

电气工程文本机器翻译的译后编辑例析

陈明芳, 向羽琪

武汉工程大学外语学院, 湖北 武汉

收稿日期: 2023年11月17日; 录用日期: 2023年12月8日; 发布日期: 2023年12月21日

摘要

近年来以百度神经网络为代表的机器翻译发展迅猛, 对于某些正式文本的翻译质量高达70%, 但机器译文仍需人工审校与修改, 因此以机器翻译为基础的译后编辑尤为重要。本文以电气工程学术论文的英汉机器翻译译文为研究对象, 从词汇、句法和语篇三个方面对该类文本机器译文和人工译文进行对比分析。研究发现, 机器译文在词汇层面的错误主要是术语错译、缩略词漏译; 句法层面的错误主要是长难句错译、短语错译、分词错译; 语篇层面的错误主要是句内句际连贯不当。上述问题在译后编辑时通过逆序法、主述分切法等策略运用均得到解决, 改进了电气工程文本英汉机器翻译译文的质量。

关键词

机器翻译, 译后编辑, 电气工程文本

Case Study on Machine Translation and Post-Editing of Electrical Engineering Texts

Mingfang Chen, Yuqi Xiang

School of Foreign Languages, Wuhan Institute of Technology, Wuhan Hubei

Received: Nov. 17th, 2023; accepted: Dec. 8th, 2023; published: Dec. 21st, 2023

Abstract

In recent years, machine translation, represented by Baidu Neural Network, has made rapid advancements, achieving translation quality of up to 70% for certain formal texts. However, machine-translated texts still require manual proofreading and editing, making post-editing based on machine translation particularly important. This study focuses on comparing and analyzing machine-translated and human-translated texts of academic papers in the field of electrical engineering from three aspects: vocabulary, syntax, and discourse. The research findings indicate that vocabulary errors in machine-translated texts primarily involve terminology misinterpretation and omission of abbreviations. Syntax errors mainly include mistranslation of long and complex

sentences, phrases, and word segmentation. Discourse errors mainly manifest as improper coherence within and between sentences. These issues are effectively addressed during post-editing through strategies such as order reversing and subject-predicate separation, thereby improving the quality of English-Chinese machine-translated texts in the field of electrical engineering.

Keywords

Machine Translation, Post-Editing, Electrical Engineering Text

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/>



1. 引言

Warren Weaver 首次提出机器翻译至今已有 70 余年, 随着信息技术不断进步, 近年来机器翻译进入了神经网络机器翻译时代。而科技文本等非文学文本, 用词较为固定, 词汇语义要求精确, 语句结构重复率较高(Hutchins & Somers, 1992) [1], 通常以描述事实、传递信息、知识和观点等为主要交际目的(胡开宝, 李翼, 2016) [2], 适合采用机器翻译。但机器译文仍需人工审校与修改, 这就使机器翻译的译后编辑尤为重要。

国外学者对机器翻译及其译后编辑的研究主要集中在英语和法语、德语、俄语、日语几个语种, 其他散见于英语和越南语、斯洛伐克语、旁遮普语、泰米尔语等。国内一些学者(如罗季美、李梅, 2012 [3]; 李梅、朱锡明, 2013 [4]; 崔启亮、李闻, 2015 [5]; 杨玉婉, 2020 [6]; 陈胜、田传茂, 2021 [7]; 杨文地、范梓锐, 2021 [8]; 夏玲、李宜蔓、李弘武, 2022 [9])对不同类型科技文本的机器译文和译后编辑译文进行了对比研究, 但鲜有以电气工程文本为对象的研究。而目前智能电网在不断建设中, 技术人员国际交流越来越频繁, 因此电气工程领域内的神经机器翻译工作十分必要(陈媛, 陈红, 2023) [10]。鉴于此, 本文以电气工程文本英汉机器翻译译文为研究对象, 从词汇、句法和语篇三个方面对机器译文和人工译文进行对比分析, 并提出相应的译后编辑策略, 以期改进电气工程文本英汉机器翻译译文质量。本文选取的语料均来自电气工程领域的学术论文, 如表 1 所示。

