

钱多秀 袁丽珺 夏 历 著

基于语料库的科技论文摘要 汉英翻译研究与对比研究



南京大学出版社

基于语料库的科技论文摘要 汉英翻译研究与对比研究

钱多秀 袁丽珺 夏 历 著



南京大学出版社

图书在版编目(CIP)数据

基于语料库的科技论文摘要汉英翻译研究与对比研究/
钱多秀,袁丽珺,夏历著. — 南京:南京大学出版社,
2019.1

ISBN 978-7-305-20848-5

I. ①基… II. ①钱… ②袁… ③夏… III. ①科学技
术—论文—文摘—英语—翻译—研究 IV. ①N533
②H315.9

中国版本图书馆 CIP 数据核字(2018)第 194871 号

出版发行 南京大学出版社

社 址 南京市汉口路 22 号 邮 编 210093

出 版 人 金鑫荣

书 名 基于语料库的科技论文摘要汉英翻译研究与对比研究

著 者 钱多秀 袁丽珺 夏 历

责任编辑 张淑文 编辑热线 025-83592401

照 排 南京理工大学资产经营有限公司

印 刷 南京玉河印刷厂

开 本 787×960 1/16 印张 15.25 字数 205 千

版 次 2019 年 1 月第 1 版 2019 年 1 月第 1 次印刷

ISBN 978-7-305-20848-5

定 价 75.00 元

网 址: <http://www.njupco.com>

官方微博: <http://weibo.com/njupco>

官方微信号: njupress

销售咨询热线: (025)83594756

* 版权所有,侵权必究

* 凡购买南大版图书,如有印装质量问题,请与所购
图书销售部门联系调换

目 录

第一章 引 言	1
1.1 现状简介	2
1.2 本书的主要内容、基本思路、研究方法	3
1.3 全书内容简介及章节介绍	6
第二章 语料库方法在对比研究、翻译研究和翻译技术研发中的应用	7
2.1 引言:语料库与语料库语言学	7
2.2 语料库与对比研究	9
2.3 语料库与翻译研究	10
2.4 语料库与翻译技术研发	11
2.5 共性与个性	13
2.6 研究前景	15
第三章 中英文科技论文摘要特点综述	17
3.1 引言	17
3.2 科技论文摘要的语篇特征及分析	18
3.3 摘要语言特征研究	26
3.4 小结	37

第四章 科技论文摘要汉英翻译研究综述	41
4.1 引言	41
4.2 科技论文摘要英译研究的现状与特点	42
4.3 具体实践中的问题及对策	51
4.4 小结	52
第五章 科技论文摘要语料库建设及总体分析	55
5.1 语料库来源及规模说明	56
5.2 语料库预处理	60
5.3 语料库内容分析	62
5.4 小结	76
第六章 基于语料库的科技期刊论文摘要语步的对比研究	79
6.1 引言:语步	79
6.2 研究设想与方法	81
6.3 语料库建设	82
6.4 研究发现	84
6.5 小结	88
第七章 基于语料库的科技论文摘要中的名词化研究	91
7.1 研究设计	91
7.2 研究发现	105
7.3 小结	131
第八章 计算机辅助翻译系统建设与应用	135
8.1 ParaConc 的使用	135

8.2 利用 Wordfast 建设计算机辅助翻译系统及其应用	152
8.3 小结	171
第九章 研究的教学应用价值——以《计算机辅助翻译》课程为例	173
9.1 语料搜集和术语确定	174
9.2 翻译记忆库与术语库处理	175
9.3 处理新翻译项目及库更新	182
9.4 小结	200
第十章 总结	201
参考文献	205
附 录 四种期刊摘要语步标注示范	219
索 引	227
后 记	233

第一章 引言

早在1990年,Swales就指出,第三世界国家的科学家由于受到英语表述能力的制约,其科研成果得不到应有的承认。近年来,中国科学家陆续在国际顶尖的权威杂志(如 *Nature* 和 *Science*)上发表有影响力的论文,中国科技期刊也渐渐受到国际社会的关注。但在国际高级别期刊中,由中国本土编辑出版的非常少,论文及其摘要被检索的数量也很少,这在某种程度上说明我国出版的科技期刊国际化程度还很低。在今天,Swales(1990)所指出的问题仍有认真研究的必要。

