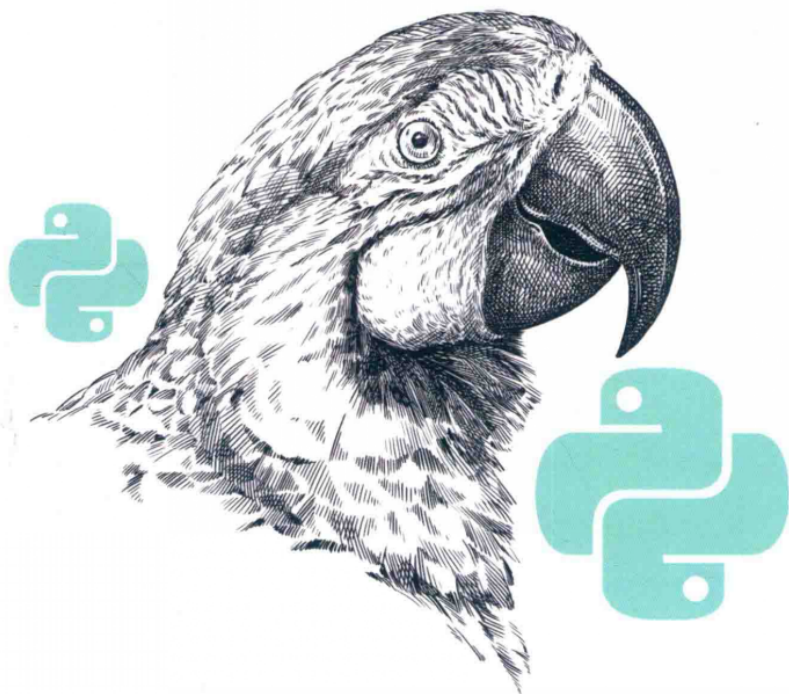


语言数字人文系列

基于Python的 语料库翻译

数据分析与理论探索

管新潮 陆晓蕾◎著



上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书旨在依托 Python 编程语言就语料库翻译这一方向展开理论深化与技术创新方面的融合式探索,其内容不仅涉及学界的语料库翻译学,也涉及业界的语料库与翻译。全书分为九章内容:语料库翻译概述、多变量协同效应、译文可读性、翻译对等的短语特征、翻译知识库的构建与应用、语言结构识别与译后编辑、翻译质量评估、文本情感对比分析以及双语工具开发与应用。第 1 章是对语料库翻译的学界研究和业界应用的综合性概述与分析;第 2~8 章均以翻译理论为原点,探索翻译理论与当下技术实现融合的可能性。第 9 章呈现了多种双语 Python 技术解决方案。全书既有理论描述,也述及实践应用,强调了理论与实践并驾齐驱的可行性。本书适合高等院校翻译学等专业的师生以及从事语言服务或翻译实践活动的爱好者阅读使用。

