

Corpus-based Studies of English Translation

基于语料库的英语翻译研究



张付花
著



中国水利水电出版社
www.waterpub.com.cn

基于语料库的英语翻译研究

张付花
著



中国水利水电出版社
www.waterpub.com.cn

·北京·

内 容 提 要

本书以英语翻译教学理论为基础,从英语语言特征、英汉搭配特征、文化词汇及习语翻译、书套语及幽默文本翻译、实用文体翻译、记者招待会口译、机器翻译等方面进行了详尽的研究。

本书语言简洁凝练、结构明了、实用性强,可作为语言学及应用语言学专业的教师、研究生、语料库研究者、语篇研究者的参考资料。

图书在版编目(CIP)数据

基于语料库的英语翻译研究/张付花著. —北京:
中国水利水电出版社, 2017. 6
ISBN 978-7-5170-5592-1

I. ①基… II. ①张… III. ①英语—翻译—研究
IV. ①H315.9

中国版本图书馆 CIP 数据核字(2017)第 167432 号

书 名	基于语料库的英语翻译研究 JI YU YULIAOKU DE YINGYU FANYI YANJIU
作 者	张付花 著
出版发行	中国水利水电出版社 (北京市海淀区玉渊潭南路 1 号 D 座 100038) 网址: www. waterpub. com. cn E-mail: sales@ waterpub. com. cn 电话: (010)68367658(营销中心)
经 售	北京科水图书销售中心(零售) 电话: (010)88383994、63202643、68545874 全国各地新华书店和相关出版物销售网点
排 版	北京亚吉飞数码科技有限公司
印 刷	三河市天润建兴印务有限公司
规 格	170mm×240mm 16 开本 16.75 印张 217 千字
版 次	2017 年 11 月第 1 版 2017 年 11 月第 1 次印刷
印 数	0001—2000 册
定 价	78.00 元

凡购买我社图书,如有缺页、倒页、脱页的,本社营销中心负责调换

版权所有·侵权必究

前 言

语料库语言学的迅速发展给语言研究带来了重大的影响,翻译学也不例外。20世纪50年代,翻译学研究强调规范研究,到了20世纪80年代,开始转向描写研究,这一研究范式的转变与语料库语言学有很大的关系。再加上20世纪90年代多种语料库的不断发展,语料库翻译学应运而生。语料库翻译学的研究对象为真实的双语语料或翻译语料。语料库翻译学建立在语料库的基础上,采用数据统计与理论分析的研究方法,对翻译本质、过程以及相关现象进行分析与研究。语料库翻译学充分利用其技术优势,对翻译共性、翻译规范、译者风格、口译规范、机器翻译等问题展开深入探讨。采用语料库方法来研究翻译增加了译学研究的科学性与客观性。基于此,作者在对语料库翻译学相关资料进行梳理的基础上,结合自己近年来对语料库翻译学以及英语翻译研究的心得与体会,精心策划并撰写了《基于语料库的英语翻译研究》一书,希望为语料库与翻译学研究的发展贡献绵薄之力。

本书共有十章。第一章对语料库进行了概述,内容涉及语料库的内涵与特点、语料库语言学的发展与前景以及基于语料库的研究方法。第二章对基于语料库的英语翻译研究进行了综述,主要讨论了基于语料库的英语翻译的内涵、起源和发展以及译学研究语料库的基本类型。第三章从英语词汇与语法两个层面研究了基于语料库的英语语言特征。第四章分别介绍了基于语料库的英语、汉语搭配研究,并对二者进行了对比分析。第五章对基于语料库的文化词汇和习语翻译进行了探讨。第六章重点研究了基于语料库的说书套语和幽默文本翻译研究。第七章将视角