Table 1. Text numbers and titles

表 1. 文本序号与标题

文本序号	论文标题
1	Systematic electromagnetic interference filter design based on information from in-circuit impedance measurements
2	Design and implementation of a 1-V transformer magnetic feedback low-noise amplifier (LNA) at 5-6 GHz, in a 90 nm complementary metal-oxide-semiconductor (CMOS) process
3	DLL-based 4-phase duty-cycle and phase correction circuit for high frequency clock tree
4	Single-ended amplifier-based touch readout circuit with immunity to display noise

2. 电气工程文本机器翻译的词汇错误

2.1 术语错译

术语翻译错误是指句子中术语的翻译出现较大偏离或完全不相关, 在机器翻译中, 多是由词对齐不

当、一词多义造成的。

例 1: (文本 1)

原文: In order to extract the **noise** source impedances, two approaches are available.

百度翻译: 为了提取**噪声**源阻抗, 有两种方法可用。

译后编辑: 为了提取**干扰**源阻抗, 有两种方法可用。

机器译文中出现了一处术语翻译错误, 即将“noise”译为“噪声”, “noise”在此处是一个术语, 查阅 Cambridge Dictionary 后发现“noise”属于一词多义, 在声音、信号、信息三个领域分别有不同的意思, 如“噪声, 杂音, 嘈杂声, 喧闹声(声音); 尤指电子设备的干扰, 干扰讯号(信号); 信息噪音(信息)”, 而例句来自一篇信号处理相关的学术论文, 因此可以确定“noise”在原文中译为“干扰”最为合理。由于机器翻译技术目前的语义处理还不成熟, 所以将多义词判断错误难以避免(崔启亮, 李闻, 2015) [5], 在译后编辑时必须查阅相关术语词典或文献, 对机器翻译中错误的术语翻译进行纠正, 保证科技文本英汉机器翻译译文的准确性、科学性。

2.2. 缩略词漏译

神经机器翻译不再需要词对齐、短语切分等步骤, 但也正由于缺乏“硬对齐”模块, 语义忠实度有所牺牲, 漏译问题普遍存在(郭望皓, 胡富茂, 2021) [11]。

例 2: (文本 1)

原文: To comply with the international regulatory **EMI** requirements, an **EMI** filter is necessary to lower the conducted **EMI** level of the **SMPS** below the limit.

百度翻译: 为了符合国际监管的 **EMI** 要求, **EMI** 滤波器是将 **SMPS** 的传导 **EMI** 水平降低到极限以下所必需的。

译后编辑: 为了符合国际监管的**电磁干扰(EMI)**要求, 需要使用**电磁干扰(EMI)**滤波器来降低**开关电源(SMPS)**的传导性电磁干扰水平, 使其在规定限值以下。

机器译文明显漏译了三处“EMI”和一处“SMPS”, 该论文的关键词已给出“EMI”和“SMPS”的全称分别是“electromagnetic interference”和“switched-mode power supply”, 所以“EMI”和“SMPS”这两个缩略语应当分别译为“电磁干扰”和“开关电源”。科技英语文章常使用缩略词, 构成方式多样, 可以由单词首字母、前几个字母或关键字字母混合组成。要求机器自动还原缩略词生成准确译文, 可行而且比较有效的方法就是建立缩略词辞典(罗季美, 李梅, 2012) [3]。

3. 电气工程文本机器翻译的句法错误

除了改进和优化机器翻译软件本身, 提高机器翻译质量的另一个关键是人进行人工翻译, 因为机器翻译系统在处理长句或复合句时往往会产生句法不通顺、诘屈聱牙的翻译结果。(胡开宝, 李翼, 2016) [2]。

3.1. 长难句错译

对比研究发现, 机器翻译往往遵从原文词序, 忽视了英汉两种语言的特点和差异, 一味地效仿英语句子结构进行汉译。

例 3: (文本 2)

原文: The effect of the product MI_D , **where** M stands for the mutual inductance between the primary (L_D) and the secondary (L_S) inductors of the transformer and is given by Equation (2), **in which** k is the transformer's coupling coefficient, denotes a negative feedback **that** counterbalances the positive feedback through the parasitic capacitance C_{gd} and therefore

neutralises its effect.

