

【主持人语】(北京外国语大学 许家金教授) 本专栏4篇文章力图在语料库翻译研究的广度和深度上有所推进。许家金一文通过文献考订,指出语料库翻译研究不应限于翻译共性假说;在翻译语言特征上,不应止于词句的频数统计,更应关注语篇特征和语义特征;在统计方法上,主张采用多因素、多维分析法。黄立波、石欣玉一文通过比较《到灯塔去》的两个汉译本发现,宏观词频数据未必能有效区分翻译风格,应深入文本细读,挖掘核心叙事特征。朱一凡、陶庆、郭鸿杰一文将多维分析法运用于翻译质量评价之中,在67个语言特征和6个体裁维度上,对3组学习者译文进行了全面考察。徐秀玲一文一改此前翻译语言特征中过于偏重代词、名词使用频次统计的倾向,采取小切口、深挖掘的做法,详尽分析了汉语译文中“这/那”的回指功能。本专栏的几篇论文努力拓展语料库翻译研究在选题范围、语言特征、分析手段上的视野,特别注重深度描写,期待贴近翻译语言特征成因,希望能引起学界对语料库翻译研究的持续关注。

语料库翻译研究遗珠*

许家金

(北京外国语大学中国外语与教育研究中心,北京 100089)

摘要:目前的语料库翻译研究多限于对翻译共性的讨论,并将其发端归于 Baker (1993)。本文首先梳理了语料库翻译研究的核心内涵,并结合文献分析指出:1) 理论上, Baker 借鉴了 Blum-Kulka (1986) 的“明晰化假说”、Blum & Levenston (1978) 的“简化假说”和 Toury (1978) 的“翻译规范”理念等;2) 在翻译语料库建设方面,加拿大 Hansard 语料库的创建早于 TEC 语料库;3) 在研究方法上, Johansson & Hofland (1994) 基于英语-挪威语双向平行语料库提出的 ENPC 模型更为全面均衡;4) 实证方面, Gellerstam (1986) 对瑞典语翻译文本的研究,称得上语料库翻译研究的先声。概言之,本文拾取了文献中部分语料库翻译研究的“遗珠”,希望有助于明确前贤和后继者之间的学术联系。同时,我们也主张重视和发扬我国本土的语料库翻译研究理念和方法。

关键词: 语料库翻译研究; 翻译共性; 学术史

中图分类号: H059 **文献标识码:** A **文章编号:** 1002-722X (2018) 02-0001-10

0. 引言

借助双语或多语文本,以人们熟知的文字解读其他文字的做法,古已有之。有关罗塞塔石碑的研究便是个广为流传的例子。藏于大英博物馆的这块古埃及石碑,刻立于公元前 196

收稿日期: 2017-08-18; 作者修订: 2017-12-29; 本刊修订: 2018-01-23

基金项目: 教育部人文社会科学重点研究基地重大项目“服务国家战略的外国语言与外语教育创新研究”(17JJD740003) 子课题“大数据视野下的外语及外语学习研究”

作者信息: 许家金 (1977-), 男, 安徽芜湖人, 教授, 博士, 博士生导师, 研究方向为话语分析、二语习得、语言对比与翻译及语料库语言学, E-mail: xujiajin@bfsu.edu.cn。

年。石碑自上而下用 3 种文字刻写了一段诏书：上端是古埃及象形文字，属于当时的殿堂雅语；中间是百姓所用的埃及俗文字（Demotic）；下部是当时埃及的希腊统治者用古希腊文译出的诏书内容。正是通过这段三语平行文本，英法学者找到了释读古埃及象形文字的钥匙（Parkinson 2005：7、33-46）。在我国，借助外汉对齐文本解读佛教经典的梵汉对勘法，至少在隋唐便已存在。以严格的梵汉对勘法研究佛经文献和佛教汉语的历史，少则也有近百年。近 20 多年，通过梵汉对勘研究佛教汉语，更成为古汉语研究的一条重要路径（朱庆之 1990，2001）。罗塞塔石碑的三语对释和梵汉对勘之所以能够助力原典解读和译语剖析，关键在于双（多）语文本虽然采用的符号体系不同，但其传达的内容和含义是共通的。因此，平行对译文本是“以形索义”的重要资源（Johansson 2007：28）。而翻译的实质正是基于对原文含义的理解，进而通过目标语对原文再现的过程（Nida & Taber 1982：12）。