转向了译者,对基于语料的译者风格展开了讨论,以《水浒传》四英译本为例进行了详细分析。第八章研究了基于语料库的实用文体翻译,主要涉及基于语料库的电子商务翻译、中医英语翻译两方面的内容。第九章为基于语料库的记者招待会口译研究,内容包括基于语料库的记者招待会口译语言特征、口译规范以及口译策略。第十章涉及基于语料库的机器翻译研究,主要讲述了机器翻译的现状与前景以及机器翻译系统的评价机制。

本书从语料库视角对英语翻译进行了系统的研究,借助语料库方法,对很多实际存在的翻译事实或现象进行了定性与定量的分析与总结,内容充实,论述深入浅出,理论研究与实际运用有机结合,希望本书可以帮助读者更全面地了解语料库翻译学,也希望为翻译研究提供新的视野,同时为语料库研究带来启示。

在本书的写作过程中,引用了很多专家、学者的学术见解以及相关论述,在此向他们表示诚挚的谢意。

由于作者写作水平有限,书中难免有疏误之处,恳请广大读者不吝指正。

作 者
2017年4月

目 录

前 言

第一章 语料库概述	1
第一节 语料库的内涵与特点	1
第二节 语料库语言学的发展与前景	7
第三节 基于语料库的研究方法	18
第二章 基于语料库的英语翻译研究综述	26
第一节 基于语料库的英语翻译的内涵	26
第二节 基于语料库的英语翻译的起源与发展	34
第三节 译学研究语料库的基本类型	43
第三章 基于语料库的英语语言特征研究	50
第一节 基于语料库的英语词汇特征研究	50
第二节 基于语料库的英语语法特征研究	70
第四章 基于语料库的英汉搭配特征研究	76
第一节 基于语料库的英语搭配特征研究	76
第二节 基于语料库的汉语搭配特征研究	88
第三节 基于语料库的英汉搭配特征对比分析	100
第五章 基于语料库的文化词汇、习语翻译研究	103
第一节 基于语料库的文化词汇翻译研究	103
第二节 基于语料库的习语翻译研究	115
第六章 基于语料库的说书套语、幽默文本翻译研究	130
第一节 基于语料库的说书套语翻译研究	130
第二节 基于语料库的幽默文本翻译研究	146

第七章 基于语料库的译者风格研究——以《水浒传》	
四英译本为例	159
第一节 《水浒传》英译研究现状	159
第二节 《水浒传》四英译本翻译特征宏观对比研究 ..	165
第三节 《水浒传》四英译本翻译特征微观对比研究 ..	177
第八章 基于语料库的实用文体翻译研究	183
第一节 基于语料库的电子商务翻译研究	183
第二节 基于语料库的中医英语翻译研究	199
第九章 基于语料库的记者招待会口译研究	212
第一节 基于语料库的记者招待会口译语言	
特征研究	212
第二节 基于语料库的记者招待会口译规范研究	225
第三节 基于语料库的记者招待会口译策略研究	230
第十章 基于语料库的机器翻译研究	233
第一节 语料库与机器翻译	233
第二节 机器翻译系统的评价机制	238
第三节 机器翻译的现状与前景	250
参考文献	256

第一章 语料库概述

基于语料库和语料库语言学的研究对语言研究和英语翻译研究产生了积极的作用和深远的影响,语料库和语料库语言学已经逐渐步入人们的视野,并成为当前语言教育界和翻译界的一大热门话题。同时,语料库语言学凭借其自身的优势对许多语言研究领域产生了越来越重要的影响,因而其在研究中也得以发展并呈现出良好的发展前景。那么,究竟何为语料库?语料库语言学究竟发展得如何呢?本章将结合语料库的内涵与特点、语料库语言学的发展与背景以及基于语料库的研究方法进行研究和分析,以期对语料库及其相关内容有深入、系统的认识。

