



汉译英机器翻译错误类型统计分析

——以外宣文本汉译英为例

蔡欣洁, 文 炳

(浙江理工大学外国语学院, 杭州 310018)

摘 要: 通过平行测评国内常用的机器翻译平台, 研究现有机器翻译平台在汉译英转换上的不足之处。通过搜狗、有道、谷歌和百度四个机器翻译平台对选取的案例文本进行汉译英转换, 并通过和人工译文的比较, 统计并分析国内机器翻译平台在汉译英转换中易出现的典型机器翻译问题。经对 *China Daily* 网站上一则演讲稿汉译英案例文本机器翻译译文和人工译文的平行比较, 发现所测试四个机器翻译平台在汉译英转换上, 存在词和句层面翻译错误、漏译、大小写格式错误等共性化的机器翻译问题。这些翻译问题有待机器翻译平台的完善和译者的译后编辑来处理。

关键词: 机器翻译; 汉译英转换; 演讲稿; 译文比较; 翻译错误; 漏译

中图分类号: H059

文献标志码: A

文章编号: 1673-3851 (2021) 04-0162-08

Statistical analysis on types of C-E machine translation errors: A case study on C-E translation of publicity text

CAI Xinjie, WEN Bing

(School of Foreign Languages, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: The study aims to analyze deficiencies of existing machine translation platforms in Chinese-English translation by means of parallel evaluation of common machine translation platforms in China. Four common machine translation platforms namely Sogou, Youdao, Google and Baidu were used to carry out Chinese-English translation of the selected case. Through comparing C-E machine translation with the human translation, the study counts and analyzes typical problems that easily occur in the C-E machine translation by machine translation platforms. Based on the parallel comparison between the C-E machine translation and human translation of a speech draft case on the website of *China Daily*, it is found that the tested four machine translation systems have common limitations in C-E machine translation including syntactic translation errors, omissions and format errors, which need to be further solved by improvement of machine translation systems and post-translation editing.

Key words: machine translation; C-E translation; speech draft; comparison of translation versions; translation errors; omission

20 世纪 40—50 年代, Weaver^[1] 首次正式提出并阐述了“用现代计算机进行自然语言转化”的设想, 这标志着基于现代电子计算机技术的机器翻译

思想问世, 为机器翻译理论的发展奠定了坚实的基础^[2]。自那以后, 虽然机器翻译理论研究、应用开发和实践研讨曾于 20 世纪 60—70 年代遭遇寒冬, 后

收稿日期: 2020-06-14 网络出版日期: 2020-11-05

基金项目: 浙江省教育厅一般科研项目(专业学位研究生专项)(14090031661911)

作者简介: 蔡欣洁(1996—), 女, 山东青岛人, 硕士研究生, 主要从事科技翻译方面的研究

通信作者: 文 炳, E-mail: wen2bing@126.com

又几度面临瓶颈,但总体上处于快速蓬勃发展态势。特别是进入21世纪以来,“机器翻译方法历经了由生成语言学为基础的规则方法向以数据驱动为基础的语料库方法的转变”^[3],现如今,鉴于机器翻译“不仅能够节省译者的时间和精力,而且能为需求翻译的客户节省资金”^[4],其已被广泛应用于语言服务行业 and 高校翻译专业教学与实践。

机器翻译的理想目标是借助计算机技术实现没有人工参与的自然语对转换,输出达到终端用户所要求翻译质量和语言表达水平的译文。进入21世纪以来,虽然机器翻译“正在向实用化、商品化的方向迈进”^[5],且机器翻译理论和应用研究取得了蓬勃发展,但不可否认的是,由于“翻译本身的复杂性”、“自然语言的复杂性”和“机器自身的局限性”^[6],机器翻译在现阶段仍存在词句翻译转换上的不足,往往需要机器翻译错误识别、译后编辑审校等后加工环节来提升译文质量。为厘清现行机器翻译平台存在的不足,不少学者对机器翻译开展了测评和比较研究,如:贺文照等^[7]以谷歌机器翻译为工具,以机译文可接受度问卷调查和机译文案例抽样对比为研究方法,考察了机器翻译在英语关系从句汉译转换上的效果和存在的问题,发现英语关系从句中存在定语关系从句、核心词位于介词短语中并构成复杂的各种逻辑关系、多项位于并列、省略关系代词等情况时,英译汉机译文易出现转化问题;Klubička等^[8]以译文准确度和翻译错误数为衡量指标,对传统机器翻译和基于神经网络与深度学习的机器翻译展开了比较研究。