近些年，通过收集和观察大量原文和译文文本，开展翻译语言量化描写的研究成果相当丰硕。现有文献（Laviosa 2002；王克非、黄立波 2008；黄立波、王克非 2011；黄立波、朱志瑜 2012；胡开宝 2011，2012；胡开宝、毛鹏飞 2012）多指出语料库翻译研究发端于 Baker（1993）。诚然，就对语料库翻译研究的影响而论，Mona Baker 称得上个中翘楚。然而，将语料库翻译研究的起点定于 Baker（1993），在理论和实践方面都未必完全符合史实，这多少有割裂学术发展的承前继后和推崇英语中心主义之嫌。

本文将梳理此前谈及不多的相关文献，拾取一些具有语料库翻译研究性质的“遗珠”，希望有助于对整个领域建立更为全面的认识。

1. 语料库翻译研究界定

本文所谈的语料库翻译研究具备以下特征：

1) 理论与应用目的。相关研究的目的应是考察原文和译文之间的对应关系，或者是探究翻译语言自身的特质，即翻译语言的普遍性特征。

2) 翻译文本形态。研究数据必须是达到一定规模的（通常在 10 万字词以上的）双语或多语文本。其中的核心语料必须是翻译文本。语料可由原文译文对应文本构成（未必实现句级对齐），也可是不具有翻译关系的翻译文本和原创文本构成的可比语料，甚至是由平行、可比、历时等类型语料库构成的复合语料库（王克非、秦洪武 2017）。

3) 量化与对比方法。研究方法应当以量化为主，定量分析是语料库研究的必要条件之一（许家金 2014：35，2017：52）。量化统计可由手工完成，也可由计算机完成。统计量以频数为主，也包括长度（如词长、句长）等连续变量。对比是语料库翻译研究的基本方法，可以是原文和译文语际对比、译文和原创文本语内对比、不同译文之间的对比，或不同时期语料的对比等。

典型的语料库翻译研究应具备上述 3 项特征。下文将涉及的个别理论探讨型成果，其理念若指向上述 3 项特征，也属本文的讨论范围。综上，语料库翻译研究是指基于大规模翻译文本，采取量化和对比方法开展的有关翻译语言的理论（或者翻译语言本体研究）和应用研究。

2. 语料库翻译研究理念溯源

Baker（1993：243-245）倡导将语料库方法运用于翻译语言研究之中，并概括出以下共性假说：1) 明晰化，2) 消歧与简化，3) 语法规范化，4) 避免重复，5) 对目标语特征的夸

大, 6) 源于中介过程的特定翻译语言特征。Baker (1996: 176-177) 开始使用“语料库翻译研究”(Corpus-based Translation Studies) 这一称谓, 并将此前提出的翻译共性假说凝练为简化、明晰化、规范化及保守性、均质化, 逐步将语料库翻译研究向学科化方向推进。

就理论原创性而言, Baker 声明了自己的理论源头。在宏观上, Baker 指出她直接受益于 Even-Zohar (1979) 的“多元系统理论”和 Toury (1985) 的“描写译学”理论。而在理论假说的表述上, Baker (1996: 177) 承认, 除了“均质化”外, 其他几个假说都引自前人, 并非首创。

概括而言, Baker 的翻译共性假说关注的是“显隐”和“规范”两个方面: “显隐”是一币两面, 因此可归并讨论, 其研究的视野较微观; “规范”研究的视野更宏观, 它假定社会规约的存在, 其核心是将翻译语言置于言语社团语用规约的大背景下。在这两方面, Baker 受 Blum-Kulka (1986) 和 Toury (1978) 影响最大。然而, 正因为翻译共性框架中糅合了微观和宏观两个理论体系, 从而造成 Baker (1993, 1996) 共性假说之间相互重叠, 逻辑不清。譬如, 有的假说坚持形式标准, 有的坚持意义标准。孰不知, 形式明晰化常常会造成意义简化, 两者很难截然分开, 因此明晰化和简化极易混淆。“规范化或保守性”中也可能蕴含“明晰化”现象。

Baker 各项假说中有关“显隐”的部分基本来自 Blum-Kulka 的研究, Baker (1993: 243) 援引了 Blum-Kulka 的“明晰化假说”(the explication hypothesis)。Baker (1993, 1996) 未谈及的其他相关研究中, Blum & Levenston (1978) 专文论述了“简化假说”^①。Toury (1978, 1995) 的“翻译规范”理念对 Baker 也有直接影响, 它是指整个时代翻译行业的现状与具体翻译实践之间的互动关系。两相比较, Blum-Kulka 的“显隐”比 Toury 的“规范”更适合对特定的翻译文本进行微观研究。各项翻译共性假说中, 聚焦具体词句特征的“明晰化”最受关注。