图书在版编目(CIP)数据

基于 Python 的语料库翻译:数据分析与理论探索/
管新潮,陆晓蕾著. —上海:上海交通大学出版社,
2022. 8

ISBN 978-7-313-27035-1

I. ①基… II. ①管…②陆… III. ①软件工具—程
序设计—应用—语料库—翻译学—研究 IV. ①H059-39

中国版本图书馆 CIP 数据核字(2022)第 110810 号

基于 Python 的语料库翻译——数据分析与理论探索

JIJU Python DE YULIAOKU FANYI——SHUJU FENXI YU LILUN TANSUO

著 者:管新潮 陆晓蕾

出版发行:上海交通大学出版社

邮政编码:200030

印 制:上海万卷印刷股份有限公司

开 本:710mm×1000mm 1/16

字 数:226 千字

版 次:2022 年 8 月第 1 版

书 号:ISBN 978-7-313-27035-1

定 价:78.00 元

地 址:上海市番禺路 951 号

电 话:021-64071208

经 销:全国新华书店

印 张:14.75

印 次:2022 年 8 月第 1 次印刷

版权所有 侵权必究

告读者:如发现本书有印装质量问题请与印刷厂质量科联系

联系电话:021-56928178

总序 ▶▶▶

数字人文是将数字技术运用于传统人文学科研究与教学的新兴交叉研究领域。20世纪50年代,意大利耶稣会牧师罗伯托·布萨神父在工程师的帮助下将一千余万词拉丁语著作做索引标注,此项标注工作被普遍看作数字人文的发端。20世纪六七十年代,研究者们接受了具有跨学科属性的实证社会科学,这也标志着注重思辨的传统人文研究与注重实证的硬科学开始结合。20世纪后半叶则见证了数字人文研究的飞速发展,数字人文领域的两本期刊《计算机与人文》(*Computers and the Humanities*)和《文学与语言计算》(*Literary and Linguistic Computing*)相继创刊,标志着“人文”研究离不开数字“计算”已成为业界共识,“人文计算”或数字人文研究渐成燎原之势。21世纪以来,数字人文进入2.0阶段,研究者开始突破传统学科界限,“生产、管理和交互‘天生数字化’的知识”。近年来,国内的数字人文研究也方兴未艾。2019年被称作国内“数字人文元年”,首份专业期刊《数字人文》创刊,学术会议交流等活动在各地举行,多所大学成立数字人文研究中心、开设数字人文课程,数字人文在国内俨然呈现井喷发展态势。

语言学研究者,特别是计算语言学、语料库语言学等领域的研究者,似乎与数字人文研究有着天然的亲近感,他们积极参与数字人文研究。究其原因,计算语言学或语料库语言学研究与数字人文研究都基于问题和数据驱动,大多数数字人文研究也与语言或文本分析与处理紧密相关,因此很多数字人文学者已然将语言学相关研究视作数字人文研究的有机组成部分。

在此大背景下,语言数字人文应运而生。语言数字人文是采用数字技术与方法以提出或解决语言学及其相关问题的研究领域。针对上述定义,我们

至少可以明确三点。①语言数字人文研究聚焦于语言学及其相关问题研究,即传统语言学问题的研究和语言相关的其他学科问题研究;②语言数字人文研究须采用数字技术与方法来进行研究;③语言数字人文研究不但采用数字技术与方法来解决已有问题,更是通过数字技术与方法,突破语言学的学科藩篱或界限,发现或提出新的问题。

语言数字人文丛书旨在推动国内语言数字人文的蓬勃发展,并助力数字人文领域相关研究。丛书首批将出版语言数字人文理论建设、语言数字人文方法论等基础论著,后续将陆续出版语言数字人文与文学、新闻传播学、心理学、计算机科学、信息科学、图书馆科学等交叉应用研究相关论著。我们相信,语言数字人文作为数字人文的分支研究方向,在研究内容、研究方法、知识创新等诸多方面具有新文科属性和特征,语言数字人文及本丛书完全有理由助力新文科建设与发展。

是为总序。

雷蕾

2022年6月26日

前言 ▶▶▶

起始于2017年的Python语言与翻译应用写作体验,随着本书的完成应该可以告一段落。从初级版的《语料库与Python应用》到系统版的《Python语言数据分析》,再到专业版的《基于Python的语料库翻译》,三种图书的顺序出版构成了我学习与工作实践的心路历程——所得到的乐趣远多于所付出的辛苦。本书的写作体验使我再一次明白,语料库翻译技术的发展永无止境,但相关技术必须融入语言数据或语料库的翻译概念之中,否则只能视其为邻家领域的技术。

本书的写作初心源于对翻译的喜爱。既然Python能够融入语言学的各类概念之中,而且已呈现出良好的发展势头,自然也能够融入翻译学和翻译实践之中。于是,本着这一理念开启了本书的写作之旅。可能有人会提出疑问:机器翻译就是编程实现的,你的Python翻译编程是什么啊?其实,有过机器翻译应用体验的人都会有这样的感受——机器翻译已经步入可用阶段,但其译文质量在很多应用场景下仍有待商榷,尤其是与高水平人工翻译质量相比照。我想说的是,我的Python翻译编程主要集中于机器翻译无法实现的诸多翻译实践与应用场景,同时也聚焦于语料库翻译的研究领域,更确切地说,是语料库翻译理论与翻译实践的Python编程探索。一个个独立的Python第三方库已经为各种实际应用场景提供了诸多可能性,其应用价值毋庸置疑。但在语料库翻译领域场景下,其应用仍须注意相应技术可能存在的各种不确定性:一是每一种第三方库仅仅展现自身的优点,很少谈及不足之处;二是特定的技术都有其特定的应用场景,并不一定具有普适性;三是第三方库的组合应用难度更大。

