

基于统计方法的多译本翻译质量评测

张英贤¹, 蒋跃^{1*}, 杨延龙²

(1.西安交通大学 外国语学院, 陕西 西安 710049; 2.西安建筑科技大学 文学院, 陕西 西安 710055)

摘要: 研究以《傲慢与偏见》的人机译本为例, 使用问卷调查法和统计学的方法, 试图寻找一种有效对比和评测机器翻译与人工翻译质量的方法。将定性定量分析相结合, 通过测算统计学上的基本概念, 如均值、标准差、中数、众数、极值、偏度、峰度等, 分析了各个译本的得分集中趋势、离散程度和分布情况等。结果表明, 机器翻译在处理短句时, 能产生质量非常高的译句, 但整体上与人工译本仍存在较大差距, 其总的翻译得分基本处于最底端。人工译本得分基本呈正态分布, 各译本相互之间差异不显著。通过讨论分析文学翻译的限制因素, 发现时代因素对翻译的影响很大。同时, 研究发现, 在翻译文学作品应综合考虑原作的语义和语篇信息结构, 较完整地保留原有风格, 并适应读者群的阅读习惯, 方能实现较好的翻译质量。此外, 设计有效的问卷可以客观评估翻译质量, 对改善机器翻译和人工翻译质量具有借鉴意义。

关键词: 机器翻译; 翻译评价; 问卷调查; 傲慢与偏见; 统计

中图分类号: H 059 **文献标志码:** A **文章编号:** 1008-7192(2016)04-0078-06

虽然语言不同阻碍了人类的交流, 但重建“巴别塔”的人类共同梦想却从未停止过, 翻译因此也是一项历史悠久的活动。随着时代的进步和全球化进程的加快, 翻译已从最初的精英事业转变成一种基本交流形式。上世纪中叶以来, 随着不同语种和不同文化的快速融合以及个体交流的日益频繁, 人工翻译已不能满足现代大区域和高效率的翻译需求, 机器翻译(以下简称“机译”)逐渐迈入历史舞台, 并随着计算机和大数据的研究快速发展为大众不可或缺的一项技术。机译已经在当今的国际政治、经贸、教育、文化等各个方面发挥愈来愈重要的作用。但是, 无论是机器翻译还是人工翻译, 自翻译活动产生起, 对于翻译质量评估的研究就应运而生。正如翻译学家辜正坤^[1]所说“翻译标准问题是翻译理论的核心问题, 也是一个哥德巴赫猜想式的问题”。因此, 正确评价翻译作品质量优劣, 将对翻译(尤其是机译)的发展具有重要的导向作用。

一、翻译评估的相关研究

中西方的翻译研究史中, 各个流派的翻译原则

和标准不尽相同。语文学派通常以原语为导向, 注重字词句的对应, 如“泰特勒三原则”等。而语言学派认为语言的作用就是交流, 因此翻译作品应以忠实、通顺、文雅等作为衡量标准, 如Newmark^{[2]82-98}区分了翻译的两种导向——交际翻译和语义翻译(前者重视语言的形式和作用, 后者主要指作品的文化和风格对等)。功能学派指出翻译是有目的的行为, 并将“忠诚原则”引入功能主义模式, 指出译者应同时为原文和译文环境负责^{[3]179-200}。描写学派指出翻译研究应描写人类普遍的翻译现象, 并总结一些普遍规律来解释和预测该现象^{[4]269-280}。基于此, 描写学派建立了多元系统翻译理论^{[5]283-288}。而文化学派认为, 翻译活动受多种方面限制, 如赞助人、诗学规范、语用、经济、意识形态等, 因此文学翻译本身就是一种“折射”和“改写”^{[6]354-374}, 在传播的过程中为了妥协各方制约, 一定会有“改写”的行为。之后翻译研究进入后翻译学时代, 最大的特点就是多种翻译标准共存, 译者身份也开始凸显, 不再是之前所倡导的“透明”、“隐身”状态, 翻译模式也走向多元化。