与上述讨论最相关的研究中, Blum-Kulka (1986) 是一个相对被忽视的作品, 是本文所谈的语料库翻译研究在理念方面的一个重要“遗珠”。这篇文章主旨是谈篇章衔接方面的显化, 但其内容实际已涵盖 Baker 相关假说的多个方面。Baker 的各项假说中所谈的具体例子, 比如代词使用的多少、实词的使用、连接成分使用的多少, 不少都跟衔接有关(许家金、徐秀玲 2016), 旨在帮助建立文本的前后照应。比如, 所谓“显”, 要么是内容显, 要么是形式显, 前者会增加实词衔接, 即 Halliday & Hasan (1976) 所谓的词汇衔接; 后者会增加功能词在篇章衔接上的作用, 即 Halliday & Hasan (1976) 所谓的语法衔接。

Blum-Kulka (1986: 19) 明确提出了“明晰化假说”, 在此文文末她特别倡导开展相关实证研究, 即主张进行大规模对比实证研究(Blum-Kulka 1986: 32-33)。言下之意即基于大规模语料库验证“明晰化假说”。通过对比分析, Blum-Kulka 认为可以发现: 1) 某个语域的文本在 A 语言中的显化情况; 2) 该语域文本在 B 语言中的显化情况; 3) 相同语域的文本在 A、B 两种语言之间的显化情况。Blum-Kulka 希望据此了解到: 1) 译文可能存在的偏向目标语规范的衔接特征; 2) 译文可能存在的偏向源语规范的衔接特征(即所谓“迁移”过程); 3) 独立于源语和目标语而自成体系的衔接显化特征。在分析过程, Blum-Kulka 还指出要分清哪些是“强制性的”, 哪些是“自由选择”的衔接显化特征。这一区分早于 Klaudy (1998: 82-83) 的相关讨论。Blum-Kulka (1986) 一文不但是 Baker (1993) 的理论先导, 更是

把后继的“语际显化”“语内显化”，以及强制性和非强制性的显化都论述清楚了。所谓的源语透过效应，她也有所涉及。

我们认为 Blum-Kulka (1986) 对翻译语言特征的认识较为均衡。她清楚地认识到，翻译是“周旋在两种语言文化之间”的 (王克非 2010: 1)，翻译语言应当是源语、目标语，以及可能存在的共性特征共同作用的结果。我们认为所谓源语型共性和目标语型共性是伪命题，共性特征应当是不受源语和目标语影响的。但凡是受源语或目标语制约的特征，都会受到具体语言对的影响，故而很难推及共性。

另外，从翻译中意义的解读和传译角度来看，Blum-Kulka (1986: 18) 指出只有从语篇角度才能抓住翻译意义的转换。这一点值得我们思考。当前很多语料库翻译研究，似乎过度关注词汇、句式等形式特征的使用频率，而忽略了对意义的考察 (许家金 2016: 17)。在 Blum & Levenston (1978) 所提的“词汇简化共性” (universals of lexical simplification) 中，他们也是立足语义的，其具体表现为二语 (包括中介语、翻译等) 中使用上位词、近义词替代、迂回和释义等方法。

Blum-Kulka 在谈论“显”和“隐”时，强调都是发生在源语/母语和目标语之间的特征。而 Baker 的翻译语言共性更强调不看原文，采用可比语料库研究翻译语言特征，旨在确立翻译语言作为合法语言变体的学术地位。然而，翻译仍然不可避免地受到源语和目标语影响，是否存在脱离语际影响的共性，至今尚无定论 (Mauranen & Kujamaki 2004)。有关此类共性的论说，在 Baker (1993) 之前，Blum-Kulka (1986) 的框架中已有涉及，更早则有 Duff (1981) 的“第三语言观” (The Third Language) 和 Frawley (1984) 的“第三语码说” (The Third Code)。

我们主张，语料库翻译研究不应只局限于翻译共性假说，应当将母语文本、原文文本、目标语文本都纳入视野之中。因此，基于翻译相关语料库的对等 (equivalence) 和对应 (correspondence) 研究、翻译共性特征研究、跨文化研究、翻译社会变异研究、社会认知研究、翻译的批评话语研究都是语料库翻译研究的合理部分。在更抽象意义上来说，翻译语言研究应当兼顾共性论和相对论两个方面 (House 2016: 41)，因此理论视角和研究方法也应丰富而多样。