表1 案例文本各测试平台机器翻译错误类型统计总表

一级错误 类型	搜狗翻译		有道翻译		谷歌翻译		百度翻译	
	统计数/个	占比/%	统计数/个	占比/%	统计数/个	占比/%	统计数/个	占比/%
词汇翻译错误	29	64.5	28	35.9	47	54.7	42	59.2
语句翻译错误	11	24.4	7	9.0	28	32.5	17	23.9
其他翻译错误	5	11.1	43	55.1	11	12.8	12	16.9
合计	45	100.0	78	100.0	86	100.0	71	100.0

从表1、表2可以看出:第一,不计单词的字母大小写错误,从案例文本汉译英转换的准确度来看,对案例文本的翻译结果错误少的依次为搜狗翻译、有道翻译、百度翻译、谷歌翻译;第二,测试的四个机器翻译平台存在一些个性化的汉译英转换问题,比如错误类型B普通多义词的义项选择、错误类型E成语熟语的翻译、错误类型G和H长短句的语句结构和逻辑混乱等;第三,各个平台在进行案例文本汉译英转换时,产生的具体各类错误数量不同,占总错

上述研究梳理了特定机器翻译平台对特定语言现象的翻译转换效果,不足在于:一是平行测评、比较的平台数量较为有限,二是对机译错误类型的梳理和划分较为粗浅。鉴于此,本研究以习近平主席《深化文明交流互鉴,共建亚洲命运共同体》演讲稿汉译英文本为研究案例,以搜狗、有道、谷歌、百度作为进行汉译英转换的机器翻译平台,以China Daily网站给出的官方译文为平行对照的人工译文(网址 <https://language.chinadaily.com.cn/a/201905/16/WS5cdeb785a3104842260bbe59.html>),通过统计分析和译例对照,分析国内机器翻译平台在汉译英转换中易出现的典型机器翻译问题。

一、案例文本机器翻译错误类型统计

本研究将案例文本的China Daily人工译文和搜狗、有道、谷歌、百度四个平台的机译文作了逐句平行对照。原汉语案例文本《深化文明交流互鉴,共建亚洲命运共同体》共计3608中文字符;使用四个机器翻译平台获取译文的时间为2019年10月7日。对于对照过程中发现的机器翻译错误类型分类,参考了罗纪美等^[9]和李梅等^[10-11]研究中采用的二级分类体系和错误类别归纳框架,同时结合案例文本比对实际,对所涉及错误类型具体分类有所增删。本研究所划分的机器翻译错误类型和各机器翻译平台经译文横向对比得出的错误类型统计数据汇总如表1所示,统计明细如表2所示。为了陈述清晰,本研究将各类机器翻译二级错误类型依表格内统计顺序标记为错误类型A—L。

误数量的比重不同,各机器翻译平台存在个性化的翻译问题。

具体来看,在案例文本机器翻译汉译英转换过程中,谷歌翻译的出错率最高。在统计并归纳的12类汉译英机译错误类型中,谷歌翻译在其中7类汉译英机译错误的数量统计上位居前二位(不含各机器翻译平台错误数量相同的错误类型J),其中尤以错误类型H译文长句结构或逻辑混乱的出错数量最高,且远高于其他三个测试平台,这说明谷歌翻译