百度翻译: 乘积 MI_D 的效应, 其中 M 代表变压器的初级(L_D)和次级(L_S)电感器之间的互感, 由方程(2)给出, 其中 k 是变压器的耦合系数, 表示通过寄生电容 C_{gd} 抵消正反馈的负反馈, 从而抵消其效应。

译后编辑: 方程(2)表示 M 和 I_D 的乘积效应, 其通过寄生电容 C_{gd} 产生负反馈, 从而抵消了正反馈, 其中 M 表示变压器主电感(L_D)和副电感(L_S)之间的互感, k 表示变压器的耦合系数。

这是一个包含三个定语从句的复杂句, 主句抽离出来应为 “The effect of the product MID denotes a negative feedback.” (M 和 I_D 的乘积效应表示负反馈。)在这个长句中, “where” 引导的非限制性定语从句修饰先行词 “MID”, 解释了 M 表示变压器主电感(L_D)和副电感(L_S)之间的互感, 方程(2) $M = k\sqrt{L_D L_S}$ 在前文给出。“in which” 引导的非限制性定语从句修饰先行词 “Equation (2)”, 解释了 k 是变压器的耦合系数。“that” 引导的限制性定语从句修饰先行词 “negative feedback”, 说明 MID 的乘积效应通过寄生电容 C_{gd} 产生负反馈, 从而抵消了正反馈。

再看机器译文, 排除其中的术语错译, 可以发现译文完全仿照英文原文结构, 见了逗号就拆开, 造成语义理解问题, 不符合汉语表达习惯。在译后编辑时, 笔者对机器译文结构进行了调整, 将句末 “that” 定语从句移到最前面, 随后是 “where” 和 “in which” 引导的非限制性定语从句。在进行英汉翻译译后编辑时, 应在理解和忠实原文的基础上, 摆脱原文束缚, 不拘泥于词句, 而是按意思和逻辑关系, 大胆改变原文形式, 通过合并和重组来改善译文。

3.2. 短语错译

3.2.1. 名词短语错译

科技英语广泛采用名词短语使语言简洁。名词短语错译主要发生在翻译结构比较复杂的名词短语时, 如 “形容词 + 多个名词构成定语” 的名词短语及由 “多个名词和现在分词构成定语” 的名词短语等, 如:

例 4: (文本 3)

原文: One of the solutions is using multi-phase signals. But there are some unwanted factors such as jitter, skew, and PVT variations that make distortion of duty-cycle and/or a phase-skew between multi-phase clocks.

百度翻译: 其中一个解决方案是使用多相信号。但也有一些不必要的因素, 如抖动、歪斜和压力、体积、温度变化, 造成占空比失真和/或多相时钟之间的相位歪斜。

译后编辑: 解决方案之一是使用多相信号。但也有一些不必要的因素, 如抖动、倾斜和压力、体积、温度变化, 这些因素都会导致占空比失真和/或多相位时钟之间的相位偏差。

机器译文中将 “skew” 译为了 “歪斜”, 还将 “phase-skew” 译为了 “相位歪斜”, 显然不符合此处科技文本语境和内涵, 查阅词典可以发现 “skew” 一词作为名词有以下含义: 斜角, 倾斜; (对特定人或对象的)偏见; (统计)偏态, 偏斜性。进一步查阅, “Clock Skew” 应该译为 “时钟偏移; 时钟脉冲相位差; 时钟偏斜”, “phase skew” 应该译为 “相位偏差”。

3.2.2. 介词短语错译

in、on、for、with、as 等介词引导的短语在句子中可担任不同语法角色, 如表语、定语、补足语和状语。然而, 在作为定语和状语时, 介词短语容易被误译, 主要表现在位置处理错误、定语误译为状语以及介词短语前的被修饰词错误分离等。如:

例 5: (文本 3)

原文: DLL-based 4-phase duty-cycle and phase correction circuit for high frequency clock tree.

