

科技术语汉英机器翻译质量研究

——以测绘学术语为例

周海涛

中国矿业大学外国语言文化学院, 江苏 徐州

收稿日期: 2023年4月10日; 录用日期: 2023年5月22日; 发布日期: 2023年5月31日

摘要

本研究选取谷歌翻译这一主流机器翻译引擎, 以测绘学术语为例, 通过翻译实验考察了科技术语汉英机器翻译的质量。研究发现: 所选术语的机器翻译准确率达到七成, 表明机器翻译确有相当的应用价值。研究重点考查了科技术语汉译英机器翻译中主要存在的五种翻译错误, 分别为术语含义判断错误、译文近义词混用、欠译、词性判断错误以及译文不合行业规范, 并结合实例进行了详细分析。本研究在一定程度上揭示了现有主流机器翻译引擎在科技术语汉英翻译方面的整体翻译水平和问题所在, 能够在译前编辑和译后编辑方面为机器翻译的广大用户提供一定参考。

关键词

科技术语, 汉译英, 机器翻译质量, 谷歌翻译

Exploring the Chinese-English Machine Translation Quality of Technical Terms

—A Case Study of Terms in Surveying Sciences

Haitao Zhou

School of Foreign Studies, China University of Mining and Technology, Xuzhou Jiangsu

Received: Apr. 10th, 2023; accepted: May 22nd, 2023; published: May 31st, 2023

Abstract

Employing Google Translation, one of the mainstream Machine Translation (MT) engines, and taking terms in surveying sciences as an example, translation experiments were conducted in this study to examine the quality of target texts in the Chinese-English machine translation of technical terms.

文章引用: 周海涛. 科技术语汉英机器翻译质量研究[J]. 现代语言学, 2023, 11(5): 2282-2288.

DOI: 10.12677/ml.2023.115308

It is found that: around 70 percent of the terms we selected were accurately translated into English, indicating the considerable value of MT in practice. Particularly, this study focuses on the five major types of translation errors in the Chinese-English machine translation of technical terms as revealed in the translation experiments, including mis-decoding the meaning of the source terms, misuse of synonyms, inadequate translation, mis-decoding of part of speech and translated texts inconsistent with traditions of a specific field, and detailed exposition are also provided in the form of case analysis. This study is meaningful in that, it reveals the merits and demerits of current mainstream MT engines in terms of C-E translation of technical terms. Hopefully, it will shed some light on such aspects as pre-editing and post-editing references for users of MT.

Keywords

Technical Term, Chinese-English Translation, Machine Translation Quality, Google Translation

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

机器翻译(Machine Translation, MT), 是指利用计算机将一种语言符号转换成另一种语言符号的技术[1]。21世纪以来, 机器翻译技术不断进步, 并被广泛应用于以科技翻译为代表的翻译领域, 机器翻译的译文质量在某些领域达到了媲美甚至超过人工翻译的水平[2]。机器翻译的主要用户群体是外语或翻译水平不足, 但有实际翻译需要的外语使用者。以笔者所在的江苏某高校为例, 出于各种原因, 习惯于使用机器翻译将中文写成的学术论文初译成英文, 再委托第三方对机译文进行润色的学生和教师不在少数。笔者在承接的数次理工科学术论文译文润色项目中发现, 机器翻译的译文在英语语法、句法方面已经达到相当高的准确性。然而, 机器翻译译文仍存在一些不完善之处, 其突出表现为受各种原因影响, 暂时无法准确翻译部分学术术语, 以至于在机翻后需要审校者耗费大量时间和精力审查术语翻译的准确性并予以人工校正。科技文本(scientific writing)是“随着科学技术发展形成的一种独立的文体形式, 包括自然科学和社会科学的学术著作、论文、研究报告、专利产品说明书等”[3]。科技文本的一个重要特征是专业术语分布密集。“专业术语构成阐述科学理论的语言基础, 其语义具有严谨性和单一性”[3]。专业术语翻译得是否准确, 关乎科技文本的学术专业性和规范性。姜望琪指出, 术语翻译“应尽可能兼顾准确性、可读性、透明性三个原则”。其中, 准确性即“准确传递原文的意义”; 可读性即“译文应便于使用”; 透明性是指“读者要能从译名轻松地辨认出源词, 能轻松地回译”[4]。为了考察机器翻译能在多大程度上达到术语翻译的标准, 本文尝试选取一种主流的机器翻译引擎, 研究其进行科技术语汉英翻译的翻译质量。国内常见的在线机器翻译引擎有谷歌翻译、有道翻译、百度翻译、DeepL等。其中, 谷歌翻译起步较早, 全球市场占有率较高。同时, 调研得知, 谷歌翻译也是笔者所在高校师生进行学术翻译时主要选用的平台, 故本研究以谷歌翻译为例, 进行在线机器翻译实验。