收稿日期: 2016-04-02

基金项目: 2014年陕西省社科科学基金项目“中国学者英语学术论文中模糊限制语的缺失问题及其对策的研究”(2014J05)

作者简介: 张英贤(1989-), 女, 西安交通大学外国语学院硕士研究生, 研究方向为语料库翻译学、计算语言学; 蒋跃(1958-), 男, 西安交通大学外国语学院教授, 博士生导师, 研究方向为语料库翻译学、计量语言学。Email:yuejiang58@163.com

***通讯作者:** 蒋跃

在翻译学研究中,对于翻译作品的考察往往等同于翻译标准,主要是少量的字、词、句或者对译文的直观感受,归纳译者的翻译特点,进而评测翻译作品的优劣。这种方法通常为主观内省式的翻译研究,缺乏实际的调查,客观性欠佳。因此,学者们开始考虑引入客观评价的方法。范守义^[7]将模糊数学理论应用到翻译评价中,提出了九个层面的标准,并尝试确定其隶属度,具有一定的创新性。然而,该理论各种隶属度的确定主要通过人为制定,且各层面权重也视为均等,影响结果的可靠度。穆雷^[8]在此基础上,使用人工评测法找出更加客观的隶属度确定方法,验证了模糊统计法在翻译评估中的可操作性。随着机译的发展,翻译质量的自动评价方法也越来越多,主要是针对机译而研发的,如 BLEU、NIST、METEOR、TER (Translation Edit Rate) 等。因其对于改善机译质量有重要的意义,也受到众多研究者的关注。但是,自动评测翻译方法通常是以字、词等表层结构为基础,难以涵盖语义层面,不能客观公正地评价译文。此外,翻译本身还受到译本读者的限制,读者虽然不是翻译批评者,但却是翻译程序中的终结点^[9]。自动翻译评测中,将读者排除在外,不利于作品的传播。因此,这类纯粹的定性分析就显得“刚性有余,柔性不足”。

翻译评估模式基本分为两类:量化模式和非量化模式。翻译标准的确立,应当融合客观与主观层面,前者可以降低主观性,有利于实现更高层次的客观性,而这正是翻译质量评估的最核心要求。后者则有利于对翻译本质的探讨,从而促进翻译质量的提升,推动翻译作品的传播。针对中文译本的研究很多,但大多为主观定性分析。近年来,随着语料库翻译学的发展,有研究者开始使用语言的计量特征来对比不同的译本^{[10][56-61][11]}。这种方法比传统的翻译研究更加客观,论述更加合理。但也仅限于对语言单位的解构,忽略了译作的整体效果,且未将读者对译文的感觉融合起来。

在翻译评估中,发展较为成熟的评测体系,主要是针对机器翻译的自动评测系统,不太适用于文学作品的翻译。尤其是在评估大规模翻译作品质量时,通常会耗费较多时间和精力,通常都是特殊句子的例证法,而且样本量较小,未能找到客观有效的评估方案。问卷调查是人文社科研究的经典方法之一,也经常用于翻译质量评估中^[12]。其主要特点是

在纯粹状态下考察某一因素,易于做出明晰的结论;而综合的着眼点是众多因素的相互联系、相互作用及其整体效应^[13]。因此,本研究尝试将问卷调查法和客观统计学方法相结合,选用较多的样本量,来评估整个译本的翻译质量。因为问卷调查法能够综合译本的总体翻译效果,并将读者的认同度直接加入到评测系统中,大样本量调查研究具有统计学意义,可在不同视角下实现对译文的客观公正评价。该研究通过主观评价和客观计算分析相结合,对人工和机器译文的翻译质量进行评测,为翻译风格的测量提供一种可操作的方法。而且使用数理统计算法,能够使译本对比和翻译风格的定量研究更加科学客观,有一定的推广价值。