第一节 语料库的内涵与特点

随着现代计算机技术的普及和发展,人们开始频繁地接触语料库,并对语料库使用价值的认识逐渐加深。因而,系统地对语料库的内涵和特点进行分析就非常必要。

一、语料库的内涵

(一)何为语料库

虽然关于语料库的研究已经有了不短的历史,但是关于语料库,至今并没有一个普遍公认的界定。下面结合一些国内外比较

有代表性的观点进行介绍。

1. 国外学者的观点

根据辛克莱(Sinclair)的观点,语料库是依照明确的语言学标准选择并进行排序的语言运用材料的汇集,语料库主要是用来作为语言的样本存在的。

根据利奇(Leech)的观点,语料库是一个由大量在真实情况下使用的语言信息集合成的专门用于研究使用的资料库。

根据克里斯特尔(Crystal)的观点,语料库是语言材料的收集,资料来源不仅可以以书面语篇的形式存在,而且还可以以话语的记录脚本形式存在。它可以被用作语言描述的起点或者用作对有关语言假设的方法进行验证。

2. 国内学者与权威出版的观点

国内一些学者和权威出版也对语料库提出过自己的见解和看法。

李文认为,语料库是按照相应的语言学原则,采取随机抽样的方法,收集一些自然出现的、连续的语言,运用文本或者话语片段的形式进而建成的具有一定容量的大型电子文库。

彭勤认为,语料库(corpus)是利用数据库技术把大量的口头和书面交际的真实语言材料以电子文本形式存储,并通过计算机自动处理、检索以及进行各种统计分析的大型语言资料库。^①

根据黄昌宁、李娟子的观点,语料库其实就是用来存放语言材料的仓库或者数据库。

从传统观点上来看,语言学家用语料库这一术语表示可用作语言研究的、大量出现的、基础的语言数据。这些语料库通常是由书面语或者口语的样本组成,并且通常用来代表一种特定语言或者语言变体。计算机还没出现时,一些研究者尤其是词典编纂

^① 彭勤. 语料库概述及在英语教学中的应用[J]. 科技信息, 2010, (16): 123.

者,也涉及一些语料库的内容,之所以没有得到密切关注,主要是因为规模还比较小,所涉及范围也比较狭窄,因而很难在学术界形成气候。近四十年来,语料库这一术语通常是以电子形式保存的语言材料,并在语言研究和语言工程方面得到了广泛的运用。伴随着计算机功效的飞速增长,语料库在规模、多样性以及使用便利化方面都发生了剧烈变化。与此同时,人们为了对语料库所拥有的信息进行存取和加工,就开发了一些专门化的、大量的专用软件。这也就使得计算机语料库很快成为一种用于语言研究的资源。

从当前来看,世界上已经建立了诸多规模很大的语料库,还有一部分属于国家级的,有一部分属于大学与词典出版商联合研发的。此外,随着个人电脑的迅猛发展,用于存储数据的硬磁盘的造价持续下降,研究者个人也开始建立了一些贴合自己研究兴趣的小型语料库。

在《语言学 and 语音学词典》中,也有关于语料库的定义,其认为语料库是语言资料的收集,通常是经过加工后带有语言学信息标注的语料文本,可以被用作语言描述的起点或验证语言假设的方式。

(二) 语料库的类型

按照不同的划分标准,可以对语料库进行不同的分类。

依据语言所涉及的类型,可以将语料库分为三大类型:(1)单语语料库;(2)双语平行语料库;(3)多语语料库。

依据语料选取的时间进行划分,可将语料库分为两大类型:(1)历时语料库;(2)共时语料库。

依据语言所涉及的题材,可将语料库分为两大类型:(1)普通语料库;(2)专门用途语料库。普通语料库的设计与创建通常需要考虑对语料库的各种可能的应用需求,并且是以大多数人都能够接受的语言理论作为指导的。通常来说,普通语料库基于其是建立在广泛收集的基础上的,因而有比较大的库容量。专门