从某种意义上说,本书的写作也是对学界语料库大咖秦洪武老师和许家

金老师的一次个人响应。语料库语言学和语料库翻译事业的发展生机勃勃,令人振奋,但可供使用的语料库技术却并非有想象中的那么多选项。仅仅依靠纯技术人员的技术“施舍”无法满足语料库研究和实践应用的需要,再说纯粹的技术若不涉及语料库概念也无其用武之地。其实,有许多知名大企业拥有很多语料库技术,但企业的逐利性使相关技术的公开成为不可能。学界不能因此气馁,这一问题的解决必须依靠语料库学界的自身力量。令人欣喜的是,这一尝试行动已经体现在上海外国语大学胡开宝教授领衔的语言数据团队等身上。如若本书的尝试也能在一定程度上促使上述问题的解决,那就深感欣慰,荣幸之至。

本书的写作所能产生的体验感与前两本书的多有不同。一是本书的写作时间近乎延续一年,这是文献综述和案例编程费时费力所致;二是写作本书所带来的乐趣远超前两本书,这是模式创新给人的体验,尤其是感触到能以编程方式转化之前所积累的翻译研究知识和翻译实践经验;三是 Python 技术的介入会给语料库翻译的发展带去更多可能性,使翻译研究更具个性化。

鉴于上述感受与体验,备感本书所涉内容有其三个独特而具有新意的特点:

(1) 变量设置的多样性。语料库翻译的相关研究通常会以不同的变量加以表征,如标准类符形符比、平均句长、词簇等,并依据变量数值的大小对语言特征进行相应判别。这一变量处理模式迄今有效,但所面临的问题是究竟需要多少变量才能恰如其分地表征相应的语言特征?毫无疑问,以一个变量或少数变量所表征的语言特征只能是局部特征,而非总体特征。这是一个需要多方参与探讨的专业性和逻辑性问题。其实,更多的变量设置也会带来另一个问题,即变量之间的关联性,其协同效果如何。若两个变量的作用彼此相互抵消,那么将两者设置在同一个维度之内可能并非明智之举。因此,有必要采用统计学方法进行变量的有效性检验,以此确定合理的变量个数,使其能够有效发挥表征作用。语料库翻译研究领域是如此,语料库翻译实践领域也同样需要有效的变量设置。以译后编辑为例,如何发现特定语言结构模式下的机翻译文总是会出错这一现象,其明显需要新的变量,如虚词密度等。变量设置的多样性业已呈现,希望本书在此方面的尝试能够激发出现更多可有效表征的变量。

(2) 技术应用的组合性。与经典的语料库工具相比,新技术所能提供的

技术种类已越来越多,这不仅是技术发展的必然,也是多样性变量设置的需要。将不同技术组合在一起,服务于特定目的,须遵循一定的逻辑。以 FastText 或 Word2Vec 模型应用为例,模型本身并不复杂,其已封装,所能调用的代码只有两行——模型加载和语料训练。训练语料之前如何将原始语料转换成模型所需格式却是关键,因为原始语料的格式可能互不相同,其所需清洗方式和清洗程度也彼此有别。是否需要词形还原?采用何种标注器进行词性标注?是转换为句子列表还是段落列表?用于表征变量的模型须加载多少?等等。所有的这一切都以不同模型或第三方库的组合为前提,因此对不同技术优缺点的了解与把握显得至关重要,过程之中的数据传递与迁移模式亦发挥决定作用。技术的有效组合在于厘清理论思路,精简算法,提高效率。

(3) 双语文本的对比性。这是本书的一大特点,即以双语平行或双语可比或双语混合语料作为考察对象,实现双语语境下的 Python 应用。有些模型看似彼此之间并无明显关联性或关联性不强,但经过双语平行语料的测试,发现两类模型之间的确存在可比性。这一作用的体现极大地拓展 Python 第三方库或模型的可应用性,如汉语情感词典与英文情感极性工具之间,spaCy 的中文与英文语言模型之间,同一个语言模型应用于不同语种之间,等等。与翻译相关的双语技术工具的开发使翻译技术从“拿来主义”模式转化为极具个性的技术创新模式,使翻译专业的技术学习与应用不再受限于固化模式,使双语文本的彼此对比更具可行性和多样性。