梳理相关文献可知, 国内针对机器翻译译文质量的研究多从词汇、句法、风格等层面展开, 专门考察术语翻译质量的研究较少。罗季美对汽车技术文献英汉翻译中的术语翻译错误进行了归纳和描述[5]; 也有学者结合计算机技术研究了石油术语英汉机器翻译的质量评价的词向量空间模型方法[6]。可见学界针对科技术语英汉翻译的机器翻译质量有一定涉足, 但对于汉英翻译的研究还比较匮乏。因而, 本文特

对科技术语的汉英机器翻译质量进行考察。

本研究试图回答：1) 谷歌翻译这一机器翻译引擎进行科技术语汉英翻译的英文译文准确率有多高？2) 英译文的翻译错误可以分成哪些类型？3) 根据问题 1)、2) 的答案，可以从译前编辑和译后编辑两个角度为科技文本汉英机器翻译提出怎样的建议？

2. 研究语料来源和步骤

2.1. 语料来源

为评判术语翻译的准确性，研究者不仅要具备良好的双语能力，最重要的是必须具备所研究术语相关学科的专业知识。考虑到笔者本人有过测绘学专业相关课程的学习背景，本研究选取来自《测绘工程专业英语》[7]一书中收录的测绘学术语作为机器翻译实验源语语料。从该书每一章节后所附测绘术语表中随机选取共计 400 条测绘学术语及其标准译文，并整理成各术语及其正确译法一一对应的语料库备用。

2.2. 研究步骤

使用 Python 语言编写简单的计算机程序以调用谷歌提供的谷歌翻译应用程序编程接口(Application Programming Interface, API)，并导入纯文本格式的 400 个中文测绘学术语，通过该自编程序实现调用谷歌翻译进行批量翻译。之后，按“术语原文 - 术语标准译法 - 术语机器译文”的格式制作 Excel 电子表格，然后将全部术语的机译译文和标准译法逐一人工对比，从而评判机译译文的质量。对机译译文与标准译法不一致的术语进行分析，判断其存在的翻译问题的类型，并将涉及同类翻译问题的术语在工作表中进行标记，最后整理归纳各种类型翻译错误所涉术语的数量。

3. 术语翻译错误类型及分析

通过将 400 个术语词条的机译结果与正确译法逐一进行对照，并对准确率情况进行统计可以发现，译文完全准确的术语为 270 个，约占测试所用术语总数的 70%。对于其他存在翻译错误的术语进行排查可知，术语翻译错误主要有五种表现形式，分别为：术语含义判断错误、译文近义词混用、欠译、词性判断错误以及译文不合行业规范(表 1)。下面将结合具体的例子分别对五种翻译错误类型的含义及表现予以解释说明。

Table 1. Different types of terminological translation errors and their percentages

表 1. 术语翻译错误类型及其百分比

	术语含义判断错误	译文近义词混用	欠译	术语词性判断错误	译文不合行业规范
术语数目	58	28	16	13	15
百分比	14.5%	7%	4%	2.5%	3.75%

3.1. 术语含义判断错误

分析发现，对术语含义的误判是术语翻译错误中最主要的类型。在线机器翻译系统生成译文的原理主要有两种：一种是依托后台强大的语料数据库，即对于能在数据库中精确匹配到的词条会直接调用已有的正确译文；而对于数据库中无法准确匹配的词条，如新词、合成词等，则会通过基于概率原理的统计机器翻译模型对要翻译的文本进行语义识别和拆分，然后再对拆分后的短语按目标语的特定规则进行排列，从而实现翻译。表 2 给出了一些术语含义误判的典型例子。

Table 2. Examples of mis-decoding of the meaning of source terms
表 2. 术语含义错误判断示例

中文术语	机译结果	正确译法
① 常差	often poor	constant error
② 碎部测量	crushed measurement	detail survey
④ 边角测量	corner measurement	triangulation
⑤ 导线折角	wire chamfer	traverse angle
⑥ 严密平差	tight adjustment	rigorous adjustment
⑦ 电子手簿	electronic handbook	data recorder