二、研究设计

1. 语料选取

所选用的语料为《傲慢与偏见》,是英国作家 Jane Austin 的经典著作,曾被毛姆列为世界十大小说名著之一。这部小说以幽默的语言再现了英国中世纪的风土人情及婚姻价值观,其文学价值和社会价值对读者的婚姻观和人生观产生了巨大的影响,颇受读者欢迎。小说的汉语译本多达几十种,男性女性译者均有,还有部分合译本,时间跨度颇大,翻译风格迥异。上世纪 50 年代我国出版了王科一译本,为后来的译者提供了宝贵经验,后于 1980 年重印(上海译文出版社,1980)。孙致礼译本(译林出版社,1985)清新流畅、准确传神,目前销量较好。张玲、张扬译本(人民文学出版社,1995)由两位译者合译,很有特点。张经浩译本(浙江文艺出版社,2004)中译者主体性较强,译风自由。综合考虑时间、翻译风格、译者性别等因素,本研究使用的人工翻译语料选定为上述四个译本。此外,随着互联网和人工智能技术的飞跃式发展,机译尤其是网络在线翻译,在克服不同语言间的障碍方面发挥了很大作用。目前在线翻译网站有很多,而百度在线翻译依托强大的搜索实力,是国内比较先进的在线机译系统,应用很普遍,且已获得 2015 年国家科学技术进步奖,代表了机译技术发展的方向。因此,研究选定百度翻译作为机器在线译本的代表。

2. 问卷设计

《傲慢与偏见》全书共计 61 章,由于人力物

力所限,且考虑到评分员在评分时要保持良好的状态,不能承受太重负荷,因此无法对文中所有译句都进行评估。但问卷是针对整个译本设计的,需具有系统性,要尽可能涵盖整个译本,要求合理抽样。因此,研究根据采用等距抽样的原则,每隔一章随机选取一句英文及其对应的五句汉语译文作为抽样,共计31组、186个完整句子(英语句子有长有短,句型差别较大,具有代表性)。之后,根据随机性的原则,借助EXCEL中的随机数表功能,打乱每组内部译文的排列顺序(英语原句均在译句前面)。如,第一组排列顺序为:英语原句、王科一、孙致礼、张玲、张经浩、百度;第二组顺序随机为:英语原句、孙致礼、张玲、百度、王科一、张经浩等。采取乱序有助于排除评分员在打分过程中受到前面分数和译者的影响,使评分过程更加客观。此外,在问卷中还隐去了译者名字,且无任何指示译者身份的信息。

通常翻译评估中的问卷采用两种评分法,即分析法和综合法。分析法是指设定若干项目,按照规定的评分标准给每个项目都进行评分,然后再比较最后总分^[12]。但是这类评分标准的设置是主观的,有时几项之间会出现交叉,评分员可能无法在有效时间内正确区分两项之间的差异,容易割裂读者的整体印象,因为部分之和并不等同于整体,易导致“只见树木,不见森林”的片面翻译评价观。分析法通常表述较为复杂,操作性不强,具有盲目性^[12]。因此,本研究采取综合评分法,评分员根据忠实、通顺等基本原则,综合考虑译文的整体可接受性,分别给每个译句打分。每个译句后面都设置有5个打档次可供选择,即2分、4分、6分、8分、10分。分数越高,代表译本的可读性越强,则说明翻译质量越好。问卷中的级别差设置为2分,可以方便识别译本之间的差距,且更便于评分员操作。

3. 确定评分员

采用综合法虽然可以考察译文的整体状况,但也有些弊端。评分员通常具有主观色彩,调查结果中相应地会掺杂诸多不确定性因素等,为了消除这些因素的影响,需要增加样本数量,即选用较多的评分员。而且评估的是文学作品翻译,因此评分员需要具备较高的英汉双语水平,且对翻译标准有一定的了解。本研究选取的评分员均为西安几所高校的英语教师(所占比例60%)或英语专业的研究