上述三大特点的前后顺序或交叉呈现是构成本书章节设置的前提。各章节虽仅以序号区分,但根据潜在的内容显示亦可将全书划分为三大部分:第一部分即第 1 章,介绍语料库翻译的前因后果以及学界业界之间的语料库恩怨情仇;第二部分起始于第 2 章,结束于第 8 章,涉及具体的双语对比,如第 2 章的不同变量设置、第 3 章的译文可读性、第 4 章的翻译对等概念、第 5 章的翻译知识库、第 6 章的语言结构、第 7 章的翻译质量评估、第 8 章的文本情感;第三部分由最后一章即第 9 章构成,聚焦于双语工具的开发与应用,旨在拓展 Python 与语料库翻译的可融合性。

写作本书的一个重要目的就是设想将自己多年积累的翻译实践经验转化为可实证描述的翻译知识。这一设想已在我的第一本专著《语料库与翻译》

中有所体现,现如今结合了 Python 技术,使我更有可能通过大数据的方式去挖掘翻译实践知识。以此方式所能获取的翻译知识将更有可能助力于中译外的翻译实践,助力于“讲好中国故事、传播好中国声音”的伟大实践。本书的呈现使我的设想在一定程度上得以实现,感谢技术进步所赐予的各种可能性和可行性——愿与技术同行。

本书的写作得到了诸多同事、同行、学生和朋友的帮助、鼓励和启发。首先感谢上海交通大学外国语学院和同事,是他们充满创意和新意的研究为我开阔了诸多可资写作的想象空间。同时,感谢学界和业界从事语言研究和应用的同行,是他们的论文或发言激励我运用 Python 进行持续不断的探索。其次感谢上海交通大学、复旦大学、上海外国语大学等学校的学生,是他们富于创意的课堂提问或练习作业使我有更多可资思考的来源。因此,本书的写作参与者更为多元,其中的第 7 章由厦门大学外文学院的陆晓蕾老师写作完成。她在翻译质量评估方面探索多年,硕果累累,由她完成第 7 章可以显著提升内容的学术性和实践应用性。厦门大学外文学院 2020 级翻译硕士陈晨同学贡献了 7.2.1 节的部分案例。上海交通大学外国语学院 2021 级翻译硕士马佳宁同学和 2020 级翻译硕士秦恺誉同学分别贡献了 8.2.1 节“企业风险要素文本的情感分析”和 8.2.2 节“汉英对比与中国形象”,他们对问题的思考方式使我深受感动。最后,还要感谢诸多随我一起舞动 Python 学习热情的老师和同学们,你们的不懈支持是我继续前行的动力源泉;同时也感谢上海交通大学出版社为我出版了 Python 图书三部曲。关注微信公众号“Python 语言数据”,可直接复制按图书章节排序的相应代码。

本书的写作意在将 Python 技术应用于语料库翻译研究与翻译实践,同时期望本书能够成为“计算翻译学”的出发点,抛砖引玉,共同进步。因此,这一写作过程充满着探索式体验,其中必有不足之处,还望各位不吝指正。愿以本书为新起点,共同推进语料库翻译的理论深化与技术创新。

管新潮

2022 年 1 月 10 日于上海

目录 ▶▶▶

第1章 语料库翻译概述	1
1.1 语料库翻译学	1
1.1.1 概念与定义	1
1.1.2 多变量设置	2
1.1.3 总结与展望	7
1.2 语料库与翻译	8
1.2.1 语料库翻译的缘起	8
1.2.2 业界的语料库与翻译	9
1.2.3 语料库翻译对比	10
参考文献	11
第2章 多变量协同效应	13
2.1 多变量表征概述	13
2.1.1 译者风格表征	13
2.1.2 显化与隐化的变量表征	15
2.1.3 语际多变量表征	19
2.2 多变量协同路径	21
2.2.1 著作权法英译风格分析	21
2.2.2 词汇维度的多元线性回归	26
2.2.3 双语文本语义迁移对比	29
参考文献	35