如表 2 所示, 术语“常差”是指测量学中数值和正负号均恒定不变的误差[8], 即译文应采用表示“恒定不变的”之意的形容词 constant 以及表示误差的 error, 然而, 谷歌翻译将该术语的含义错误理解为“经常的贫穷”, 以至给出 often poor 这样的毫不相干的译文。术语“碎部测量”是指测定决定地物地貌轮廓的特征点的平面位置和高程, 并将其绘制成地形图的工作, 也是测绘学中的常见概念[8]。然而, 从译文来看, 谷歌翻译将该术语的含义理解成“破碎的测量”(crushed measurement), 一方面完全曲解了原术语的含义, 另一方面这样的译文本从英语语法角度来看也是毫无根据的。术语“边角测量”是一项利用三角测量(Triangulation)和三边测量(Trilateration), 推求各个三角形顶点平面坐标的测量技术的简称[8], 且英语中有专门表示该概念的合成词 triangulation [7], 但是从谷歌翻译给出的译文来看, 其后台数据库明显未收录“边角测量”这一术语的正确译法, 而是将其含义误读为“角的测量”(corner measurement), 从而导致翻译错误。术语“导线折角”实际上是由两个术语构成的复合词。其中, “导线”是指在导线测量法(traversing)中将一系列测点依相邻次序相连的折线[8], 然而从译文来看, 谷歌翻译显然是将这里的“导线”错误判断为电路中的导线(wire); “折角”则是指折线的夹角(angle) [8], 然而译文中的 chamfer 一词是指切割直角形成的削角并非是夹角, 可见谷歌翻译同样错误判断了原术语的含义。术语“严密平差”表示高精度的、严格的平差测量[8], 因此“严密”对应英文的 rigorous 一词, 而谷歌翻译给出的译文则使用了表示空间上的严密、紧密的 tight 一词, 同样是错误判断了术语的含义。最后一个例子“电子手簿”是指测绘外业测量工作中, 用于存储观测数据并能将数据按规定要求输出的电子记录装置(data recorder) [7] [8]。谷歌翻译在无法调用已有翻译的情况下将其按字面直译成“electronic handbook”(电子手册), 这完全违背了术语的原意。

综合对上述六个例子的分析, 可以看出: 一方面, 谷歌翻译的数据库在术语收录方面还存在不完善之处, 即诸如“常差”、“碎部测量”等测绘学术语目前依然无法通过后台匹配直接得出正确的译文; 另一方面, 可以发现, 谷歌翻译对于未收录的术语目前仍采用的是通过自然语言处理将术语从字面上进行分割后翻译, 然而这种方式弊端较为明显, 或由于语义分割不当导致翻译错误, 或只是机械地采取字面翻译, 对于需要意译、释义或具有约定俗成译法的专业术语仍难以准确译出。

3.2. 译文近义词混用

研究发现, 谷歌翻译在进行科技术语汉英翻译过程中的另一大突出问题是在译文的选词中混用英文的近义词, 从而导致术语翻译错误。中英两种语言都存在“一词多义”和“多词一义”现象, 同一个词放在不同的上下文语境中, 含义也会不同。例如, 中文的“看”在英文中可能对应 see、look、watch 等多个动词; 英文中 lead 一词作动词有“带路”、“通往”、“导致”、“引导”等多个完全不同的意思,

此外还可以作名词，指铅这种物质。像这样的例子不在少数。对于人工翻译来说，“一词多义”和“多词一义”往往只要求译者在思考的基础上选取恰当的译词，因而一般不会导致翻译错误。相比之下，机器翻译引擎主要是依据开发者预设的规则进行匹配，无法像人工译者一样进行思考，因此在翻译数据库中未收录的术语时，译文中混用近义词的现象时有发生。表 3 是该类翻译错误的四个实例。

Table 3. Examples of misuse of synonyms in translation

表 3. 译文近义词混用示例

中文术语	机译结果	正确译法	所涉及的近义词
① 平面测量	plane measurement	plane surveying	survey 和 measure
② 高度角	height angle	elevation angle	elevation 和 height
③ 目标高	target high	target height	height 和 high
④ 人为误差	human error	personal error	personal 和 human