表2 案例文本各测试平台机器翻译错误类型统计明细表

一级错误类型	二级错误类型	错误类型统计							
		搜狗翻译		有道翻译		谷歌翻译		百度翻译	
		统计数/个	占比/%	统计数/个	占比/%	统计数/个	占比/%	统计数/个	占比/%
词汇翻译错误	术语和准术语翻译错误(错误类型 A)	6	13.3	2	2.6	10	11.6	11	15.5
	普通多义词义项选择错误(错误类型 B)	3	6.7	5	6.4	12	14.0	9	12.7
	机械选词引起的翻译不恰当(错误类型 C)	5	11.1	7	8.9	9	10.5	6	8.5
	译文重复选词(错误类型 D)	4	8.9	2	2.6	5	5.8	3	4.2
	成语和熟语翻译错误(错误类型 E)	3	6.7	6	7.7	5	5.8	7	9.9
	词性或时态错误(错误类型 F)	4	8.9	4	5.1	3	3.5	2	2.8
语句翻译错误	词素或词块切分错误引起的短句结构或逻辑混乱(错误类型 G)	4	8.9	2	2.6	3	3.5	4	5.6
	译文长句结构或逻辑混乱(错误类型 H)	5	11.1	3	3.8	23	26.7	12	16.9
	译文语句的语境语义翻译错误(错误类型 I)	4	8.9	2	2.6	3	3.5	3	4.2
	无主句主语还原错误(错误类型 J)	2	4.4	2	2.6	2	2.3	2	2.8
其他翻译错误	漏译(错误类型 K)	3	6.7	17	21.8	10	11.6	8	11.3
	译文大小写错误(错误类型 L)	2	4.4	26	33.3	1	1.2	4	5.6
合计	不计算大小写错误	43	95.6	52	66.7	85	98.8	67	94.4
	计算大小写错误	45	100.0	78	100.0	86	100.0	71	100.0

在原文的语句切分、语义识别、逻辑判断和译文的语义还原、意群与结构搭建上相较其他机器翻译平台同比出错率更高,有待技术改善。百度翻译产生的主要汉译英机译错误类型和谷歌翻译产生的错误类型类似,集中于错误类型 A 术语和准术语翻译错误、错误类型 B 普通多义词义项选择错误和错误类型 H 译文长句结构或逻辑混乱,同时其对成语和熟语的翻译效果不佳,在案例文本翻译过程中对成语或熟语的翻译出错量最大,这在一定程度上说明,百度翻译在规范化术语库和准术语库、特色汉语表达平行语料库或云端语料的建设和大数据匹配上有提升空间。搜狗翻译和有道翻译在对案例文本进行汉译英转换时,出错数量较少。但有道翻译存在两个较明显的译文生成问题:一是在译文中存在较多漏译现象;二是生成的译文存在较多的大小写格式错误。对于前一问题,在翻译过程中发现,若译员减少输入文本的单位长度,如将按段输入原文至机器翻译平台改为按句群或按句输入原文至机器翻译平台,这将减少译文漏译现象的产生,但同时也将一定程度上加大译者的工作量。对于后一问题,目前

可行的解决方案是进行译后编辑处理,但同样也存在增加译者工作量的矛盾。下文将对本次案例文本机器翻译汉译英转换过程中,所测试的四个机器翻译平台体现出的共性翻译问题,以译例比较的形式作梳理分析。

二、机器翻译词汇翻译错误

英国翻译学家 Newmark 曾在 *A Textbook of Translation*^[12] 一书中指出:译者在翻译时,需要在意义、所指、衔接和自然表达四个层面对译文负责。即,译文需和原文语义等值、所指对等,译文的衔接和语言表达需要自然、连贯、规范、流畅。词汇是最小的语义单位,也是语句的基本组成元素。词汇的翻译转换是否合适,能极为直观地考察出译文是否做到了和原文语义等值、所指对等。现行机器翻译平台标配有云数据辞典库和数据量庞大且不断更新的文本语料库,这使得机器翻译平台具备了较高水准的词语抓取、识别、意义辨析和转换能力。但受制于机器翻译平台的高度机械性或不完善性、各类不同语种语言的多维个性以及语言在语义和语用上所