3. 翻译语料库的建设

国际范围内，早期的双向平行计算机语料库中，Hansard 语料库影响最大，这可能也是最早的平行语料库之一。该语料库的准确名称应是“加拿大 Hansard 语料库” (the Canadian Hansard Corpus)，否则容易同英国议会单语 Hansard 语料库 (目前规模已达 16 亿词次) 混同。Hansard 一词系 1803 年以来英国议会官方会议纪要印刷和出版者的姓氏，英联邦各国议会辩论文本都称 Hansard。鉴于加拿大 Hansard 语料库的英法双语属性，文献中有时也会以复数 Hansards 指称。20 世纪 80 年代末开始，加拿大 Hansard 语料库由 IBM 托马斯沃森研究中心和贝尔实验室采集加工。据 Church & Gale (1991)，初期的 Hansard 平行库规模大约在 100 万词左右。1995 年通过宾州大学的“语言数据联盟” (LDC) 发布的加拿大 Hansard 语料库版本 (LDC95T20) 说明文档中未见总库容信息，但其中 IBM 所创建的子库就达 287 万英法平行句对。

Gale & Church (1991) 较早运用加拿大议会英法双语语料库进行英法句对齐实验和跨语

言搭配词的抽取。该语料库的设计目的和最终使用领域主要是在机器翻译领域，这顺应了20世纪90年代初基于大规模语料库进行统计机器翻译的潮流。

Hansard 系列语料库可以理解为天然语料库，或者说监控语料库，只要议会会议进行，语料库就会扩增。稍晚出现的欧洲议会多语平行文本库和联合国决议多语平行文本库，甚至包括如今谷歌神经机器翻译背后的海量多语平行库，都是此类天然平行语料库。这些由计算语言学家主导的语料库建设工程，其目的不在翻译研究，而是为了开发和优化机器翻译系统。当然，其中的对齐算法、平行索引工具等给语料库翻译研究者带来了不少便利。

早期服务于翻译理论研究的电子化语料库中，仍属 Baker 主持创建的 TEC (Translational English Corpus) 翻译英语语料库为早，且影响较大。TEC 库大约于 1996 年开始建设，到 2000 年左右规模稳定在 1000 万英文单词 (Luz & Baker 2000)。该语料库收集了从各个语种翻译到英语的英语译文，包括飞行杂志文章 (128 篇)、报纸文章 (166 篇)、名人传记 (13 部)、小说 (81 部)，如果把名人传记和小说都计作文学类的话，那么该库实质上是一个英译文学库，其中的报纸杂志译文大致可以忽略不计。BNC 语料库被设定为 TEC 的参照语料库，然而就取样代表性而言，两者并不具有太多可比性。

我国双语语料库的建设大致跟西方同步，在语料库翻译研究项目中，1997 年开始筹建，2001 年左右完成的平行双向“中日对译语料库” (徐一平、曹大峰 2002) 规模为 2000 万字。该语料库从取样看较为平衡，并不限于文学文本。中日对译语料库实现了句级对齐，全部实施了词性标注，其中有 10% 的语料进行了句法剖析。由于句对齐和校对工作，平行语料库建设的人力成本数倍于单语语料库。从产生的年代、语料规模、均衡取样和技术实现等多方面看，中日对译语料库的建设都称得上是一项极重要的工作。可能由于语种所限，该语料库影响范围并不广，这可以算是翻译语料库建设方面的“遗珠”之一。

在英汉平衡翻译语料库研制方面，王克非团队的工作在国际范围内也是领先的。有关翻译语料库建设的更多信息，可参阅王克非、黄立波 (2012) 关于国外 8 大双语语料库的概括。另外，在语料库翻译研究活动中，自始至终都不缺少专书翻译研究，譬如特定作家作品翻译研究或特定译者研究。此类研究更多具有个案分析的性质，本文不再赘述。

总体来看，早期的翻译语料库建设多数采取方便取样 (convenience sampling) 原则，较少按随机或分层等方法进行科学抽样。包括 TEC 这样的知名翻译语料库，取样都相当偏态。取样不佳的语料库，纵使规模数百万，其研究所得结论终究难以推广。语料库建设中，取样代表性始终优先于语料库规模。在均衡取样的大规模平行语料库方面，我国起步较早，成效可观，如“中日对译语料库”“北外通用平衡语料库”。当然，除了知名团队的重大课题之外，像肖忠华创建的 ZCTC (ZJU Corpus of Translational Chinese) 和 COTE (Corpus of Translational English) 这样的布朗家族平衡翻译语料库，则是个人研究者值得学习的范例。

4. 语料库翻译研究方法的提出

语料库翻译研究方法与相关语料库的建设相辅相成、不可分割。本节所谈的研究方法包括语际和语内的对比分析以及适应翻译研究的计量分析方法。

在对比分析方法方面，北欧、西欧学者较早给出了一些系统的方案，其中较有代表性的是 ENPC 分析模型 (the ENPC model) (Johansson & Hofland 1994: 26; Oksefjell 1999: 199; Johansson 2007: 11)。ENPC 是挪威奥斯陆大学和卑尔根大学合作研制的英语-挪威语双向平行语料库