如表 3 所示，术语“平面测量”的本意实际上是指在区域较小的可视地面上开展的测量工作，实际上这一概念的内涵为“测绘”[8]，故这里的“测量”应该译成使用表示“勘测、测绘”的 *surveying*。然而，中文“测量”二字在英文中直接对应的词是 *measure*，但 *measure* 是指对速度、强度、尺寸、含量等具体指标的测量[9]，其含义范围不如“测绘”广，因而“平面测量”翻译成 *plane measurement* 是错误的。术语“高度角”指测绘中从一点至观测目标的方向线与水平面间的夹角[8]。英文中 *height* 和 *elevation* 两个词都可以对应中文的“高度”。但是，这两个词仅仅是字面意思相近，但是其使用场景完全不同。*height* 是普通用词，用于指任何物体可测量到的高度，不受物体本身的高低的限制，既可以用于指人或动物的身高、物体的高度这样的自然属性，也可以用于指具体的高度值，如用作 *at a height of 3000 meters*；相比之下，*elevation* 主要指山丘或耸立在地面上物体的海拔高度，表示(某地的)垂直于海平面的高程[9] [10]。可见，两个概念的内涵和外延均存在明显差异，必须区分使用。考虑到测绘学中主要考察物体相对于参考面的高度，而非的绝对高度，高度角中的“高度”应译成 *elevation*，而不是 *height*。类似地，在翻译术语“目标高”中的“高”这一概念时，谷歌翻译的译文中同样混用了 *height* 和 *high* 两个英语名词。英语中 *high* 作名词用时指“最高水平；最大数量；最高气温”等抽象概念，而 *height* 才是指具体的物理高度[9] [10]，因而“目标高”翻译成 *target high* 也是错误的。最后一个术语“人为误差”又叫个人误差，是指测绘中由个体差异导致的观测误差，其包含“不同的观测者”这层含义[8]，英语的 *personal* 恰能表达这种个体差异性；谷歌翻译给出的译文用的 *human* 一词，*human* 是指相对于动物、机器、上帝等概念而言的人的概念，强调的是人作为物种的同质性，因而无法准确传达“人为误差”这一概念的内涵，因此也属于译文中混用近义词导致的翻译错误[9] [10]。

3.3. 欠译

欠译，又称漏译，是指字面上的翻译程度不够，即源语中该翻译的内容在目标语中未全部译出的翻译错误[11]。例如，术语“地形测量”的被谷歌翻译译作 *topography*，即名词“地形；地貌；地势；地形学”[9]，原术语最核心的概念“测量”(survey)完全没有译出，因而是典型的欠译。

3.4. 词性判断错误

词性判断错误，是指机器翻译引擎错误判断术语的词性或构成术语的各分子概念的词性引起的翻译错误，例如把动词概念误判为名词概念。崔启亮、李闻指出，词性判断错误是机器翻译系统常犯的错

误[11]。例如,在本研究中,测绘学术语“平差值”被谷歌翻译译为 *adjustment value*, 但该术语的正确译法为 *adjusted value*。在测量控制网当中,同一个未知点通过不同的线路推算得到的坐标观测值会有所不同,全部观测值的加权平均值就叫做该点的坐标平差值,即平差值是对数据经过人为平差处理(*adjust*)后得到的数值[8],因而原术语中“值”为名词概念,对应英语的 *value*,“平差”为修饰前者的动词概念,对应英语的动词 *adjust*。在译为英语时,根据两个概念之间存在被动关系,应当用 *adjust* 的过去分词作为定语修饰 *value*,即应该翻译成 *adjusted value*,机器译文中用名词“平差”(adjustment)修饰 *value* 这一译法未能直接体现“平差后的数值”这一准确含义,亦不符合术语翻译约定俗成的原则,故仍然是错误的译法。

3.5. 译文不合行业规范

本研究发现,谷歌翻译在翻译某些未被收录的术语时,未能遵循测绘行业的规范,使用专业的用词,存在“机械直译”的情况。各学科、各行业中的一些复杂概念都常用已有规约的专有词汇进行表达,因而在科技翻译中必须遵守行业的术语规范。例如人口学中 *demographics* 一词虽然只是一个独立的单词,却可表达中文中“人口统计数据”这一由三个子概念复合而成的复杂概念的含义。因此,如果在汉英翻译中将“人口统计数据”机械地直译为 *statistical data of population*,而不直接使用英语固有的 *demographics* 一词,就产生了译文不符合行业规范的问题。在本研究中,译文存在未按照行业惯例使用专有词汇翻译的问题,表 4 给出几个典型例子。