具备的超符号性,机器翻译在处理词汇翻译时,仍不免存在一些错误或不足,亟待系统的完善,或依赖于译者的译后编辑。现对本次机器翻译考察过程中发现的典型词汇翻译错误作总结分析。

(一)术语和准术语机器翻译错误

现行基于大数据语料库和深度学习机制的机器翻译平台,其大数据语料库的语料量、扩充速度和自我更新能力是影响翻译功能的重要指标项。上述语料库的语料量不足或功能缺陷,会导致译文质量的下降或显著的翻译错误。由此导致的词汇机器翻译错误主要表现为术语和准术语翻译错误、词语词性或时态错误等几类,这些错误将导致译文对原文语义和所指的偏离。

就术语和准术语而言,术语是“某一专业特有的词汇”,其“特点是一词一义”;准术语是“由普通词汇演变成的专业词汇”,且“在不同的领域,可能会演变为不同的意思”^[13]。理论上来说,如果机器翻译平台的语料数据库在深度、广度上足够,同时及时扩充,规范地收纳各学科或行业的专业词汇,机器翻译是能够高正确率地实现术语和准术语翻译的。但实际上,现行基于大数据和深度学习机制的机器翻译语料库搭建模式,在极大地促进语料的更新速度、拓宽语料来源面的同时,也造成了所收纳术语和准术语表达的潜在出错。网络高频出现的术语和准术语译法并不一定等价于正确规范的行业或学科译法,多样的语料来源导致了原词-译词关系中可匹配译词项的分支化。如本次实践所用文本为一则外宣演讲稿。所测试的四个机器翻译平台在翻译如例1、例2所示“亚洲命运共同体”、“亚洲文明对话大会”等原文中国特色话语术语时,就出现了译文各行其是、不够规范的问题。术语和准术语的一大特点是其表达具有固定性、规范性、严谨性和专业性,是某一行业、学科或领域统一的词汇或词组指称。因此,在翻译术语或准术语时,应依照行业、学科或领域的规范来翻译,实现译文和原文语义的等值、所指的对等,且维持其固定性、规范性、严谨性和专业性。测评的四个机器翻译平台受制于其机械性,在例1和例2的术语翻译上,都出现翻译谬误,未实现规范且专业的翻译转换。

例1 亚洲命运共同体

搜狗: An Asian Community of Destiny

有道: An Asian community of Shared future

谷歌: The community of Asian destiny

百度: The Community of Asian Destiny

China Daily: An Asian Community with a Shared Future

例2 亚洲文明对话大会

搜狗: Asian Dialogue of Civilizations

有道: The conference on dialogue among Asian civilizations

谷歌: The Asian Civilization Dialogue Conference

百度: The Asian Dialogue of Civilizations Conference

China Daily: The Conference on Dialogue of Asian Civilizations

(二)普通多义词义项选择错误

Catford^[14]指出:“词语意义实际是一个有层次的语义结构或网络,它由词语的多个义项构成”,这使得对实际语篇或会话中词汇意义的理解需参照词汇所在的语句及上下文关联语句。机器翻译的词义选择是基于辞典库、平行语料和数理统计模型来执行的,因而对分支义项少的词汇或词汇的高频应用义项识别转化能力强,对多分支义项、含蓄语义或词汇的低频应用义项识别转换能力稍弱。此外,机器翻译的语境辨析功能弱,也制约了其普通词汇义项选择的准确性,如例3中,原句隐去主语为“中华文明”,其中的“积淀”实际含义为“代表、象征”,而非物理意义上的“增叠数目”。测试的四个机器翻译平台中,只有有道翻译译出了原词在该上下文语境中的正确含义,而其他三个机器翻译平台生硬地将原词译为“accumulate”,是一处误译,未能将词语在该语篇情境中的节点义项正确译出,造成了译文在语义和所指上有悖于原文。