一般而言,国际化是指对在某地用某语言准备的某一产品或服务进行改变,使其能够被其他地区其他语言的市场用户接受和使用。其第一步通常涉及文档的翻译,与语言有关。科技期刊国际化的指标可以分为期刊内容的国际化和内容以外其他诸方面的国际化。前者的重要性自不待言,而后者就包括语言载体的使用。对比国内期刊与国外期刊可以发现,我国期刊国际化过低的一个主要原因就是它们绝大部分使用的是中文,有的期刊没有英文题目和摘要,或有英文摘要但不符合国际检索系统的要求。这说明摘要的英文写作或汉英翻译非常关键,是期刊国际化的一个重要因素。因此,对该领域进行研究具有及时性和迫切性。

本书将从语言层面对这一领域进行对比性实证研究,研究将以自建的平行语料库和可比语料库为基础,以科技期刊论文摘要的汉英翻译为切入点。

1.1 现状简介

自 20 世纪下半叶以来,语料库及语料库语言学发展迅速,许多国家和地区已陆续建立起规模不等、类型各异、功能不同的语料库,开辟了多层次、多维度研究的新格局。基于语料库的研究已经成为相关学科的主流研究范式之一。

双语平行语料库对语言对比、翻译研究、语言教学等有着非常重要的价值。在翻译研究领域,它的应用尤其广泛,如 Baker(1993,2000)等人的研究,涉及文学翻译产品的各种特性、译者风格探索、译者培训等多个领域。可比语料库同样对上述领域贡献颇多,如 Laviosa(1997,2002)等人的研究,通过词汇变化、词汇密度、句长等指标,探讨了翻译文本的普遍性特征。

这些基于语料库的翻译研究通过统计数据讨论出相应的结果,有助于在整体上把握文本,具有一定的客观性和可靠性。

国内双语平行语料库的建设与研究起步较晚,已建成的平行语料库根据用途大致可分为两类:面向语言研究的平行语料库和面向自然语言处理研究的语料库。前者如上海交通大学的中国英语学习者语料库、北京外国语大学的大型通用汉英平行语料库 PCEE、燕山大学建立的《红楼梦》汉英平行语料库、南京大学建设的南大-国关平行语料库和外语教学与研究出版社建设的英汉文学作品语料库等。后者不在本课题研究范围之内,在此从略。在可比语料库构建和研究方面,肖忠华(2003)等已经做了有益的尝试,提出了一些构想,并取得初步的研究成果。

国内外研究界对已经建成的大型语料库及其相关研究多以理论探讨为主,关注文学和法学等领域,对科技领域涉及不多,尤其缺乏具有

高度说服力和推广价值的应用性研究,几乎没有系统关注过中国科技期刊国际化中的语言问题,尤其是在科技期刊的论文摘要这一特定领域进行的对比研究和翻译研究,并无专门设计的语料库,缺少规模化的实证性应用研究,这在某种程度上影响了科技期刊的国际化。根据国内外的学术现状,这类研究是急需的,中国科学家的研究成果应该得到应有的承认。

初步观察我国 1999—2009 年间对科技期刊摘要汉英翻译的研究,如于建平(1999)、段平与顾维萍(2002)和施蕴中(2003)等,整体仍倾向于翻译原则介绍、翻译方法指导、具体错误分析等,以经验讨论为多,缺乏客观数据及讨论,难以摆脱主观、局部和印象式的特点。而事实究竟如何,这些结论有无科学考据,是否引起了误读和滥用,是否影响了论文摘要英译文的语言质量,从而影响了期刊的国际化?上述问题都亟待投入人力和物力,进行实证研究并做出相关理论总结。

本书将基于语料库的实证研究,尝试对此做出部分回答,并探讨在以英文为主要交流工具的国际科学界,如何帮助中国科学家将成果在国际上发表。

1.2 本书的主要内容、基本思路、研究方法

1.2.1 主要内容

本研究采集真实的双语语料和可比语料(图 1.1),构建规模适中、通用型和定制型结合的语料库,为科技论文的汉英摘要翻译研究及对比研究奠定坚实的数据基础,为开展大规模汉英科技翻译提供丰富的材料,从语言层面促进科技期刊国际化。具体包括以下 4 个方面:

- 1) 语料库建设:包括可比语料库和平行语料库,容量为 100 万字词,进行以句子为主的对齐处理;
- 2) 对比研究和翻译研究:主要在科技期刊论文摘要的语篇结构、语言特征等层面进行;
- 3) 指出中国科技期刊论文摘要汉英翻译中存在的问题和改进策略,在实践中验证应用性研究结果;
- 4) 建设针对科技期刊论文摘要汉英翻译的计算机辅助人工翻译数据库。