生及博士生(所占比例40%),他们对于英汉双语的驾驭能力较高。翻译课程是英语专业的一门基础课程,因此,教师和学生对于翻译标准的认识都较为成熟,评分过程比较客观可靠。本研究共发放纸质调查问卷120份,回收113份,回收率为94%。观察发现,有个别问卷评分不完整或不符合规范,需要剔除这类无效问卷。经过筛选后共得到有效问卷107份,有效率达95%。之后对调查结果进行数据转录,并使用统计分析工具SPSS 22.0进行了数据整合分析。

三、结果和讨论

经过统计,31组句子中每组内5个译句的平均得分都在5.33~6.47之间,最大均值比最小均值高出1分左右,相差不大。说明评分员在打分过程中严肃认真,评分较为稳定,调查结果的可信度较高。研究了一些统计学常用计量指标,来进一步考察各译本的得分的集中趋势、离散程度和分布情况等。

1. 译本在五个档次的得分统计

问卷中总共设置5个分数档,我们分别统计了每个译本各自在5个档次的得分数量(表1)。为了更直观地对比各得分档次的数量差异,我们使用了折线图来表示(图1)。

表1 各译本在不同得分档次的数量统计

	王科一	孙致礼	张玲	张经浩	机译
2分	264	110	172	225	2406
4分	774	409	647	779	492
6分	979	860	883	995	284
8分	849	1102	998	858	103
10分	451	836	617	460	32

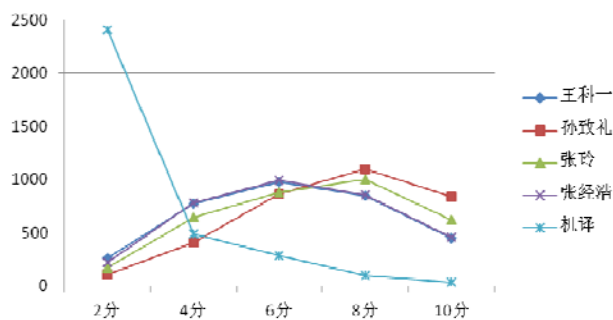


图1 各译本在不同得分档次的数量折线

从表1可知,人工译本和机器译本在5个档次上均有得分,机译最高得分为10分,但数量极少。说明机译有些句子的质量非常好,一些评分员给予

满分, 结果令人满意。如:

例句 1: It is a truth universally acknowledged, that a single man in possession of a good fortune must be in want of a wife.

[机译]这是一个举世公认的真理, 在一个有钱的单身男人一定需要一个妻子。(3.9 分)(此处及下文例句后面的分数是指平均分)

[王科一]凡是有钱的单身汉, 总想娶位太太, 这已经成了一条举世公认的真理。(7.1 分)

在例句 1 中, 机译能够传达原文的基本语义, 且使用了一个四字成语“举世公认”, 增加了译文的文学色彩, 可读性较强。但是与王科一译本相比, 机译仍拘泥于原文的句型结构和语序, 不如人工那样灵活自然地重整译句的顺序结构。文学作品不同于科技文章, 无固定的专业术语, 对于机译系统是个很大的障碍。且总体来说, 机器译本的高分不多, 与人工翻译仍有很大的距离。同样, 在图 1 中, 机译最低分 2 分的数量远远高出其它译本, 基本处于翻译质量的最低档次。相比于规律性较强的科技、经贸、法律等专业性语言, 使用机译处理文学作品会有更大的障碍。随着自然语言处理技术的发展以及机器神经网络的自主学习, 相信机译系统将会产出质量较高的文学翻译作品。

从图 1 可看出, 4 个人工译本的得分情况变化幅度都不如机器译本变化大。相比于机译得分集中在低端 2 分, 人工翻译基本集中在 6 分、8 分的中高档次。其中张经浩和王科一两条折线几乎重合, 说明这两个译本在 5 个 9 档次的得分分布情况非常接近, 由此可预测这两个译本的相似度或许最高。人工译本相互之间的差别都不大, 表明人工译作存在某种共性。