例3 (中华文明)积淀着中华民族最深沉的精神追求

搜狗: It has accumulated the deepest spiritual pursuit of the Chinese nation

有道: It embodies the deepest spiritual pursuit of the Chinese nation

谷歌: Accumulating the deepest spiritual pursuit of the Chinese nation

百度: Accumulating the deepest spiritual pursuit of the Chinese nation

China Daily: It represents the profound pursuit of the Chinese nation and provides a rich source of strength for its lasting development.

(三)机械选词或重复选词引起的翻译不恰当

各类语言之间存在着字面符号意义相近甚至等同、但话语口吻色彩大相径庭的近义或同义词汇表达。机器翻译平台擅长词义、词性、词型的判断和识别,而对词汇语体和情感色彩的判别较弱,这使得其在翻译选词上易出现译文字面含义乍看正确、实则相悖的情况。如在例4中,原文中的“洗礼”一词实则表示的是“长期经历、历时积淀”的含义,但测试的四个机器翻译平台,都将该词翻译为了“baptism”或其变形。“baptism”指的是“a Christian ceremony in which a few drops of water are poured on sb or they are covered with water to welcome them into Christian Church and often to name them”^[15],表示的是一种以水浸、浇或洒入教人的基督教宗教仪式,只是恰好中文也将其表述为“洗礼”。机器翻译平台未能识别判断出该点,而是机械地、字面地逐词对应翻译,因而造成了译文的选词错误,使得译文和原文在所指对象上相悖。

例4 亚洲先民历经岁月洗礼

搜狗: Asian ancestors have gone through the baptism of time

有道: Generations of Asian ancestors, through the baptism of time

谷歌: Generations of Asian ancestors have been baptized over the years

百度: Asian ancestors have gone through the baptism of years

China Daily: Generation after generation, our ancestors in Asia, with their tireless endeavors

机器翻译在选词上的机械性还表现为,其在处理语篇内多次出现的普通词汇或连续出现的近义词时,选定的译词往往重复。这虽然不影响译文的语义表达,但降低了译文的可读性,影响了译文表达的自然流畅。这一现象在本次实践的汉译英文本中表现得较为明显。汉语表达的一大特点是同义词或近义词的并置,以使语言表达实现强调、形式美、阅读节奏感强等功能。机器翻译在处理前后并置出现的同义词或近义词时,表现出统一选词的倾向性。如在例5中,原句中“对等”、“平等”的连续出现是近义词前后并置,以达到强调的作用。四个翻译平台在处理该处时都重复选定译词 *equal* 来翻译,虽语义正确,但可读性差,且有可能造成读者的阅读误解,不理解为何要重复用词。

例5 文明交流互鉴应该是对等的、平等的。

搜狗: Civilization exchanges and mutual learning should be equal and equal.

有道: Exchanges and mutual learning between civilizations should be equal and equal.

谷歌: Civilization exchanges and mutual learning should be equal and equal.

百度: The exchange of civilizations and mutual learning should be equal and equal.

China Daily: Such exchanges and mutual learning should be reciprocal, equal-footed.

(四)成语和熟语机器翻译错误

成语和熟语是结构固定、表义完整的一类习惯表达,并且往往具有文化负载性。成语和熟语的语义包含三个基本层面:字面语义、引申语义和文化内涵。考虑到即使是人工译员,其在处理语言文化内涵的转换传达上也因语言及文化异质性而面临极大困难,因此本次考察对成语和熟语机器翻译错误的统计标准为:只要翻译出原词语或熟语的引申语义即视为正确的翻译。以此标准来统计,测试的四个机器翻译平台仍存在可观数量的成语或熟语翻译错误。比较典型的如例6,在翻译成语“薪火相传”时,所测试的四个机器翻译平台都是机械地逐字翻译,导致译文语义停留在表层字面语义,而没有译出原表达的深层语义以及隐含的寓意,这易引起读者对原文语义内容理解混乱,也造成了原文和译文在语义信息完整度和表达功能上的差异。