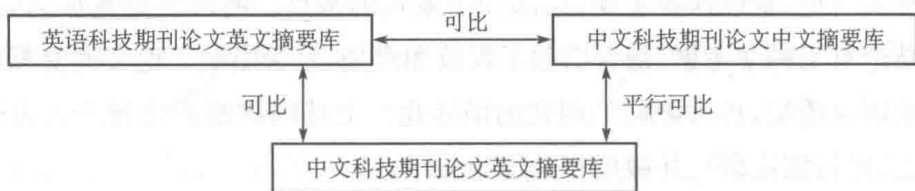


图 1.1 语料库类比及互动关系

从图 1.1 可以看出,几个语料库之间分别是可比、平行可比的关系,具有互动关系,以供各种研究使用。

1.2.2 基本思路和研究方法

本书在应用语言学、语料库语言学、对比语言学和翻译研究的框架下开展研究。选取 1991 年至 2015 年间出版的中文科技期刊的论文摘要及其英译本,构建一个原始的双语平行语料库,并选取同期欧美出版的同类科技期刊中收录的英文摘要,构建一个可比语料库。

语料库的结构和类型:语料子库名称、子库容量及所占比例如表 1.1 所示:

表 1.1 语料库构成

语料库名称	字数	占总库比例
汉语科技期刊论文摘要子库	约 40 万字	40%
科技期刊论文摘要汉英翻译子库	约 40 万字词	40%
英语科技期刊论文摘要子库	约 20 万字词	20%

1.2.3 研究所持的基本观点

本研究所持基本观点如下：

1) 标注完备的大型语料库可进行语言对比研究及翻译研究,这已成为当今研究的主流与范式(Baker, 1993;秦洪武,王克非,2004;Leech & Smith, 2006);

2) 各种电子资料的逐步完善及其可获取性,使得研究者能够根据研究目的建立小型语料库对翻译转换(Munday, 1998)、译者培训(Bowker, 1998)、语言差异(Maia, 1998)等进行考察。肖维青(2005)论证了自建小型语料库进行翻译研究的可行性,认为建立符合具体研究方向的小型语料库,在资料和数据充足的条件下,能够致力于多层面、多维度的翻译评估;

3) 通过本课题构建起来的科技论文摘要汉英翻译双语语料库和可比语料库,可以对汉英科技论文摘要进行系统的翻译研究和对比研究,并从实证角度得出具有说服力和应用价值的研究成果,有利于科技期刊的国际化进程;

4) 研究结果对目前的印象式、主观式、经验式、直觉式的汉英科技翻译论点是一次验证与修订,对汉英科技翻译的理论与实践具有借鉴作用,对我国科技期刊的国际化也具有促进作用。

1.3 全书内容简介及章节介绍

本书第一章为引言,大致叙述了研究的来龙去脉。余下章节共分为两部分。第一部分主要是理论阐述和回顾。其中第二章回顾并展望了语料库在对比研究、翻译研究、翻译技术研发中的作用,也论及了基于语料库的对比研究和翻译研究。第三章是中英科技论文摘要的特点综述。第四章是对科技论文摘要汉英翻译的回顾。第二部分是本书所进行的主体研究,从几个主要方面展开。其中第五章描述了研究所需的科技论文摘要语料库的建设,也展示了使用计算机工具对语料库进行的初步宏观分析。第六章对基于语料库的科技论文摘要的汉英对比研究做简要评述,并进行语篇层面的研究,主要以语步为研究对象,为后期研究做铺垫。第七章是对所建语料库进行的对比研究和翻译研究,主要从名词化角度进行。第八章和第九章是本书重点引申内容。第八章主要介绍基于科技论文摘要双语语料库而建成的翻译记忆数据库,并利用计算机辅助工具 Wordfast 进行新的翻译,以说明基于语料库的科技论文摘要汉英翻译在计算机工具的辅助下,具有较高的内容准确率和工作效率。第九章是在已建双语摘要库和术语库的基础上,利用 SDL Trados 来说明计算机辅助的科技论文摘要的翻译过程,以展示其对翻译实践教学的意义。第十章是总结。