2. 各译本得分的基本统计数值

研究中使用的问卷有 31 组句子, 有效问卷为 107 份。通过对问卷数据统计分析提取特征量指标, 包括平均值、中位数、众数、标准偏差等 8 项, 能够说明译本得分的集中趋势、离散程度和分布情况等。各统计特征量的值如表 2 所示。其中最大值和最小值反映得分的分布区间; 平均值、中数和众数均可从不同角度反应译本得分的集中趋势和一般水平; 而标准偏差则反应译本得分的离散水平(波动大小); 偏度表明译本得分分布偏离正态分布的程度(值越大, 越偏离正态分布); 峰度表明得分

分布的尖锐程度^[14]。

表 2 各译本的特征量统计

	王科一	孙致礼	张玲	张经浩	机译
最小值	2	2	2	2	2
最大值	10	10	10	10	10
平均值	6.27	7.29	6.75	6.33	2.90
中位数	6.00	8.00	6.00	6.00	2.00
众数	6	8	8	6	2
标准偏差	2.310	2.171	2.287	2.274	1.698
偏度	-0.046	-0.466	-0.228	-0.038	1.996
峰度	-0.845	-0.528	-0.838	-0.831	3.503
合计	20 800	24 192	22 384	21 000	9 628

从表 2 可看出, 无论是人工翻译还是机译, 得分中均存在 2 分和 10 分, 全句都是 8 分, 说明两类译本中均出现一些翻译较好和较差的文句。这表明机译在某些情况下也能翻译出较好的译文。然而人工译本得分的平均值均在 6.2 以上, 而机译的平均得分小于 3 (仅为 2.9), 不足人工译本平均最低分的 1/2, 这表明虽然机译在某些文句上翻译较好, 但整体翻译质量较人工翻译尚有不小的差距。研究发现, 机译在简单句翻译中句子流畅, 得分较高。例如:

例句 2: A few days after this visit, Mr. Bingley called again, and alone.

[机译]这次访问之后的几天, 彬格莱先生又来了, 独自一人。(5.1 分)

[张经浩]宾利先生这次做客后才几天又登门了, 是单独来的。(5.7 分)

例句 2 中, 机译与人工译者差别较小, 相似度极高, 两句译文得分相差仅有 0.6 分。在翻译这类短句子时, 机译能够依据英语原文, 产出流利的译文, 同时保持完整的语义。但是在处理长难句时, 机译明显逊于人工翻译, 会产生诸多语法、语义、语篇等错误。如机器译文“一个星期后在幸福的爱情和计划专业”、“和大自然的不足却通过教育或社会小辅助”等, 这类句子佶屈聱牙, 甚至无法传达最基本的源语信息。因此, 针对长难句的问题, 机译系统需要句法的深度学习, 并将语用因素考虑进去。

从中位数和众数的值进行比较也能发现相似规律, 人工翻译的中位数和众数明显高于机译。机译的中数和众数均为 2, 其大部分译句得 2 分, 表明其翻译质量有待提高。人工和机译的标准偏差接近, 说明两类数据的变化情况类似, 但该指标不能

表明哪一类翻译质量较好,仅能体现译者前后翻译质量的波动和离散程度。而在得分分布方面(偏度值和峰度值),人工翻译和机译有较大的差距,其中人工翻译的值在-1~0之间,近似于正态分布稍左偏或负偏。表明人工翻译得分处于边缘状态和极值(最大值和最小值)的得分较少,整体翻译质量较好。而机译偏度约为2,峰度3.5,均大于0,极大地偏离正态分布且右偏或正偏,极值位于高分段。同样从表1也可得出相同结论,机译大部分得分集中在2分的低分段,故造成其偏度和峰度值较大。