用途语料库则是依据某一领域的具体需要进行设计的语料库,因而其选材也就可能仅限于这一领域,语料标注原则在某些情况下也是专门为某一特殊化的应用而确定的。

依据语料来源的不同,可将语料库分为两大类型:(1)口语语料库;(2)书面语语料库。

依据是否被标注,可将语料库分为三大类型:(1)生语料库;(2)原始语料库;(3)标注语料库。其中标注后的语料库更加便于检索。

依据语料库的动态更新程度的不同,可将语料库分为两大类型:(1)参考语料库;(2)监控语料库。参考语料库从原则上来讲根本不进行动态的更新。相比之下,监控语料库通常需要进行持续不断的动态更新。

依据语料库处理方式的不同,可将语料库分为两大类型:(1)自然文本语料库(没有经过标注或者赋码处理的);(2)赋码语料库。当然,也有一些研究者认为,应注意区分语料库和文本数据库。这些研究者认为,语料库不仅是原始语料的集合,而且是一种标注语法、一定结构、语义、语音、语用等一种或者多种语言信息的语料集合,这是语料库与一般文本库的区别所在。从这种意义上进行分析,就不能将自然文本语料库称为语料库。

(三) 语料库的标注问题

有关语料库的标注问题,在学术界存在着诸多看法。在汉语书面文本中,词和词之间没有空白,这非常不利于计算机的处理。因而,汉语书面文本的语料库通常都做切词和词性标注。有一部分学者认为,应进行语料库标注,因为标注过的语料库在研究和开发时更为方便,使用时可被重复利用,功能上更加多样,并且分析时更为清晰。

还有一部分学者对语料库的标注持反对意见。他们之所以不同意对语料库进行标注,一方面是因为他们认为经过标注后的语料库失去了客观性,所得到的标注语料库缺乏纯粹性,带有标

注者对语言的主观认识。另一方面,手工标注的语料库准确性虽然高,但是缺乏一致性;自动或者半自动的标注一致性虽然高,但是准确性会比较差。因而,在语料库的标注问题方面至今尚未达成共识。

二、语料库的特点

(一)语料库的真实性

语料库中的语言材料通常都是作为人们在具体使用中的语言材料,因而语料库具有真实性的特点,这一特点也决定了语料库语言学的研究成果具有真实性和可靠性这两大特性和优势。不仅如此,语料库的真实性还使语料库的研究对语言教学的研究有着积极的作用和意义。根据克拉申(Krashen)的可理解性输入假设和斯温(Swain)的可理解性输入假设,英语学习者为了实现可理解的产出这一目的,并达到知识的内化,往往也需要接触大量真实的目标语或者使其身处目标语的语境中。但是,能够在目标语环境中学习语言对于大部分英语以及外语学习者而言,通常都很难实现,只有通过语料库这一途径才能够使学习者接触更多真实的目标语。当然,这种类型的输入通常是有选择性地输入,外语教师通常应能够对语料库中的丰富的、具有代表性目的的语料以及相关知识加以充分利用,并结合教学的实际、知识的深度和广度等,引导学生按照语言真实使用的情况来运用语言,力求实现最优化的输入。

(二)语料库的代表性

在探讨语料库时,通常很难回避语料库样本的有限性,因而在进行语料的选材时,应尽可能地追求语料的代表性。尽量在样本语料有限的情况下,使其尽可能地反映无限真实语言现象的特征。语料库的代表性这一特点具体体现在以下三个方面。

其一,要求样本应源于与语言文字规范相符合的真实的语言材料。

其二,要求样本应源于正在“使用中”的语言材料,可涉及各种环境下的、规范的或者不规范的语言应用。

其三,语料通常应体现时代性的特点,并且能够很好地反映语言的发展变化以及当代语言的生活规律。

计算机只有对具有代表性的语料库进行处理,才能了解更加真实的语言应用规律,才不仅能对规范的语言进行理解和处理,而且还能处理那些不规范但是被广泛接受的语言进行处理,当然也能处理那些包含些许错误的语言。