第 3 章 译文可读性	38
3.1 文本/译文可读性概述	38
3.1.1 英文可读性	38
3.1.2 汉语可读性	43
3.1.3 领域特色可读性	46
3.1.4 译文可读性	47
3.2 译文可读性探索路径	48
3.2.1 著作权法英译文的可读性	48
3.2.2 可读性公式对专业译本的适用性	52
3.2.3 译文句子复杂性对比	57
参考文献	63
第 4 章 翻译对等的短语特征	65
4.1 翻译对等概述	65
4.1.1 共选理论	65
4.1.2 翻译共选	67
4.1.3 翻译对等	69
4.2 翻译对等的短语实现路径	72
4.2.1 双语文本对应词的相互对应率	72
4.2.2 文本互信息共现	78
4.2.3 译文与原文的术语对等	83
参考文献	87
第 5 章 翻译知识库的构建与应用	89
5.1 知识库概述	89
5.1.1 语义知识库	89
5.1.2 翻译知识库	91
5.1.3 讨论与小结	93

5.2	专业通用词翻译知识库	95
5.2.1	专业通用词概念	95
5.2.2	翻译知识库概念与定义	97
5.2.3	专业通用词提取方法	98
5.2.4	翻译知识库构建模式	105
5.2.5	译后编辑应用	113
	参考文献	117
第6章	语言结构识别与译后编辑	119
6.1	语言结构概述	119
6.1.1	语言共性与翻译共性	119
6.1.2	语言结构及其作用	121
6.1.3	译后编辑模式	126
6.2	语言结构及其译后编辑探索	128
6.2.1	语言结构对译后编辑的影响	128
6.2.2	新闻话语与中译外	132
6.2.3	复杂语言结构的识别	135
	参考文献	138
第7章	翻译质量评估	141
7.1	翻译质量评估概述	141
7.1.1	翻译质量评估理论	141
7.1.2	基于测试的翻译质量评估	143
7.1.3	计算机辅助翻译质量评估	144
7.2	机器翻译质量自动评估指标与方法	146
7.2.1	基于 N-gram 相似性的评估	146
7.2.2	基于翻译编辑率的评估	153
7.2.3	基于向量相似性的评估	156
7.2.4	无参考译文的机器翻译质量自动估计	159

说明	161
参考文献	162
第 8 章 文本情感对比分析	165
8.1 情感对比分析概述	165
8.1.1 评价理论与情感分析	165
8.1.2 双语可比与情感分析	167
8.1.3 翻译关系与情感分析	170
8.2 情感对比分析案例	172
8.2.1 企业风险要素文本的情感分析	172
8.2.2 汉英对比与中国形象	179
8.2.3 基于情感值的机翻译文分析	184
参考文献	187
第 9 章 双语工具开发与应用	189
9.1 工具演变概述	189
9.2 改进型工具	192
9.2.1 双语上下文语境关键词	192
9.2.2 双语平行对齐	195
9.2.3 平行文本格式互换	198
9.3 新思路工具	202
9.3.1 平行文本信息提取	202
9.3.2 双语文本清洗	207
9.3.3 原文与译文的相似性对比	210
参考文献	215
附录 汉英术语对照表	216

第1章 语料库翻译概述

语料库翻译是指依托语料库所能实现的翻译研究和翻译实践。与翻译研究相关的学科是语料库翻译学,诸多已出版或发表的文献(胡开宝 2011: 3;王克非 2012a: 5;肖忠华 2012: 1;黄立波 2021: 1-2)皆表明,这一学科缘起于1993年 Mona Baker 发表在论文集 *Text and Technology: In Honour of John Sinclair* 中的一篇论文 (Corpus linguistics and translation studies — Implications and applications)。其随后在世界各地发扬光大,尤其是以英汉语言对为主的学术研究将语料库翻译学提升至新的高度。这一点已得到国内外出版/发表的诸多专著或论文的印证。语料库翻译学出现的意义在于翻译学术研究的创新,即为翻译学研究指出一条新的路径。与翻译实践相关的语料库翻译研究极有可能将翻译实践中不自觉运用的翻译规律呈现给世人,可彰显理论与实践相结合的检验意义。

1.1 语料库翻译学

1.1.1 概念与定义

语料库翻译学是指“以语料库为基础,以真实的双语语料或翻译语料为研究对象,以数据统计和理论分析为研究方法,依据语言学、文学和文化理论及翻译学理论,系统分析翻译本质、翻译过程和翻译现象等内容” (胡开宝 2011: 1),或者是指“以语言理论和翻译理论为研究上的指导,以概率和统计为手段,以大规模双语真实语料库为对象,采用语内对比与语际对比相结合