3. 调查结果反思

通过研究数据可知,几个中文译本中,孙译本得分最高(平均分7.29)。仝亚辉^[15]用“清新流畅、准确传神”高度评价孙译本,认为其不拘于原文词汇限制,能够准确地传达原文意境。除了机器译文,目前王译本得分最低(平均分6.27)。但是,在前人的研究中,该译本也曾广受好评,曾一度被奉为经典。孙致礼^[16]就曾评价王译本“在传神、入化方面颇具特色”。但是,翻译不是简单的字词句转换,同时还所受到多种语内和语外因素的限制,如语言形式、经济、赞助人、文化、意识形态、诗学规范等^[6]。王译本早在1955年就已出版,留有很多旧时因素,打上明显的时代烙印。从简单的人名、地名、称谓等都可以看出,如“吉英”、“娘们儿”、“麦里屯”、“浪博恩”、“哈德福郡”等。但是,随着时代的变迁,人们的思想和阅读偏好发生了巨大的转变,“娘们儿”等字眼很可能会涉及到性别歧视的问题,引起女权主义者的反感。早期的译作中有些句式结构已经不符合现代的阅读习惯,如“决不定”,在现代汉语语料库中仅搜到12条检索结果。因此,多种因素叠加,会导致王译本得分较低。

此外,在文学翻译中,还应尽可能还原原著的场景和人物形象,保留原语的特点,不能随意改变原语结构。例如:张经浩的改变原作的结构和语篇,因此得分较低。例句:

例句3: Colonel Fitzwilliam's manners were very much admired at the parsonage, and the ladies all felt that he must add considerably to the pleasure of their engagements at Rosings.

[张经浩] 牧师府对菲茨威廉上校的翩翩风度非常欣赏,几个女人觉得,有了他,去罗辛斯做客会多出许多风味。(5.21分)

[王科一] 费茨廉的风度大受牧师家里人的称道,女眷们都觉得他会使罗新斯宴会平添不少情趣。(7.31分)

[孙致礼] 菲茨威廉上校风度翩翩,受到牧师家众人的高度赞赏。夫人小姐们都觉得,他定会给罗辛斯的聚会平添不少情趣。(8.51分)

在例句3中,原句是被动结构,而张将其转换成了主动结构,主位—述位结构发生了很大变化。相比于王和孙的译句,张得分最低。主位—述位信息结构对构建语篇有重要作用,张的译句对原文的形式结构改变很大,虽“得其意”,却“忘其形”,有损于原作的文学感染力。孙译本能兼顾语篇结构和语义,重现原作的风格^[17],因而在调查中会得到更好的评价。

四、结 语

文学翻译受制于多种因素,对机译而言,是个巨大的挑战。本研究通过问卷调查综合评价法,把定量分析和定性分析结合起来,使用统计学中的均值、众数、偏度、峰度等统计指标对文学翻译进行评估。研究对机译目前的水平有了客观的认识,发现机译质量仍与人类期望存在不少差距。机译基本集中在得分的最底端,偏离正态分布。而人工翻译主要在分数的高端,基本呈正态分布。实验中个别机译句子能得到满分,说明机译对大规模文学翻译有一定的潜力,在不远的将来可能会成为一种产业。但机译系统需要深层句法分析学习。此外,对于文学作品的翻译,若能兼顾原作的语篇等形式结构和意义风格,将会得到更好的效果。

通过合理的问卷设计能够客观有效地评测整个译本,但研究限于人力物力等客观因素,问卷内容和评分员的设置都难免会有失偏颇。但作为一项探索性尝试,或许能为翻译质量评估提供一个新的角度和方法。“互联网+机器翻译”的时代对于世界优秀文学的传播有重要的意义,因此,也更需要可靠的综合评测翻译质量的方案。

参 考 文 献

- [1] 辜正坤. 翻译标准多元互补论[J]. 中国翻译, 1989(1): 16-20.
- [2] NEWMARK P. Communicative and semantic translation[A]// JUNPING L. Selected Readings on Western Translation Studies. Wuhan: Wuhan University Press, 2012.