例6 薪火相传

搜狗: Passing on the fire from generation to generation

有道: Passing on the torch from generation to generation

谷歌: The passing of fire and the protection of generations

百度: The passing of firewood and fire from generation to generation

China Daily: build on its heritage from one generation to the next

三、机器翻译语句翻译错误

从句层来看 Newmark^[12]所强调的意义、所指、衔接和自然表达四个层面对译文负责,指原文语句和译文语句在意义上等值、功能上等效,且译文的语句表达能贴合于目的语群体的表达和阅读习惯,自然、流畅、表义清晰地传达给读者。根据上述标准,

在本次考察中发现的机器翻译语句翻译错误集中表现为语句结构或逻辑混乱、语句语境语义翻译错误、无主句主语还原不当三类。其中,语句结构或逻辑混乱属于平台缺陷型机器翻译错误,即作为机器翻译平台内核的计算机代码逻辑语言系统,其具备的语料检索、匹配和逻辑判断能力与稳定性仍待完善;语句语境语义翻译错误、无主句主语还原不当属于语言特性型机器翻译错误,即语言的语境性导致了机器翻译在处理原文语义上的不足。

(一) 语句结构或逻辑混乱

词素是构成词的最小语义或语法功能单位;词块是指具有完整语义、符合语法规则的词语搭配组合。机器翻译中出现的短句译文结构混乱、语义错误或矛盾、逻辑混乱往往源于对短句中各词素或词块的切分、识别错误或关系判断混淆。如在例7中,原文中的“耕耘”和“灌溉”实际为两个并列动词,表示两个动作行为,但测试的四个机器翻译平台在判断词块关系时,都将“灌溉”一词误判为“耕耘”一词的宾语,因而导致了译文的翻译错误。

例7 (人们)开始耕耘灌溉

搜狗:……began to cultivate irrigation

有道:……began to cultivate irrigation

谷歌:……began to cultivate irrigation

百度:……began to cultivate irrigation

China Daily:……tilled and irrigated

长句由于句子元素更多、句式结构更为复杂,因此其机器翻译的错误率也更高。由此,长句翻译的正确率高低也成为了检验各机器翻译平台内置计算机代码逻辑语言系统分句划分能力、意群和句子结构辨析能力、句子元素逻辑关系判断能力强弱的有力依据。从本次考察来看,谷歌翻译和百度翻译对汉语长句的语义、结构与逻辑判别及转换功能弱,出现了较多结构混乱、逻辑错误或语义不通的译句。如例8中,原句包含着两层逻辑关系:首先“没有和平,冲突不断甚至战火纷飞”是后半句“经济增长……沦为空谈”的假设性条件句;其次,分句“没有和平,冲突不断甚至战火纷飞”是一个因果句,“冲突不断甚至战火纷飞”是因,“没有和平”是果。在翻译该句时,搜狗翻译和有道翻译正确译出了第一层的假设性条件句逻辑关系,但在判断分句“没有和平,冲突不断甚至战火纷飞”中的因果关系时前后颠倒。而谷歌翻译和百度翻译语义逻辑完全错误,将“without peace”判断为总领全句的假设性条件,而把“冲突不断”到“人民往来”几个词组都作为后半主

句的主语,造成了语义的荒谬和逻辑的不通顺。上述的翻译错误使得译文在语句语义、逻辑关系上错解了原文。

例8 没有和平,冲突不断甚至战火纷飞,经济增长、民生改善、社会稳定、人民往来等都会沦为空谈。

搜狗: Without peace, conflicts will continue and even wars will rage. Economic growth, improvement of people's livelihood, social stability and people's exchanges will all become empty talks.