百度翻译: 基于延迟锁相环的高频时钟树四相占空比及相位校正电路。

译后编辑: 基于延迟锁相环的四相占空比和用于高频时钟树的相位校正电路。

原文中“for high frequency clock tree”这一介词短语修饰的是“phase correction circuit”，应该译为“用于高频时钟树的相位校正电路”，机器翻译中“for”介词短语修饰前面的“DLL-based”显然不对。用介词短语来描述方式、手段、速度、时间、步骤等是科技文本的典型特征，对这些介词短语的误译也是机器翻译有待解决的一个问题。

3.2.3. 动词短语错译

机器译文在动词短语翻译上表现出的错误包括将“动词+副词”短语误译为“动词+介词”短语、动词短语中动词的一词多义等。

例 6: (文本 3)

原文: Because of its analog and feedback loop, this type of DCCs achieves higher correcting accuracy, but suffers from relatively slow settling behaviour, loop instability, and charge pump mismatch.

百度翻译: 由于其模拟和反馈回路, 这种类型的 DCC 实现了更高的校正精度, 但存在相对较慢的稳定行为、回路不稳定性和电荷泵失配。

译后编辑: 由于它有模拟和反馈环路, 这种类型的占空比校正实现了更高的校正精度, 但存在相对缓慢的沉降行为、环路不稳定和电荷泵失配的问题。

机器译文中将“settling behaviour”一词译为“定居行为”, 显然与科技文本所阐释内容格格不入, 属于动词短语误译。该文本源自一篇高频电子线路领域的论文, 此处“settling”并不是“定居”而是“沉降”的意思, 故改译为“沉降行为”。

3.3. 分词错译

由于分词比从句更为简洁, 因此在科技英语中被广泛使用, 在句子中可以充当表语、定语、状语、补语和独立结构。

例 7: (文本 4)

原文: To suppress errors in a touch readout circuit due to display noise, differential sensing using fully differential amplifiers is widely used.

百度翻译: 为了抑制触摸读出电路中由于显示噪声引起的误差, 广泛使用使用全差分放大器的差分感测。

译后编辑: 为了抑制触摸读出电路中由于显示干扰引起的误差, 目前普遍采用全差分放大器进行差动传感。

百度翻译将“differential sensing”译为“差分感测”, 查询柯林斯词典可以发现“differential”有“(数学中的)微分”这个义项, “sensing”作动词时表示“感觉到; 察觉到”, 机器译文对这两个词进行了拼凑, 得到“差分感测”这个译文, 但“differential sensing”作为一个专有名词在电气工程领域表示“差动传感”。另外, 还有一处明显错误“广泛使用使用使用”, 这可能是由于原句中有两个表示“使用”的词汇“using”“used”, 其中“using”在此处是现在分词作定语, 修饰前面的“differential sensing”, 表示“使用全差分放大器的差动传感”, “used”在这里是被动语态, 主语是“differential sensing”, 机器译文中的三个“使用”应该是对齐问题导致的。

笔者认为在进行科技文本机器翻译的译后编辑时, 需要特别注意-ing 结构术语, 例如 soldering(焊接), circuit board etching(电路板蚀刻), signal amplifying(信号放大), power supply regulating(电源调节)等, 译