(English-Norwegian Parallel Corpus) 的缩写。该语料库建于 20 世纪 90 年代初, 专门用于翻译研究, 而非机器翻译。其在文类(文学及非文学)、翻译方向(英-挪/挪-英、英-瑞/瑞-英)等方面设计较为严谨。这个模型是集平行语料库与可比语料库于一体的研究框架。

图 1. ENPC 对比分析模型

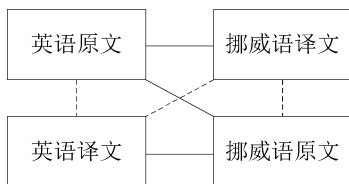


图 1 引自 Johansson & Hofland (1994: 26)。Oksefjell (1999: 199) 和 Johansson (2007: 11) 在描述 ENPC 模型时, 虚线标注跟 1994 年版有所不同。虚线的标识一般跟作者关注的研究层面或研究目的有关。例如, 1994 年版中英语译文和挪威语译文之间标注为虚线, 意味着 Johansson 和 Hofland 当时更为关注语际对应, 而不是翻译共性。ENPC 对比分析模式已指明了各种可能性, 即借助双语和多语语料库, 我们可以开展基于原生语言的对比语言学研究、不同语言对之间原文和译文的对应分析、不同译文之间的翻译语言特征研究。前一种属于类型学研究或对比语言学研究, 后两种都属于语料库翻译研究。Baker 及弟子所开展的研究主要是最后一种。从 ENPC 对比分析模型综合来看, 除了可以考察源语译语差异和译文特征, 还可以推究翻译语言特征中可能存在的“源语透过效应”(Teich 2003)。

我国语料库翻译研究由外语界学者主导, 事实上在汉语史研究领域, 针对佛教翻译汉语(朱庆之称“佛教混合汉语”)所采用的研究方法也十分值得关注。朱庆之自 1989 年前后开始的梵汉对勘研究, 便是基于大规模梵汉字句对齐语料, 考察佛教混合汉语的语言特征和译者风格。方法论上, 朱庆之精辟地概括出了研究汉语“来龙去脉”的复合语料库分析模式, 即综合运用平行语料库、可比语料库和历时语料库。朱庆之(2014: 序 3)指出可以“用译文与平行梵文对比分析的方法, 发现译文中直接来自原典的成分; 其次, 用译文与非佛教文献对比分析的方法, 发现译文中那些虽与原典无关却独见或首见的语言现象, 看看它们的产生是不是翻译时的创造”。朱庆之(1990)在其博士论文中便采用了上述方法开展佛教汉语词汇和语法范畴的研究。朱庆之及其弟子所采用的梵汉对勘法, 可以说是比较典型的平行语料库的方法(朱庆之 2014)。另外他所用的译文与非佛教文献的对比方法, 则与 Baker 等人提倡的可比语料库方法别无二致。龙国富(2013: 29)也提出了相同的方法论认识。此后, 刘云(2013)还将该方法用于满汉合璧文献的研究之中。这些具有本土原创性的观点, 也是本文希望彰显的语料库翻译研究方法论的“遗珠”。

语料库翻译实证研究方法在统计分析层面, 经常采用的型次比(type-token ratio, TTR)、罕用词(hapax legomena)、词长、句长等算法和指标, 也是语料库翻译研究学者从前人那里借鉴而来的。起先这些方法主要用于作者认定、文体计量学以及数理语言学研究(Yule 1944; Whatmough 1956; Herdan 1960, 1966)。比如, Yule (1944) 基于不同作品及不同版本作品中(常用)名词的出现情况等, 尝试判定作品的真伪, 分析不同作家作品的风格差异, 还考察了同一作者不同作品之间的词汇分布差异, 鉴定作品的作者争议等, 其中用到了词汇频数的描述性统计、卡方检验和线性回归分析等方法。除了 Yule (1944) 为探讨作品真伪、