的方法,对翻译现象进行历时或共时的描写和解释,探索翻译本质的一种翻译学研究方法”(王克非 2012a: 4)。

由此可知,语料库翻译学立足于语言数据即双语语料或翻译语料,利用数据挖掘信息提取手段,在语言学或翻译学等理论的指导下展开翻译研究。这一点与语料库语言学颇为相似,所不同的是后者主要以单语语料作为研究对象。语料库语言学是指“以真实的语言数据为研究对象,从宏观的角度对大量的语言事实进行分析,从中寻找语言使用的规律;在语言分析方面采用概率法,以实际使用的语言现象的出现概率为依据建立或然语法进行语法分析”(杨惠中 2002: 4)。其实,语料库语言学的各种单语研究方法也同样适用于语料库翻译学。

从翻译实践视角看,语料库翻译学的研究范式和结果呈现使之成为最有可能为翻译实践提供理论性指导的一种理论。如语料库翻译学的“形式化程度说”:由形式化程度较高的语言翻译成形式化程度较低的语言,显化现象发生递减,而隐化现象发生递增(柯飞 2005)。这一理论性描述可构成英译汉实践的理论性依据,如下文所述:

本书虽属翻译,但力避卅年代弱小民族自卑心理下之欧化文体。诸如“当……时候”“假若……的话”“散步着”“有着”“被成功地实验了”“房子被建好了”“快速地跳”“公然地反对”“那些花朵们”等句法文词,全避而不用。(张振玉 2009: 5)

上述例句亦可解释为一种翻译隐化行为。汉语本无此类句法文词,在提倡文化自信的当下更应隐去这些本就无须存在的表述。这其中自然也包括诸多人称代词、介词等的隐化。

1.1.2 多变量设置

早期的语料库翻译学研究多为翻译共性、译者风格等内容,迄今为止的对翻译规范、目标语语言变化、意识形态、形象构建等的研究进一步丰富了语料库翻译学的研究内容。这使得语料库翻译学的关注对象从翻译语言本身转向

至影响翻译行为的社会、文化、历史、政治等方面的限制性因素(黄立波 2021: 165)。无论是翻译共性或译者风格,还是意识形态或形象建构等,研究内容的阐释一般是以特定的变量加以表征。变量的代表性及其协同效应将直接影响到语料库翻译学研究内容的表征意义。

1) 翻译共性

翻译共性“又称翻译普遍性或翻译语言普遍特征,是指翻译语言作为一种客观存在的语言变体,相对于源语语言或目标语原创语言整体上所表现出的一些规律性语言特征”(黄立波 2021: 88)。基于语料库的翻译共性有显化、隐化、简化、范化、明晰化等,其在统计分析层面上经常采用类符形符比、罕用词、词长、句长等算法和指标(许家金 2018),即本文所指的变量。此类变量的统计结果可直接或间接说明翻译共性特征的显著性程度。

- 英语和汉语分属不同语系,形式化程度相差较大,在不同的翻译方向上显化或隐化的表现不一样。鉴于汉语的连接词(作为变量)等衔接手段和指称形式(或以代词为变量)都不同于英语等印欧语言,考虑到汉语表达习惯而做隐化处理的英译汉通常比不做隐化的仿译更为地道(柯飞 2005)。
- 显化和隐化的对称和非对称(假说)均与具体语言对和语料体裁相关。以因果连接词为变量,语言对“法语-荷兰语”双向翻译中(小说)显化与隐化之间是平衡的,呈对称关系(Denturek 2012),而英汉语言对之间的非对称翻译关系更如柯飞(2005)所述须考虑汉语的习惯表达,但欧化翻译不属此列。
- 以文字总量和平均句长为变量,说明莎士比亚戏剧《哈姆雷特》的梁实秋译本和朱生豪译本在不同程度上存在显化,但朱译本显化程度要比梁译本更高(胡开宝,朱一凡 2008)。
- 政府公文英译文的简化研究显示(于红 2016):宏观的统计数据表明译文的实义词比例偏高、词汇平均长度大、长词比例高和词汇难度高,但是 STTR 的统计数据表明译文的词汇变化度低、词汇重复率高。因此宏观上的语料库统计数据不完全支持简化假设;微观上通过对功能词