(三)语料库的结构性

语料库是有针对性、有目的地收集语料的集合,而不是任意的语言材料的堆积,因而这就要求语料库应具有特定的结构。在当下计算机的技术条件已经普及的大背景下,语料库必须是以电子文本的形式存在着的、计算机可读的语料集合。语料库的逻辑结构设计应能够对语料库子库的组成情况加以确定,并对语料库中语料记录的码、元数据项、取值范围、每个数据项的数据类型以及数据宽度等进行定义。这些都很好地体现了语料库的结构性特点。

(四)语料库的平衡性

平衡因子是对语料库的代表性这一特征产生影响的关键特征。在平衡的语料库中,语料库为了力求实现平衡,最为首要的事项就是确定语料的平衡因子。这些对语料库的平衡产生影响的因素有很多种,具体如表 1-1 所示。

表 1-1 影响语料平衡的因素

序号	因素	序号	因素
1	学科	6	登载语料的媒体
2	年代	7	文化背景
3	地域	8	语料的用途(如私信、广告、公函等)

续表

序号	因素	序号	因素
4	文体	9	阅历
5	使用者的年龄	10	性别

通常情况下,应根据实际需要,也就是平衡语料库的用途时选取其中的一个或者几个重要指标作为平衡因子,比较常用的指标有年代、地域、学科以及文体等。

(五) 语料库的规模性

大规模的语料库对于语言研究尤其是计算机语言学的研究有着至关重要的作用。然而,随着语料库的增大,垃圾语料无形中所带来的对垃圾进行统计的问题也日益严重。并且,当语料库达到相应的规模后,其功能也会伴随着规模的增大而同步增长。此时,我们就需要根据实际需要来决定语料库的规模,语料库规模的大小应根据能否满足其现实需要而决定。

(六) 语料库的可机读性

对于现代语料库来说,必须是以电子文本形式存在并能够被计算机程序检索和处理的,只有这样才能运用在计算机技术支持下的语言学研究。机读化的语料库具有以下两大优势:一是更加方便检索。使用检索软件,在上亿万字符中调阅某一个或某几个词项仅需数秒钟。二是更加方便增补一些新的语言信息。语料库大多是处在相对稳定的状态下的,并且是处在动态的更新与变化中的。

第二节 语料库语言学的发展与前景

语料库语言学是一种非常重要的语言研究的方法,它是在计

计算机科学迅猛发展的大背景下茁长成长起来的。如果追溯其发展历史,可溯及 20 世纪甚至更为久远。在其发展过程中,由于在处理 and 解决语言问题时发挥了独特的作用,因而具有良好的发展前景。在此对语料库语言学的发展与前景进行探讨和分析。

一、语料库语言学的发展

以乔姆斯基(N. Chomsky)转换生成语法的兴衰史为参照点,可以将语料库语言学的发展历史分为以下三个阶段。

(一)早期的语料库语言学

20 世纪 50 年代以前这一时期属于语料库语言学的早期阶段。这一时期是关于乔姆斯基所提出的转换生成语法理论之前的所有基于语言材料的语言研究。在 20 世纪 50 年代,语料库在语言研究中曾经得到了广泛的使用,具体体现在以下五个方面。

1. 语言习得

语言习得是最早的普遍将语料作为研究方法的一个领域。19 世纪 70 年代,在欧洲的一些国家兴起了儿童语言习得研究的第一个高潮。当时,关于儿童语言习得研究的很多研究素材都源于父母对其子女话语发展的观察日记。据了解,他们在进行研究时将这些日记当作原始资料,这些资料不仅是当时普莱尔(Preyer)和斯特恩(Stern)等人提出的理论假说的依据,而且到目前为止,这仍然是一些学者在进行研究时所运用的材料。自 20 世纪 30 年代以来,语言学家和心理语言学家提出了诸多关于儿童在其各年龄段的语言发展模式,这些语言发展模式大多也都是建立在以往对儿童自然话语所进行的大量观察、研究的基础之上的。