作者风格所采取的量化分析外, Whatmough (1956: 240-244) 特别关注了诗歌、学术型语篇和其他体裁所体现出来的“情绪渲染”(affectivity 或 emotive) 和“频量信息”(frequency、quantitative 或 informative) 等文体功能差异。在 Yule (1944) 涉及的量化统计手段之外, Whatmough 采用的一个重要指标便是罕用词或一次词。数理语言学家 Herdan (1960, 1966) 更是致力于以量化方法构建起一套全新的语言学理论, 他将语言概括为“选择”(language as choice) 与“概率”(language as chance)。Herdan 深受索绪尔影响, 认为“选择-概率”语言观是“语言-言语”(langue-parole) 二元论的另一视角。他将语言的本质视为选择, 语言的具体表现则具有概率性。这一认识除了具有普通语言学意义外, 对于文体分析、翻译研究都有价值。Herdan (1960: 304-326、348-368, 1966: 196-200) 有专门章节对英法、英西、英德翻译文体进行了统计研究。在 Herdan 的研究中, 他详细而全面地采用了各种文体变异分析指标, 包括单词音节数、平均词长、平均句长、词汇和词类(名、动、形等)分布等。另外, Herdan (1960: 26) 还特别讨论了型次比及其优化算法(如 logTTR)。“型”(Type) 和“次”(Token) 作为两个符号学术语, 在 Peirce (1906: 506) 的概念体系中, 也跟“语言-言语”和“选择-概率”暗合。二者经过数理语言学家 Herdan 的深入讨论, 再经 WordSmith Tools 这样的文本分析工具的技术实现, 成为了广泛采用的词汇丰富度测量手段。

理念、语料、方法、工具俱备, 相关的语料库翻译实证研究便顺理成章了。以下介绍一些较少谈及的重要研究案例。

5. 语料库翻译实证研究

若不以电子语料库为判断依据, 那么 1993 年之前也存在很多基于人工对翻译语料进行的计量研究。例如 Halliday (1959/1974) 对回鹘蒙古文译成的汉译本《元朝秘史》语法体系的量化描写, Herdan (1966) 对英法、英西、英德翻译文本的量化分析, Church & Gale (1991) 开展的双语对齐及平行索引技术研究等。它们要么没有将其视为翻译研究, 要么只是顺带提及, 或重在统计原理的验证和技术开发, 并非严格意义上的翻译研究。

Vanderauwera (1985) 也是文献中时常提及的一项早期语料库翻译研究, 然而它并不符合前文有关语料库翻译研究的界定, 因为该研究没有采用统计手段, 未能对翻译语言的概率特征进行考察, 全书只是通过举例论述了一系列重要观点。Baker (1993, 1996) 所谈及的翻译语言特征, Vanderauwera (1985) 中几乎都有涉及。例如, 书中谈到荷兰语小说英译本呈现出语篇组织上更连贯的总体趋势, 并且原文隐含的信息会在译文中以明确的方式表达 (Vanderauwera 1985: 97); 原本指称相对模糊的代词常被更精准的同指表述取代, 长难句式也会得到简化 (Vanderauwera 1985: 98); 为使译文行文流畅, 冗长迂回的表达和无关节节会做截省处理 (Vanderauwera 1985: 98); 具有口头叙事特色的语篇, 会调整为更为“得体”(proper) 的风格 (Vanderauwera 1985: 100), 等等。

值得一提的是, Gellerstam (1986) 对瑞典语翻译文本语言特征的量化研究完全符合前文界定的 3 项语料库翻译研究特征。Gellerstam (1986) 对 27 部瑞典语翻译小说(多数为英译瑞) 和字数规模相当的 29 部瑞典语原创小说进行了语内类比, 探讨所谓“翻译体”(translationese) 的问题。Gellerstam (1986: 88) 开篇就明确指出翻译体并非表明译文质量欠佳, 而是源语对目标语的系统性影响。Gellerstam (1986: 91-92) 通过量化统计总结出了翻译瑞典语在 5 个方面的词汇语法规律。例如, 瑞典语翻译中明显多用外来语; 英语原文的口语化表述

在瑞典语译文中变得比较中性、规范,如文中提到英语原文脏话在目标语中被以规范的形式处理,这即是洁净化。Gellerstam (1986: 92) 还提到了“词语均质化”,虽然这一提法跟后来 Baker (1996: 177) 表达的含义并非完全一致,但都强调翻译文本的语境变异性小。除了词汇因素, Gellerstam 还探讨了瑞典语译文中句法层面的源语“指纹”。可能因为 Gellerstam (1986: 94) 关注的是瑞典语翻译语言特征,而不是翻译英语研究,所以没有得到外界的足够关注,但其完全称得上语料库翻译实证研究的“遗珠”案例。

以上研究案例,可以说是翻译语言共性特征研究的先声, Baker (1993, 1996) 直接受惠于这些研究。语料库翻译研究在过去的 30 多年间承前继后,逐步壮大,如今已蔚为大观。