有道: Without peace, conflicts and even wars will continue. Economic growth, improvement of people's livelihood, social stability and people-to-people exchanges will all become empty talk.

谷歌: Without peace, conflicts and even wars, economic growth, improvement of people's livelihood, social stability, and people's exchanges will all be empty talk.

百度: Without peace, conflicts and even wars, economic growth, improvement of people's livelihood, social stability and people's contacts will become empty talk.

China Daily: When peace is interrupted by conflict or war, economic growth, decent lives, social stability and people-to-people exchanges will all be out of the question.

(二) 语句语境语义翻译错误

语句的含义脱离不开语境,具体包括语篇语境(即上下文)、语言情境和文化情境。前者是由文本字符构成的篇章环境,是一种内情境;后两者是语言生成、语义表达和功能发挥的超符号环境,是一种外情境。机器翻译平台基于语言生成转化规则和大数据语料库来实现自动化翻译,具有高度机械性。因此,机器翻译平台在处理原文内情境时,若所译语句的关联上下文离自身的文字距离越近,翻译效果越好;反之,则越差。而当所译语句的原意义和功能涉及语言情境和文化情境两大外情境的转换时,机器翻译往往力有不逮。

汉语作为意合语言,对语境具有高依赖性。这种高程度语境依赖性使得机器翻译在处理汉译英转换时,尤其是曲折表达的汉语语句时,易出现只译语句字面意义而无法译出语句内涵的情况。如例9,

该句上下文都在谈中国同亚洲各国、世界各国的交流互鉴及中国贡献,因此原句的实际含义恰如 *China Daily* 译文所表达的那样,指的是中国作为亚洲和世界一份子的融入。测试的四个机器翻译平台都只逐字翻译,造成译文生硬且语义混乱。类似地,例 10 中“美人之美、美美与共”的停顿切分是“美/人之美,美美/与共”,再联系上下文可知其所表述的是多元共生共存的世界文明观。而测试的四个机器翻译平台都以字为单位切割原文来进行翻译转化,使得译文语义错误,表达也混乱生硬。

例 9 今日之中国,不仅是中国之中国,而且是亚洲之中国、世界之中国。

搜狗: Today's China is not only China's China, but also Asia's China and the world's China.

有道: China today is not only China of China, but also China of Asia and the world.

谷歌: Today's China is not only China's China, but also China in Asia and China in the world.

百度: China today is not only China but also Asia and China.

China Daily: China today is more than the country itself; it is very much a part of Asia and the world.

例 10 第二,坚持美人之美、美美与共。

搜狗: Second, adhere to the beauty of beauty, beauty and common.

有道: Second, adhere to the beauty of beauty, beauty and together.

谷歌: Second, adhere to the beauty of the beauty, beauty and the United States.

百度: Second, adhere to the beauty of beauty, beauty and common.

China Daily: Second, we need to uphold the beauty of each civilization and the diversity of civilizations in the world.

(三)无主句主语还原

无主句是汉语语句表达的一大特色。无主句的主语不出现在字面上,而是依赖流水句式的内在联系和意合语言的语义逻辑,通过上下文的前指、后指或情景语境还原来体现主语。而英语是形合语言,强调主谓关系清晰、严谨。因此,在汉译英转换时,译者往往需要对无主句进行主语还原。而机器翻译

对语境信息的敏锐度低。通过这次考察发现,机器翻译在还原汉语无主句时,倾向于将汉语原文无主句的主语不加区分地还原为第三人称。这种较为固化的还原方式虽然在一定程度上改善了机器翻译的无主句指称还原,但一刀切式的处理方式机械生硬。以例 11 为例,原文作为一则外宣演讲稿,其隐含的语言情境是一种面对面的、强调语言共情功能的对话语境,因此其中大多数无主句的隐含主语应是陈述人所代表的第一人称。而几个机器翻译译文都将原文隐含主语还原为第三人称,这种译法在语法和语义上虽对,但削弱了原文的文本功能,仍未实现原文和译文的功能和价值对等。

例 11 有着相似的历史境遇、相同的梦想追求。

搜狗: They share similar historical circumstances and the same dream pursuit.