- [3] NORD C. A functional typology of translations[A]// JUNPING L. Selected Readings on Western Translation Studies. Wuhan:Wuhan University Press,2012.
- [4] HOLMES J. The name and nature of translation studies. [A]// JUNPING L. Selected Readings on Western Translation Studies. Wuhan:Wuhan University Press,2012.
- [5] ZOHAR I. The position of translated literature within the literary polysystem[A]// Junping L. Selected Readings on Western Translation Studies. Wuhan:Wuhan University Press,2012.
- [6] LEFEVERE A. Mother courage's cucumbers: text, system and refraction in a theory of literature[A] // JUNPING L. Selected Readings on Western Translation Studies. Wuhan : Wuhan University Press,2012.
- [7] 范守义. 模糊数学与译文评价[J]. 中国翻译,1987(4):3-9.
- [8] 穆雷. 用模糊数学评价译文的进一步探讨[J]. 外国语, 1991(2):66-69.
- [9] 翁凤翔. 翻译批评标准意义的新视角[J]. 上海翻译,2005 (S1):37-41.
- [10] 王克非. 语料库翻译学探索[M]. 上海:上海交通大学出版社,2012.
- [11] 蒋跃. 人工译本与机器在线译本的语言计量特征对比——以5届韩素音翻译竞赛英译汉人工译本和在线译本为例[J]. 外语教学,2014(5):98-103.
- [12] 穆雷. 翻译测试及其评分问题[J]. 外语教学与研究,2006(6):466-471.
- [13] 苗东升. 模糊学导引[M]. 北京:中国人民大学出版社, 1987:28.
- [14] 邵建平,邓兆卉. 分配公平性的分布偏度与峰度描述研究[J]. 统计与决策,2008(3):144-147.
- [15] 仝亚辉. 清新流畅 准确传神——评孙致礼译《傲慢与偏见》[J]. 解放军外语学院学报,1996(3):53-59.
- [16] 孙致礼. 我译《傲慢与偏见》[J]. 解放军外语学院学报,1991(4):58-62.
- [17] 李明. 得意岂能忘形——从《傲慢与偏见》的两种译文看文学翻译中主位—信息结构之再现[J]. 广东外语外贸大学学报,2009(4):88-92.

A Statistic Approach to Quality Evaluation of Translations

ZHANG Ying-xian¹, JIANG Yue^{1*}, YANG Yan-long²

(1.School of Foreign Studies, Xi'an Jiaotong University, Xi'an 710049, China;

2.School of Arts, Xi'an Univ. of Arch. & Tech., Xi'an 710055, China)

Abstract: Based on various translation versions of *Pride and Prejudice*, this paper presents a statistic approach to evaluate comparatively the qualities of both machine translations (MT) and the human translations (HT) by using data retrieved from a large-scale questionnaire. With the combination of quantitative and qualitative analysis, several statistical concepts (means, standard variance, median, mode, extremum, skewness and kurtosis) are applied to measure the concentric tendency, distribution and discrete degree of the scores. Results show that MT could produce some translations of high qualities when rendering simple and short sentences. However, most MT scores are at the lower end of qualities and have a long way to go in comparison with HT. On the other hand, all the four human translations are normally distributed and no significant difference is found between them. The analysis of the restraints in literary translation reveals that the age factor has greater impact on translation. The textual information structure, the semantic meanings, and the styles of the source text, should be taken into account to adapt to the reading habit of target readers. Besides, a questionnaire, if necessary, can serve as an efficient method to evaluate qualities of translation and provide some reference to improve both MT and HT.

Key words: machine translation; translation evaluation; questionnaire; *Pride and Prejudice*; statistics

【编辑 吴晓利】