6. 结语

我们主张梳理学科发展脉络时应重视前贤和后继者之间的学术联系,这有助于全面了解语料库翻译研究的理念、语料库建设思路和研究方法。另外,学问未必都由西来,我国本土的理论和方法也应重视和发扬。中国学者有其内在需求及技术和理论实力,有能力做出具有原创价值的语料库翻译研究。本文对语料库翻译研究的认识,不止于翻译共性相关研究。虽然说不考虑原文文本,专注于翻译文本自身特点并无不可,但若深入讨论翻译语言特点,源语和目标语仍是不可逾越的。近年翻译共性研究的选题困境从侧面说明,除去翻译语言本身,还应当充分结合跨语言、跨文化视角,这样才可以拓展更多的选题空间。技术和方法上,当前语料库翻译研究过于注重浅层语言特征的描写,对意义的考察不够。统计方法上,多因素、多维分析还比较缺乏。

* 感谢黄立波教授和智玉英同学提供有关资料。

注释:

① 这里所谈的“简化假说”并非只针对翻译语言而言。在他们的研究视野中,简化现象在二语学习者的口语和书面语、外语教师语言、外语教材和翻译语言中均有体现。

参考文献:

- [1] 胡开宝. 语料库翻译学概论 [M]. 上海: 上海交通大学出版社, 2011.
- [2] 胡开宝. 语料库翻译学: 内涵与意义 [J]. 外国语, 2012, (5): 59-70.
- [3] 胡开宝, 毛鹏飞. 国外语料库翻译学研究述评 [J]. 当代语言学, 2012, (4): 380-395.
- [4] 黄立波, 王克非. 语料库翻译学: 课题与进展 [J]. 外语教学与研究, 2011, (6): 911-923.
- [5] 黄立波, 朱志瑜. 语料库翻译学: 研究对象与研究方法 [J]. 中国外语, 2012, (6): 28-36.
- [6] 刘云. 早期北京话的新材料 [J]. 中国语文, 2013, (2): 170-177.
- [7] 龙国富. 《妙法莲华经》语法研究 [M]. 北京: 商务印书馆, 2013.
- [8] 王克非. 翻译: 在语言文化间周旋 [J]. 中国外语, 2010, (5): 1.
- [9] 王克非, 黄立波. 语料库翻译学十五年 [J]. 中国外语, 2008, (6): 9-14.
- [10] 王克非, 黄立波. 国外双语库研制与应用评析 [J]. 外语电化教学, 2012, (6): 3-10.
- [11] 王克非, 秦洪武. 基于历时复合语料库的翻译与现代汉语变化考察 [J]. 外语教学与研究, 2017, (1): 37-50.
- [12] 徐一平, 曹大峰. 中日对译语料库的研制与应用研究论文集 [M]. 北京: 外语教学与研究出版社,

2002.

- [13] 许家金. 许家金谈语料库语言学的本体与方法 [J]. 语料库语言学, 2014, (2): 35-44.
- [14] 许家金. 基于可比语料库的英语译文词义泛化研究 [J]. 中国翻译, 2016, (2): 16-21.
- [15] 许家金. 语料库研究学术源流考 [J]. 外语教学与研究, 2017, (1): 51-63.
- [16] 许家金, 徐秀玲. 基于可比语料库的翻译英语衔接显化研究 [J]. 外语与外语教学, 2016, (6): 94-102, 122.
- [17] 朱庆之. 佛典与中古汉语词汇研究 [D]. 成都: 四川大学博士学位论文, 1990.
- [18] 朱庆之. 佛教混合汉语初论 [C] // 北京大学中文系语言学论丛编辑部. 语言学论丛 (第二十四辑). 北京: 商务印书馆, 2001: 1-33.
- [19] 朱庆之. 序 [C] // 王继红. 基于梵汉对勘的《阿毗达磨俱舍论》语法研究. 上海: 中西书局, 2014: 1-8.
- [20] Baker, M. Corpus linguistics and translation studies: Implications and applications [C] // M. Baker, G. Francis & E. Tognini-Bonelli. *Text and Technology*. Amsterdam: John Benjamins, 1993: 233-250.
- [21] Baker, M. Corpus-based translation studies: The challenges that lie ahead [C] // H. Somers. *Terminology, LSP and Translation*. Amsterdam: John Benjamins, 1996: 175-186.
- [22] Blum, S. & E. Levenston. Universals of lexical simplification [J]. *Language Learning*, 1978, 28 (2): 399-415.
- [23] Blum-Kulka, S. Shifts of cohesion and coherence in translation [C] // J. House & S. Blum-Kulka. *Interlingual and Intercultural Communication*. Tübingen: Narr, 1986: 17-35.
- [24] Church, K. & W. Gale. Concordances for parallel texts [Z]. Oxford: Seventh Annual Conference of the UW Centre for the New OED and Text Research, 1991.
- [25] Duff, A. *The Third Language: Recurrent Problems of Translation into English?* [M]. Oxford: Pergamon Press, 1981.
- [26] Even-Zohar, I. Polysystem theory [J]. *Poetics Today*, 1979, 1 (1/2): 287-310.
- [27] Frawley, W. Prolegomena to a theory of translation [C] // W. Frawley. *Translation: Literary, Linguistic, and Philosophical Perspectives*. Newark: University of Delaware Press, 1984: 159-175.
- [28] Gale, W. & K. Church. A program for aligning sentences in bilingual corpora [C] // *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*. Berkeley: Association for Computational Linguistics, 1991: 177-184.
- [29] Gellerstam, M. Translationese in Swedish novels translated from English [C] // L. Wollin & H. Lindquist. *Translation Studies in Scandinavia*. Lund: CWK Gleerup, 1986: 88-95.
- [30] Halliday, M. *The Language of the Chinese "Secret History of the Mongols"* [M]. Oxford: Basil Blackwell, 1959/1974.
- [31] Halliday, M. & R. Hasan. *Cohesion in English* [M]. London: Longman, 1976.
- [32] Herdan, G. *Type-Token Mathematics: A Textbook of Mathematical Linguistics* [M]. Gravenhage: Mouton, 1960.
- [33] Herdan, G. *The Advanced Theory of Language as Choice and Chance* [M]. New York: Springer-Verlag, 1966.
- [34] House, J. *Translation as Communication across Languages and Cultures* [M]. London: Routledge, 2016.