者应当具备足够的知识储备和科技文化底蕴。

4. 电气工程文本机器翻译的语篇错误

语言的逻辑缺层是指不符合常规逻辑推理的语言表达,即在行文中没有形成完整的逻辑链(吴迪龙, 2015) [12]。科技英语中从词句到篇章各个语言层面之间存在着隐含的逻辑关系,不仅涉及语言学,还需要通过逻辑推理和实际分析来深入理解。下面将选取电气工程学术论文中语篇层面的隐含逻辑机译错误进行分析。

4.1. 句内连贯不当

句内连贯不当是指句子内部,即词组或小句之间,在翻译时,忽视了原文与译入语之间的不同,导致句子内部连贯不当。除了 cause、make、result in、result from、lead to、bring about、attribute to 等本身具有较强因果作用的动词外,还有一类动词如 increase、allow、permit、give 等表示的因果意义虽然不明显,但仍使句子表示因果关系[13]。(田传茂,许明武,2000)

例 8: (文本 2)

原文: The lower threshold voltage of the devices **allows** for the supply voltage to be reduced and power dissipation is also decreased.

百度翻译: 器件的较低阈值电压**允许**降低电源电压并且还降低功耗。

译后编辑: **由于**器件阈值电压更低, **因此**可以降低供电电压,并且功耗也会相应减少。

在电气工程学术论文中“allow”一词出现频率较高,虽然“allow”表示的因果意义弱于“cause”“make”等,但仍然表示一种因果关系。百度翻译的译文直接将“allow”译为“允许”,使句子缺少逻辑连词,意义不明确。笔者以“allow”为中心对这个句子进行了切分,译为含有“由于……因此……”的因果句,使句意分明。在处理这类句子时,译者可以采用主述分切法,即将科技术语名词、名词化或名词化结构作为主语,将后面的动宾结构或动宾补结构分开处理,以表达句子的意义。这种句子常常隐含因果关系,可以将其翻译成汉语的因果复句形式。

4.2. 句际连贯不当

所谓句际连贯不当,即翻译时,句子与句子之间由于错用连接词或信息排列次序混乱而导致的句际间的连贯不当。夸克等人(1998)把语篇关联手段分为四种类型:语用和语义暗示,词汇连接,韵律与标点,语法手段[14]。在科技英语中,有些句子之间用句号断开,没有词汇和语法连接标志,而是通过语义暗示句子间的关系。

例 9: (文本 2)

原文: The suitability of this topology for high frequency design, where the influence of the parasitics is more dominant, is stressed. This topology cancels the effects of gate-drain capacitance by adding a signal path through magnetic feedback. Beside the fact that the circuit presents an advanced overall performance, the occupied die area is significantly reduced compared to typical cascode topology.

百度翻译: 强调了这种拓扑结构对高频设计的适用性,在高频设计中,寄生效应的影响更为明显。这种拓扑结构通过磁反馈添加信号路径来消除栅极-漏极电容的影响。除了电路呈现出先进的整体性能之外,与典型的共源共栅拓扑相比,所占用的管芯面积显著减少。

译后编辑: 通过引入磁性反馈信号路径,该拓扑结构可以有效消除栅-漏电容的影响。**因为**在高频设计中寄生参

数的影响更为显著, 所以我们强调了这种拓扑结构的适用性。此外, 与典型的共源共栅拓扑相比, 该电路不仅整体性能出色, 而且所占用的芯片面积也显著减少。

可以发现机器译文一方面缺少主语, 另一方面没有展现三个句子之间的逻辑关系, 结构散乱, 句际间连贯不当。在译后编辑时, 笔者首先调整了句序, 将第二句“This topology cancels...”提到最前面。有些定语从句对先行词的限制修饰作用很弱, 实际上起着状语从句的作用, 含有原因、结果、条件、目的、时间、让步等意义(黄湘, 2002) [15]。笔者将第一句“The suitability of... where...”中“where”引导的非限制性定语从句隐含的因果关系译为含有“因为……所以……”的因果句。最后, 从语义来看第三句“Beside the fact that...”和前两句形成递进关系, 所以笔者使用了“此外”这个连词来体现逻辑关系。只有译者准确理解句子间隐含的逻辑关系, 才能在翻译时使得句子之间逻辑紧密、结构连贯、层次清晰, 这样的译文才能更好地传达原文的意思。

