的译文只有一个文本,而且其内容是有关毛泽东的传记,这种体裁在日常生活里的运用似乎不太普遍。令人惊讶的事,TEC 没有收录译自俄语、日语等语种的译文,至少在译文的源语代表性方面还存在一定的缺陷。目前 TEC 只有四种语体的译文,对于英语译文翻译特征的研究来说似乎显得有点单薄。Olohan 提到,她们还可往 TEC 里添加其他语体的素材,如:有关社会科学、政治、历史等方面的非小说翻译作品,这样一来将会与小说子库形成鲜明的对照,同样,如果她们添加更多的传记文体素材,就会为小说文体与纪实文体之间的衔接研究起到一定的弥补作用[7]。与其他大型语料库相比,TEC 的功能显得较少。如 Plug-in 插件功能目前还只能显示词频表一项,词类(POS)查询、搭配分析、图表可视化显示等其他功能还有待开发[9]。如果把译文文本像国际英语语料库英国英语分库(International Corpus of English-Great Britain,简称为 ICE-GB)那样进行复杂的语法标注就可进行句法结构层次的深层分析[14]。

并非所有学者认为对译文文本(translation product)的分析会有助于认识翻译过程(translation process)。Olohan[3]39 转述 Dominic Stewart 的话说:使用可比语料库的研究重点是放在研究译文上,而使用平行语料库翻译研究的重点是放在研究翻译过程上则具有更大的优势。她认为,要研究翻译过程就需要分析原文和译文的相互关系,而这点却不可能指望在可比语料库里研究。因此,如果 TEC 在框架设计上增加类似平行语料库的对比功能,即像英语-挪威语平行语料库(English-Norwegian Parallel Corpus)那样设计成兼有平行语料库和可比语料库的特点,其用途将会进一步拓宽。当然,这会涉及到大量的诸如库容过大、配套软件复杂等技术难题,但是,这样的多功能翻译语料库从理论上说是可以实现的。

四、翻译英语语料库的应用潜力

TEC 官方网站的资料显示它支持两大领域的研究:翻译文本与同语种的非翻译文本句式结构差异研究和译者各自翻译风格差异研究。回顾所有基于 TEC 的翻译研究后,我们可以看出,大多数基于 TEC 的翻译研究集中在研究译文与源文的词法和句法特征差异上,如:明朗化、简易化和规范化。部分研究还涉及到了类比构词现象(lexical productivity)、名词化现象(nominalization)、词语多样性(lexical variety)、词频分析、类符/形符比率、平均句长和叙事结构特点、搭配规律和语义韵(semantic prosody)、男女译者的风格差异、语义场和文化信息差异等[15]67-73。

就目前 TEC 的功能而言,作为一种新型翻译研究的利器,我们认为 TEC 有待开拓的领域其实还可包括以下及各方面[15]72:

1. 译文自身的挖潜研究

(1)译语文化中的翻译研究:包括具体文本、作者或学派本身所在的文化习俗,翻译对译语文化的影响以及译语文化中的翻译取舍问题。(2)翻译中的省略问题,尤其是非语境因素的省译现象。(3)译者性格和性别对译文的影响,特别是不同性格的男女译者对异性作者作品的措辞斟酌和调子的处理。(4)译者所在的国家政治背景、法规等翻译文本的选择和翻译策略的取舍问题。(5)不同时期、同一时期文学作品翻译的共时和历时研究。(6)女性翻译研究、主体文化的规范如何制约和影响翻译政策和策略。(7)小说、传记、新闻等不同文体的翻译技巧和翻译策略的差异研究。

2. 与其他语料库组成研究

与其他语料库组成可比语料库、多语种语料库研究语言学特征、文本的风格、新语新词的翻译特点、语言习惯如语言的冗余度、词语搭配、连贯方式、句法模式和标点符号的使用特征。

3. 翻译培训和翻译教学

既然所有的译文都是由以英语为母语的人翻译的,作为鲜活话语的一种语料形式,这些译文自然可以选用作为翻译教学的材料。学习翻译的学生可以在线免费登陆 TEC 的官方网站,进行语料的在线处理和分析,十分方便。

五、结语

尽管收集语料的局限和现有语料的先天不足导

致语料存在代表性问题,人们对于语料库在语言研究中的作用和科学性褒贬不一,但是总体来说,多数人持肯定的态度。许家金教授[16]7援引 Tognini-Bonelli 的话说,语料库语言学"为语言研究提供了一种方法论基础,同时它又给语言学的研究提供了新的哲学思路"。就语料库用于翻译研究而言,其主要的研究目的是比较语言间的差异。Tymoczko 认为"基于语料库的翻译研究如果偏重差异方面的研究,如差异、多样性、文化和语言特色等,还大有潜力可挖"[12]657。而 TEC 正好可成为进行这些新型翻译研究强有力的工具。