2. 方言学

方言学一经产生,就和语料结下了不解之缘。在西方国家,

方言学脱胎于 19 世纪的历史比较语言学,方言学的最初兴趣是采用直接法所获得的关于单音不同分布的事实对方言地图进行描绘。一些从事方言研究的人员一开始是手持笔记本,后来是手提录音机,对所听到的任何方言材料进行记录。这其实是一种广为使用的采样方式,截至目前仍在使用。这种采样方法对研究方言词汇的分布有着很强的价值。我国也有关于方言学的研究,国内采用语料研究的方法可追溯到周秦。根据南朝应劭《同俗同义序》的记载,“周秦常以岁八月遣輶轩之使,采异代方言。”

我国汉语方言学的第一部著作《方言》其实就是这一方法的产物。根据相关记载,扬雄十分喜欢方言,他经常借助考廉(相当于后代的举人)与士兵集中在首都方便,普遍走访并积累材料,长久地坚持编撰和整理,历经 27 年的努力付出,最终写成了《有輶轩使者绝代语释别国方言》。

3. 语言教学

弗里斯(Fries)、特拉弗(Traver)和邦格(Bonger)是运用语料对外语教学法进行研究的语言学家。在 20 世纪前 50 年中,语料库同外语教学存在着密切的联系。在外语教学中所使用的词汇表,大多是直接从语料中统计出来的。这些词汇表对外语学习过程的控制有着非常重要的作用。

4. 句法和语义

还有一部分语言学家运用语料库对语言的描述进行研究。例如,弗里斯在对语言进行研究时,以对语料库的调查为基础,建立了英语的描写语法。弗里斯基于这项工作的研究要比 20 世纪 80 年代后期著名语言学家夸克(Quirk)等借助语料库编撰的《英语语法大全》早 30 年。

5. 音系研究

借助自然语料开展的音系研究,在西方国家最早要数结构主

义语言学家,如萨皮尔(E. Spair)等人。这些结构主义语言学家重视“野外工作”,强调获取语料的自然性和对语料进行分析的客观性。后来,这些相关的研究都在语料语言学的相关研究中得到了很好的继承与发展。

到了20世纪50年代中前期,出现了实证主义和行为主义。受到这些思潮的影响,基于经验主义的研究在语言研究中逐渐占据了主导型地位。在这种形势下,对语料重视逐渐成为当时的热点话题之一。尤其在美国,以哈里斯(Harris)等人为主要代表人物的后布隆菲尔德结构主义语言学家,将语料看作语言学唯一的研究对象。根据这一学派的观点,直觉的证据位于次要地位,通常是靠不住的,应摒弃。

(二) 乔姆斯基的转换生成语法时期

乔姆斯基在1957年的《句法理论》和其后期的诸多论著的发表,使早期语料库语言学的发展发生了根本的改变。随之,语言学研究的主流方法也慢慢地从经验主义转向了理性主义。在此期间,笛卡尔的理性主义居于主导地位,这无形中使经验主义失去了立足之地,并且被视为经验主义产物的各种语料库也就自然而然地被全部否定了。

根据理性主义的研究方法,人所具备的大部分的语言知识是由遗传因素决定的,并且是与生俱来的。但是,同理性主义持对峙观点的经验主义则认为,人的知识仅能通过感官进行输入,并通过一些简单的联想与一般化的操作得以实现。人并不是一开始就有一套相关的语言原则和处理方法。基于乔姆斯基的内在语言和语言能力说在语言学中已经得到了普遍的接受和认可。在20世纪60年代到70年代这长达20年的时期中,理性主义方法事实上主宰了欧美等诸多国家的语言研究。乔姆斯基和其转换生成语法学派对早期语料库研究方法进行批判的观点主要有以下两大论点。