本文作者希望以此提供一种研究思路和方法,供对语料库、科技论文翻译、摘要翻译、对比研究、翻译研究等感兴趣的高校师生和其他一般读者参考使用。

第二章 语料库方法在对比研究、 翻译研究和翻译技术研发中的应用^①

现代计算机技术日益发展,其存储能力也以惊人的速度倍增。自 20 世纪 80 年代以来,现代电子化存储的语料库开始大规模使用,并对相关领域产生了革命性影响。本章将简要论述其在对比研究、翻译研究和翻译技术研发中的应用。

2.1 引言:语料库与语料库语言学

语料库 (corpus) 指语言数据的集合,而语料库语言学 (corpus linguistics) 以语料库为基础对真实文本中的语言进行观察和分析,结合自然语言中的一些抽象规则,探讨某种语言及双语或多语之间的关系,并根据真实发生的语言现象得出比较可靠的分析结果。顾曰国(2003)认为语料库语言学的研究有两层主要含义:一是利用语料库对语言的某个方面进行研究,即它并不是一个新学科的名称,而仅仅反映了一个新的研究手段。二是依据语料库所反映出来的语言事实对现行语言学理论进行批判,提出新的观点或理论。前者正是本文的关注之处。

在 20 世纪 30 年代现代计算机出现之前,语料库研究通常为手工完

^① 本章内容曾发表在《外文研究》2014 年第 3 期,页码 98 - 103,108,收入本书时有修订。

成。计算机的出现使得大规模自动化处理语言成为可能。至于是否对语料进行标注,则可以根据研究目的分别对待。比如, Sinclair(1992)认为无须对语料进行过多处理,因为文本本身就能说明问题,但 Wallis(2007)却提倡标注,认为这样可以提高语言理解和处理的质量。后者的观点基于以 Quirk (1960:40 - 61)为首进行的英语用法调查(Survey of English Usage),这项调查是欧洲范围内进行的第一次有关语料的研究。

现代语料库发展历史并不长。1959年, Quirk 在伦敦大学着手英语用法调查,收集大量风格、题材各异的语料,作为对英国英语的口语和书面语进行系统描写的基础。大洋彼岸的美国,以 Francis 和 Kučera 为首的一批语言学家和计算机专家聚集在布朗大学,于 1961 年建成了当代最早的机读语料库——布朗语料库(Brown Corpus)。英国美国的这两个库的建立可以说是现代语料库语言学及其研究的开端。

经历了第一代计算机处理的语料库和第二代电子计算机处理的语料库两个发展阶段后(王建新,1998:53 - 59),进入 20 世纪 80 年代,现代科技的迅猛发展促使语料库及语料库语言学也进入快速发展期,各国、各语种语料库的库容动辄以百万字词甚至千万字词为计。根据建库目的和功能,语料库的分类多种多样,如单语/双语/多语语料库、双语/多语平行(或类比)语料库、书面文本语料库、口语语料库、历时语料库、共时语料库、动态语料库、静态语料库、各种文体/专业领域的语料库,等等。

在中国,1986年,上海交通大学建成了交通大学科技英语语料库(JDEST)。经过 30 多年的发展,我国语料库建设获得了长足进步,建成了一批可与国际同类语料库相比的专门语料库(如 JDEST)、学习者口语语料库(如文秋芳,王立非,梁茂成,2009),以及数量众多的各种平行语料库。

事实上,语言学的各个领域几乎都可以从语言使用的角度进行研究,语料库数据正是所使用语言的样本,它所提供的海量语料使研究者能够系统地对大量的文本语料进行审视,从而有可能发现一些以前从未发现

的语言事实(Sinclair, 1991)。因此,语料库作为研究手段可用于语言学研究的诸多领域,如词汇学、语法学、语言教学、文体学、对比语言学、翻译等。目前,语料库语言学的研究话题已经从词汇、语法和词典学扩展到语言教学、二语习得、翻译、自然语言处理、话语分析、认知语言,乃至抽象的理论语言学问题探索等广泛的领域(甄凤超,2010:36-41)。这些研究成果,根据所基于的语料库,可以分为单语研究、双语/多语对比研究、翻译研究、机器翻译研究、教学研究等多个类别。

作为一种研究方法,语料库在对比研究、翻译研究和翻译技术研发这三个与其显著相关的领域得到了尤为突出的使用,本文将简要介绍这些应用,分析它们在应用时所体现的共性和个性,并对其未来走势做一展望。

2.2 语料库与对比研究

对比(语言)研究是语言学研究的重要分支(许余龙,1992),但高度抽象是20世纪语言研究的总特点(顾曰国,2003),研究需要向具体化靠拢,语料库和语料库语言学在语言研究由高度抽象转向语言实际使用的过渡中起着十分重要的作用。这种作用体现在四个方面:一是提供真实语料,二是提供统计数据,三是验证现行的理论,四是构建新的理论(顾曰国,2003)。