IN 的非典型搭配分析,从词汇使用单一化的角度可以清晰地看出译者的简化处理痕迹。

- 即便是采用问卷调查方法,从另一个角度检验近年来基于语料库的实证翻译研究中所发现的翻译共性特征或翻译语体特征是否被译入语读者作为判断翻译或原创文本的依据,也离不开变量的设置如文体、句型、词汇等变量,以此从读者的视角考察汉语翻译文本和原创文本的辨识度(胡显耀,何广宁 2021)。

以相关变量的统计数值验证翻译共性存在与否,是语料库翻译学研究的常见方法。不同的翻译共性必须配以不同的变量或组合变量,方能验证其合理性。值得思考的一个问题是,究竟需要多少变量才能恰如其分地描述某一翻译共性。变量并非多多益善,因不同的变量彼此之间可能会产生相互作用,变量过多会导致解释力下降,如何把握这一相互作用是问题的关键。对翻译共性的考察主要是识别各类语言表现形式中有哪些变量是可以借助语料库工具进行直接统计测算的,而对无法直接统计的现象亦可从语义或话语的具体形式检索结果中做进一步的人工梳理(黄立波 2021: 94)。

2) 译者风格

广义的译者风格是指“译者在语言应用方面所表现出的个性特征以及包括译本选择、翻译策略应用、序跋和译注等在内的非语言特征”(胡开宝 2011: 109);狭义的译者风格是指“译者语言应用或语言表达的偏好,或在译本中反复出现的语言表达方式”(胡开宝 2011: 109)。如同翻译共性研究一样,译者风格研究的落脚点也是语言变量特征。这一点决定了翻译共性和译者风格的研究出发点是相通的。有所不同的是,译者风格是译者自身的文笔表述风格,是译者下意识的选择或习惯性选择的产物,即以具体的变量直接表征译者的译文风格,并由此指出不同译本译者风格的差异性。

- Olohan(2004: 153 - 156)考察了翻译英语语料库(Translational English Corpus, TEC)中两位译者(Peter Bush 和 Dorothy S. Blair)使用缩写形式如 it's(作为变量)的情况,指出 Bush 使用缩写形式的频率较高,且其使用频率接近英国国家语料库(British National Corpus, BNC)中缩写形

式的使用频率。其他研究则考察了外来词(作为变量)和转述动词(say和tell)的使用情况等。(转引自黄立波 2021: 101)

- 莎士比亚戏剧汉译本“被”字句(作为变量)的对比研究(胡开宝 2011: 120)表明,朱生豪的译者风格是力求最大限度译出莎剧原著内容,并着力使译文符合中文行文习惯;梁实秋的译者风格表现为还原源语文本的人文主义精神,尽量保存源语文本结构。
- 通过考察葛浩文英译小说汉英平行语料库,发现葛氏所译五部莫言小说的英译本在标准类符/形符比、中英文文字量比值、中英文句对主要类型排列顺序、said 报道小句语序、强调斜体词使用数量方面(五个变量)都基本一致,这说明葛氏翻译风格在这些方面具有一致性(侯羽等 2014)。
- 通过对比《到灯塔去》两个汉译本的及物性、情态、指示语和人称代词(四个变量),发现两译本对原作叙述视角采用了不同的翻译方法,对再现原文叙述视角产生了不同影响,体现出两译本各自的翻译风格(黄立波,石欣玉 2018)。其译者风格剖析是以相应变量的频率统计值为依据并结合适当的定性分析。

基于语料库的译者风格描写已取得有效进展,语料库规模越来越大,分析指标(变量)越来越多,跨学科属性越来越明显。那么,究竟需要多少变量才能描述某个局部风格或总体风格呢?多维度/多特征分析法或许是一种优选,其中的关键应该是不同变量之间的协同效应及其数值映射体现。多维(多变量)分析方法使译者风格描写从静态转向多译本之间动态对比,从局部描写转向整体考察,从相对于源语的译语风格转向进入目标语参照系统后的语域变异考察(赵朝永 2020)。

3) 形象建构

形象与翻译研究的结合可促使翻译研究者在“个体—群体—整体”动态链中考察不同层面的形象形成过程,关注翻译生产、传播和发挥功能的整个过程,为翻译产品、过程和功能研究提供多维度的观察视角(王运鸿 2018)。基于双语平行语料库或可比语料库的翻译与形象研究,可将社会、文化、历史、政