有道: They share similar historical circumstances and common dreams.

谷歌: They have a similar historical situation and the same dream pursuit.

百度: They share similar historical situations and same dream pursuits.

China Daily: We went through similar historical trials and hold the same dream for the future.

四、结 语

在本次案例文本机器翻译考察中发现,所测试的搜狗、有道、谷歌和百度四个机器翻译平台在汉译英转换上存在共性化的词汇和语句翻译问题。词汇层的机器翻译错误类型主要集中于术语和准术语翻译错误、普通多义词义项选择错误、机械性选词不当或重复选词、成语和熟语翻译错误等;语句层的机器翻译错误类型主要集中于短句及长句的结构和逻辑混乱、语句语境语义翻译错误、无主句主语还原错误等。除此之外,测试的四个机器翻译平台还存在较严重的漏译问题和大小写格式问题。对于测试的机器翻译平台,其漏译程度除和语料性质及平台本身性能有关外,也和译者输入原文时的单位输入长度有关。译者在输入原文时,以较小单位量逐批向机器翻译平台输入原文,可降低漏译率,但会增加人工工作量。而译文大小写格式问题在有道翻译中最为突出,增加了译后文本清洁的工作量。总的来说,在本次案例文本翻译实践和考察中,搜狗翻译的翻译

效果最佳,其次为有道翻译,再次为百度翻译,最后为谷歌翻译。为进一步提高测评数据的可信性和测评的深度广度,后续研究有待提高测试文本的语料容量和类型多样程度。

参考文献:

- [1] Weaver W. Translation [M] // Locke W, Booth D. Machine Translation of Languages. Cambridge: MIT Press, 1955: 15-23.
- [2] Hutchins W J. Milestones in machine translation No.2: Weaver's 1949 memorandum [J]. Language Today, 1998, 13(6): 22-23.
- [3] 王湘玲,杨艳霞.国内60年机器翻译研究探索:基于外语类核心期刊的分析[J]. 湖南大学学报(社会科学版), 2019,33(4):90-96.
- [4] Chesterman A, Wagner E. Can Theory Help Translators?[M]. Beijing: Foreign Language Teaching and Research Press, 2006: 108.
- [5] 冯志伟. 机器翻译与人工智能的平行发展[J]. 外国语, 2018,41(6):35-48.
- [6] 张政. 机器翻译难点所在[J]. 外语研究,2005,93(5): 59-62.
- [7] 贺文照,李德凤. 英语关系从句机译汉语评价:以谷歌机器翻译为例[J]. 中国科技翻译,2018,32(3):30-34.
- [8] Klubička F, Toral A, Victor M S. Fine-grained human evaluation of neural versus phrase-based machine translation [J]. The Prague Bulletin of Mathematical Linguistics, 2017, 108(1): 121-132.
- [9] 罗季美,李梅. 机器翻译译文错误分析[J]. 中国翻译, 2012,33(5):84-89.
- [10] 李梅,朱锡明. 英汉机译错误分类及数据统计分析[J]. 上海理工大学学报(社会科学版),2013,35(3):201-207.
- [11] 李梅,朱锡明. 译后编辑自动化的英汉机器翻译新探索[J]. 中国翻译,2013,34(4):83-87.
- [12] Newmark P. A Textbook of Translation [M]. New York: Pearson Education, 1988: 22-23.
- [13] 代永强. 浅谈纺织科技英语词汇和翻译的关系[J]. 太原科技,2002(5):58-59.
- [14] Catford J C. A Linguistic Theory of Translation [M]. Oxford: Oxford University Press, 1978: 79.
- [15] Hornby A. Oxford Advanced Learner's English-Chinese Dictionary[M]. 7版. 北京:商务印书馆,2009: 141.

(责任编辑:陈丽琼)