5. 结语

本文选取电气工程文本的英汉机器翻译译文为研究语料, 从词汇、句法和语篇三个方面对机器译文和人工译文进行对比分析, 发现电气工程文本的英汉机器翻译在词汇层面的错误主要是术语错译(如将“noise”译为“噪声”)、缩略词漏译(如“EMI”和“SMPS”), 句法层面的错误主要是长难句错译、短语错译(如将“phase-skew”译为“相位歪斜”)、分词错译(如将“differential sensing”译为“差分感测”), 语篇层面的错误主要是句内句际连贯不当。上述错译现象与罗季美、李梅; 崔启亮、李闻; 杨文地、范梓锐; 田传茂、许明武等的研究结果互补, 这些相似性主要是由科技文本的语言特点、神经网络机器翻译的工作原理等因素决定的。因此, 在译后编辑时应注重电气工程文本英汉术语查询、中英文语序、隐含逻辑等问题, 通过逆序法、主述分切法等翻译策略对机器译文中的错误进行纠正。还应提高电气工程学科素养, 扩充电气工程领域背景知识, 提高译后编辑准确度和效率。目前仍有很多问题有待解决, 比如选取语料较少, 电气工程学科特征融入不足等, 针对以上问题, 可进一步搜集语料, 并结合电气工程领域内的语言学特征进行研究, 以便改进电气工程文本的译后编辑质量。

致 谢

本文是武汉工程大学研究生创新基金(项目编号: CX2022485)、武汉工程大学研究生教育重点项目(编号: 2022ZDXM10)、湖北省教学研究项目(编号: 2022308)、教育部产学研合作协同育人项目(编号: 01020201125EB6Z005)以及“E+”国家人才培养模式创新实验区阶段性研究成果。

参考文献

- [1] Hutchins, W.J. and Somers, H.L. (1992) *An Introduction to Machine Translation*. Academic Press, San Diego.
- [2] 胡开宝, 李翼. 机器翻译特征及其与人工翻译关系的研究[J]. 中国翻译, 2016, 37(5): 10-14.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 杨玉婉. 神经机器翻译的译后编辑——以《潜艇水动力学》英汉互译为例[J]. 中国科技翻译, 2020, 33(4): 21-23+42.
- [7] 陈胜, 田传茂. 在线翻译平台汉英翻译的问题及译后编辑——以石油地质文献为例[J]. 中国科技翻译, 2021, 34(1): 31-34+49.
- [8] 杨文地, 范梓锐. 科技语篇机器翻译的译后编辑例析[J]. 上海翻译, 2021(6): 54-59.
- [9] 夏玲, 李宜蔓, 李弘武. 人工智能背景下科技论文摘要的机器翻译与译后编辑[J]. 编辑学报, 2022, 34(4): 396-401+406.

- [10] 陈媛, 陈红. 融合底层信息的电气工程领域神经机器翻译[J]. 河南科技大学学报(自然科学版), 2023, 44(6): 42-48+6-7.
- [11] 郭望皓, 胡富茂. 神经机器翻译译文评测及译后编辑研究[J]. 北京第二外国语学院学报, 2021, 43(5): 66-82.
- [12] 吴迪龙. 科技英语文章中的逻辑偏移及其汉译处理[J]. 中国科技翻译, 2015, 28(4): 1-4.
- [13] 田传茂, 许明武. 试析科技英语中的隐性逻辑关系及其翻译[J]. 中国翻译, 2000(4): 57-62.
- [14] (英)Randolph, Q., 等. 英语语法大全[M]. 苏州大学《英语语法大全》翻译组, 译. 上海: 华东师范大学出版社, 1998.
- [15] 黄湘. 科技英语隐含因果关系及其翻译[J]. 中国科技翻译, 2002(1): 15-18.