Table 4. Examples of translated texts inconsistent with traditions of specific field

表 4. 译文不合行业规范示例

中文术语	机译结果	原文译法
① 气压水准测量	<i>air pressure leveling</i>	<i>barometric leveling</i>
② 大地方位角	<i>Earth position angle</i>	<i>geodetic azimuth</i>
③ 重力水准测量	<i>gravity leveling</i>	<i>gravimetric leveling</i>

如表 4 所示,“气压的”这一语义在英语中有固定的对应词 *barometric* [9],谷歌翻译将其机械直译为 *air pressure*,未能采用测绘行业已经约定俗成的习惯表达;术语“大地方位角”是指大地测量学中的方位角[8],而“大地(相关的)”这一概念在测绘学中有 *geodetic* 这一固定表达[7],谷歌翻译将其按字面直译成 *Earth* 是不合规范的,另外“方位角”也有专用术语 *azimuth* [7],将其直译成 *position angle* 同样是未采用行业专用表达的体现。对于“重力水准测量”这一术语,测绘学中有专门的形容词 *gravimetric* 表示“重力(分析)的”这一概念[7],而谷歌翻译将其直译为名词 *gravity* (重力),以至最终译文不符合测绘行业已有的表达惯例。

4. 结语

根据上述分析,本研究可以得出的结论主要有:1) 谷歌翻译这一机器翻译引擎在进行科技术语汉英翻译时的整体正确率良好,尤其是对于收录进其后台数据库的术语的翻译准确性较高,因而在翻译实践中具有一定的利用价值;2) 机器翻译进行科技术语汉译英翻译时主要存在五种翻译错误,分别为术语含义判断错误、译文近义词混用、欠译、词性判断错误及译文不合行业规范。其中,前两种是较突出的翻译问题,出现频次相对更高。

基于本研究的发现,我们可作如下推测:现有的机器翻译系统在术语库的建设方面仍有较大完善空

间,尤其是要加强对部分专业性极强的行业术语的收录;现有的机器翻译引擎在术语原文的语义识别方面仍存在不足,以至于经常出现机械直译问题,这需要相关平台的开发者付出更多的努力进行改进。

另一方面,本研究亦能够为采用在线机器翻译引擎进行科技文本汉英翻译的广大用户在翻译原文的预处理方面(译前编辑)提供一定指导。机器翻译用户可以利用好自身的专业优势,在使用机器翻译平台前,先自行标记出或翻译自己可以确定的专业术语,从而能减少以后编辑的工作量。此外,从译后编辑的角度来讲,本研究对术语翻译错误类型的描述、分类,可以为从事科技翻译的译员或译文审校者提供参考,有效提升识别和纠正术语翻译错误的效率。

参考文献

- [1] Locke, W.N. and Booth, A.D. (1956) Machine Translation. *Journal of the Iee*, **2**, 109-116.
- [2] 罗季美, 李梅. 机器翻译译文错误分析[J]. 中国翻译, 2012, 33(5): 84.
- [3] 冯志杰. 汉英科技翻译指要[M]. 北京: 中国对外翻译出版公司, 1998.
- [4] 姜望琪. 论术语翻译的标准[J]. 上海翻译, 2005(S1): 80-84.
- [5] 罗季美. 机器翻译中的术语错译分析[J]. 中国科技术语, 2013, 15(1): 41-45.
- [6] 陈柯, 柴启栋. 基于词向量空间模型的机器翻译质量评价分析——以石油术语有道翻译为例[J]. 中国科技术语, 2022, 24(2): 21-25.
- [7] 尹晖. 测绘工程专业英语[M]. 武汉: 武汉大学出版社, 2013.
- [8] 测绘学名词审定委员会. 测绘学名词[M]. 第3版. 北京: 科学出版社, 2010.
- [9] Turnbull, J. (2012) *Oxford Advanced Learner's Dictionary 9th International*. Oxford University Press, Oxford.
- [10] 陈用仪. 英语常用词疑难用法手册[M]. 杭州: 浙江大学出版社, 2010.
- [11] 崔启亮, 李闻. 译后编辑错误类型研究——基于科技文本英汉机器翻译[J]. 中国科技翻译, 2015, 28(4): 19-22.