- [35] Johansson, S. *Seeing through Multilingual Corpora* [M]. Amsterdam: John Benjamins, 2007.
- [36] Johansson, S. & K. Hofland. Towards an English-Norwegian parallel corpus [C] // U. Fries, G. Tottie & P. Schneider. *Creating and Using English Language Corpora*. Amsterdam: Rodopi, 1994: 25-37.
- [37] Klaudy, K. Explicitaion [C] // M. Baker & K. Malmkjaer. *Routledge Encyclopedia of Translation Studies*. London: Routledge, 1998: 80-84.
- [38] Laviosa, S. *Corpus-Based Translation Studies: Theory, Findings, Applications* [M]. Amsterdam: Rodopi, 2002.
- [39] Luz, S. & M. Baker. TEC: A toolkit and API for distributed corpus processing [C] // S. Bird & G. Simmons. *Proceedings of Exploration-2000: Workshop on Web-Based Language Documentation and Description*. Philadelphia: LDC, 2000: 108-112.
- [40] Mauranen, A. & P. Kujamaki. *Translation Universals: Do They Exist?* [M]. Amsterdam: John Benjamins, 2004.
- [41] Nida, E. & C. Taber. *The Theory and Practice of Translation* [M]. Leiden: Brill, 1982.
- [42] Oksefjell, S. A description of the English-Norwegian parallel corpus [J]. *International Journal of Corpus Linguistics*, 1999, 4 (2): 197-219.
- [43] Parkinson, R. *The Rosetta Stone* [M]. London: The British Museum Press, 2005.
- [44] Peirce, C. Prolegomena to an apology for Pragmaticism [J]. *The Monist*, 1906, 16 (4): 492-546.
- [45] Teich, E. *Cross-Linguistic Variation in System and Text: A Methodology for the Investigation of Translations and Comparable Texts* [M]. Berlin: Mouton de Gruyter, 2003.
- [46] Toury, G. A rationale for descriptive translation studies [C] // T. Hermans. *The Manipulation of Literature: Studies in Literary Translation*. London: Croom Helm, 1985: 16-41.
- [47] Toury, G. *Descriptive Translation Studies and Beyond* [M]. Amsterdam: John Benjamins, 1995.
- [48] Toury, G. The nature and role of norms in literary translation [C] // J. Holmes, J. Lambert & van den Broeck. *Literature and Translation*. Leuven: ACCO, 1978: 83-100.
- [49] Vanderauwera, R. *Dutch Novels Translated into English* [M]. Amsterdam: Rodopi, 1985.
- [50] Whatmough, J. *Poetic, Scientific, and Other Forms of Discourse: A New Approach to Greek and Latin Literature* [M]. Berkeley: University of California Press, 1956.
- [51] Yule, G. *The Statistical Study of Literary Vocabulary* [M]. Cambridge: Cambridge University Press, 1944.

(责任编辑 陈 勇)