Baker 教授被公认为是世界上最早利用语料库来进行翻译研究的先驱之一,她在繁忙的翻译教学和研究中筹建了这个世界上目前最大的翻译英语语料库,既为翻译研究的重要转型做出了巨大的贡献,也为全世界对这种新型翻译研究有兴趣的学者和学生建立了一个大家可以免费登陆、共同切磋的平台。她本人极力推举这种定性和定量相结合的翻译研究,她[10]235在论文中曾引述 John Sinclair 的话说:"基于语料库的翻译研究前景广阔,不仅切实可行,而且不久会成为规范(norms)"。日益兴起的基于语料库的翻译研究(CTS)促使了翻译研究方法从规约式转向描述式。因为 CTS 既重视研究翻译过程和译文文本[17]67,[18]7,也重视其他方面,小到具体译者选择文本的细节,大到文本的内外文化特点。此外,语料库的变通性、可适应性和建造语料库的开放性使得这种新型的翻译研究具有一定的优势[12]652-653。由于 TEC 是世界上首个翻译英语语料库,自从它建成以来,一些学者利用它取得了不少成果,显示出运用这种利器进行新型翻译研究的威力,而且人们相继模仿翻译英语语料库建起了其他语种的翻译语料库,如翻译芬兰语料库,为翻译研究的重要转型确立了一种新的研究范式[15]73。

参 考 文 献

[1] TOURY G. In Search of a Theory of Translation[M]. Tel Aviv:the Porter Institute for Politics and Semiotics,1980.

[2] 廖七一. 当代英国翻译理论[M]. 武汉:湖北教育出版社,2001.

[3] OLOHAN M. Introducing Corpora in Translation Studies [M]. London:Routledge,2004.

[4] BAKER M. Corpus-based Translation Studies:The Challenges That Lie Ahead [A]. SOMERS H. Terminology, LSP and Translation:Studies in Language Engineering, in Honour of Juan C. Sager[C]. Amsterdam: John Benjamins, 1996: 175-186.

[5] LAVIOSA-BRAITHWAITE S. The English Comparable Corpus (ECC):A Resource and a Methodology for the Empirical Study of Translation [D]. Manchester:UMIST,1996.

[6] LAVIOSA S. How Comparable Can Comparable Corpora Be? [J]. Target,1997(9):289-319.

[7] OLOHAN M. Comparable Corpora in Translation Research: Overview of Recent Analyses Using the Translational English Corpus [OL]. [2004-12-03]. http://www. ifi. unizh. ch/cl/yuste/postworkshop/repository/molohan. pdf.

[8] LUZ S,BAKER M. TEC:A Toolkit and Application Program Interface for Distributed Corpus Processing[OL]. [2004-12-03]. http://www. ldc. upenn. edu/exploration/expl2000/papers/luz/.

[9] LUZ S. Web-based Corpus Software[OL]. [2004-12-03]. http://ronaldo. cs. tcd. ie/tec/CTS_SouthAfrica03/notes/wbcslides. pdf. 2004.

[10] BAKER M. Corpus in Translation Studies:An Overview and Some Suggestions for Further Research [J]. Target,1995(2): 223-243.

[11] 杨惠中. 语料库语言学导论[M]. 上海:上海外语教育出版社,2002.

[12] TYMOCZKO M. Computerized Corpora and the Feature of Translation Studies [J]. Meta,1998(4):652-659.

[13] OLOHAN M. Leave it out! Using a Comparable Corpus to Investigate Aspects of Explicitation in Translation[OL]. [2004-12-03]. http://www. cadermos. ufsc. br/download/9/pdf/Maeve-cadermos9. pdf.

[14] 陈伟. 国际英语语料库英国英语分库解析[J]. 外语学刊,2006(5):71-75.

[15] 陈伟. 翻译英语语料库与基于翻译英语语料库的描述性翻译研究[J]. 外国语,2007(1):67-73.

[16] 许家金. 语料库语言学的理论解析[J]. 外语教学,2003(6): 6-9.

[17] HOLMES J. The Name and Nature of Translation Studies [M]// HOLMES J. Translated! Papers on Literary Translation and Translation Studies. Amsterdam: Rodopi, 1998: 67-80.

[18] BASSNETT S. Translation Studies [M]. London:Routledge, 1991.

(责任编辑:侯之学)