对比语言研究与其他语言学学科一样,其研究日益成为基于数据的或数据驱动的,依靠对语料数据的统计来验证假设或揭示语言之间的相似和差异。这些数据从句法、模态、语篇、语法化、语用功能、修辞功能等(参见Marzo *et al.*, 2012)方面提供定性研究的证据,说明定量与定性研究之间是一种互补关系。利用语料库方法的语言对比研究取得了丰硕成果,以语言具体层次的比较为主,如词汇(连带语义韵)层面、短语层面、语

法层面等,同时辅以个案研究,说明具体问题,陈述研究发现及其教学应用上的意义。

通过概括总结大量的个案研究,可以进一步深入探讨语言背后的原因,如民族思维习惯和文化心态(许文胜,张柏然,2006:292-296; McNery & Xiao, 2010),也可以找到它们对其他领域的启发意义,如语言和翻译教学(赵世开,1990:145-147;卫乃兴,2006:50-54,2011:32-42)等。

但是,语料库语言学也有其局限性。因为语料库数据提供的只是所使用语言的事实例证,它本身无法对数据进行解释并对语料库数据进行推理。这些数据在辅助人们的语言直觉和内省判断的时候,离不开研究者本人的语言直觉和内省(冯志伟,2007),这也是对20世纪以高度主观抽象为特征的语言学研究的一种重要补充和返璞归真。

2.3 语料库与翻译研究

这个领域的研究以 Baker(1993:233-250)的“语料库语言学与翻译研究:启示和应用”为起始,各国学者将语料库研究方法迅速推广应用到各自语境下的翻译研究和理论阐述,使语料库翻译研究得到大力发展,逐步成为一种连贯、综合和丰富的研究范式,并成为描述翻译研究中的一种新范式(Laviosa, 1998:474-479, 2002;王克非,2006:8-9;吴昂、黄立波,2006;王克非、黄立波,2008)。

国内的语料库翻译研究起步并不晚。早期规模较大的双语对应语料库是北京外国语大学开发研制的3000万字/词的“通用汉英对应语料库”。此后,基于语料库的翻译研究一直吸引着国内广大学者的关注,研究成果主要集中在翻译语料库的建设及研究,翻译普遍性/共性研究,翻译过程研究和翻译文体研究。

就语料库对翻译研究的贡献, Tymoczko(1998: 652 - 659)的观点颇具代表性。她认为,语料库方法是一种强有力的工具,是翻译学保持活力和动力的核心途径。它可以使人们充分认识到翻译转换中产生的差异,这种差异性经归纳总结后发现的规律能起到形成规范的作用。

近年来通过大量的实践,基于语料库的翻译研究促使翻译研究方法从规约式转向描述式,翻译研究从最初的语言学和对比语言学的角度,发展到后来从社会、政治、文化等多角度来解读译文的内外特点,这标志着翻译研究从单纯的描述研究向描述、阐释研究的发展。黄忠廉(2009: 101)认为,翻译研究要在理论和实践层面处理和解决跨语言、跨文化交际问题。翻译语料库是译论研究现代化的有效手段,其定量研究可催生译论观点的萌发,还可对译论观点证实或证伪。

对比后可以发现,国内语料库实证研究的范围与国外相比还是相对狭窄(刘康龙、穆雷,2006: 59 - 64)。自 Baker(2000: 241 - 266)提出用语料库方法研究译者文体以来,国外目前基于语料库的译者文体研究不仅有对译者偏爱的语言模式的研究,还有对构成译者文体的宏观因素的量化分析,如翻译选材、翻译策略、叙事结构、意识形态,等等。但国内基于语料库与文体相关的研究还不是很多,对翻译过程的研究仍显不足(周小玲、蒋坚松,2010: 54 - 58)。

2.4 语料库与翻译技术研发

翻译是从一种语言转换成另一种语言的活动,目的是求得原语与译入语最大限度的对等对应,人工翻译和机器翻译都是如此。机器翻译是按照指定的程序让计算机进行自动化的对比操作,从而自动产生译文(王挺等,1996: 52 - 55)。但由于机器翻译技术研发遇到各种困难,全自动、高质量的自动翻译系统一直未能实现。目前,机器翻译包含在翻译技术