Modern Linguistics 现代语言学, 2023, 11(6), 2406-2412
Published Online June 2023 in Hans. https://www.hanspub.org/journal/ml
https://doi.org/10.12677/ml.2023.116323

Hans汉斯

法治新闻机器英译的译前和译后编辑研究

——基于自建语料库

王宇尧

华北电力大学(保定)英语系，河北 保定

收稿日期：2023年4月18日；录用日期：2023年6月5日；发布日期：2023年6月15日

摘　要

法治新闻的英译传播对我国法治国家形象“自塑”具有重要价值。机器翻译在人工成本、速度等方面具有突出优势，但目前仍与高质量人工译文有较大差距。本文将采用定量研究和定性研究方法，利用自建语料库探讨词汇丰富度、可读性、受动主语和汉语中的零形回指来探讨译前和译后编辑工作需面对的问题，并提出相应对策。

关键词

机器翻译，法治新闻，语料库，译前编辑，译后编辑

“Pre-Editing + MT + Post-Editing” as an Applicable Strategy for Rule of Law News Translation

—Based on Self-Built Corpora

Yuyao Wang

Department of Foreign Languages, North China Electric Power University, Baoding Hebei

Received: Apr. 18th, 2023; accepted: Jun. 5th, 2023; published: Jun. 15th, 2023

Abstract

The English translation and dissemination of law news plays a significant role in China’s self-portrayal as a country under the rule of law. Machine translation has outstanding advantages in labor cost and efficiency while there is still a considerable gap between it and high-quality human trans-

文章引用: 王宇尧. 法治新闻机器英译的译前和译后编辑研究[J]. 现代语言学, 2023, 11(6): 2406-2412.
DOI: 10.12677/ml.2023.116323

lation. This paper will employ quantitative and qualitative research to discuss the issues that editors may face in pre-editing and post-editing, including lexical richness, readability, patient subjects, and zero anaphora in Chinese by self-built corpora. Corresponding strategies will also be provided.

Keywords

Machine Translation, Rule of Law News, Corpus, Pre-Editing, Post-Editing

1. 引言

20 世纪三十年代初，法国科学家阿尔楚尼(G.B. Artsouni)就提出了使用机器进行翻译的想法，并以此申请了一项名为“机械脑”(mechanical brain)；但真正意义上的“机器翻译”这一概念是由瓦伦 •韦弗(Warren Weaver)于 1947 年提出[1]。机器翻译在经过十年左右的初步发展期后，由于在此领域的收入和产出明显失衡，机器翻译研究进入了一个较长的停滞阶段。之后在加拿大、法国于机器翻译应用实践的推动下，机器翻译缓慢发展。90 年代，在语料库这一辅助手段进入了人们的视野后，机器翻译在一定意义上实现了突破性的进展，基于语料库的机器翻译逐渐得到了应用。

语料库翻译学是指采用语料库方法，在观察大量翻译事实或翻译现象并进行相关数据统计的基础上，系统分析翻译本质和翻译过程的研究[2]。将二者进行结合，语料库可以汇总海量法治新闻语言使用情况，并且通过语料分析软件的辅助下多角度展开分析。

2. 研究目的

依法治国，为社会繁荣发展和长治久安是国家治理领域方面的一场深刻改革。在这条道路上，法律类新闻作为信息输出的主要渠道，在全球化进程中扮演着重要角色。而海外媒体在西方政治话语的影响下，在新闻话语中常对我国法治层面表现出意识形态偏见。因此，翻译我国法治新闻，让更多人了解中国是我国推进国家形象话语能力建设的必经之路。

近年来，随着人工智能的发展，神经网络机器翻译能够“利用人工智能模仿大脑神经元进行语言翻译，以端到端的方式进行翻译建模”[3]，大大提高了机器翻译译文质量，但与人工译文在一些垂直领域仍有差距。虽然机器翻译存在不足，但如果以纯人工形式对法治新闻话语进行翻译，那么必然需要耗费巨大的人工和时间成本。因此，有学者提出，“人机共译”这一将机器智慧与人工结合，平衡机译高效率与人译高质量，组成“稿件－机译－人译”的产出链的一种智能翻译模式可有效满足翻译行业大规模高效工作的需要[4]。而这种“译前编辑 + 机器翻译 + 译后编辑”的模式也逐渐成为语言服务行业的重要工作模式。所以，本文将通过自建语料库定量研究中文原文机器译文和原生法治新闻话语差异，同时通过汉英平行语料库辅助找出机器译文存在的不足，发掘计算机翻译存在的共性问题，以期帮助译前及译后编辑的工作人员针对典型问题更高效地开展审校工作。

3. 研究方法

为后续开展研究工作，笔者收集了来自央视网和人民网近三年的中文法治新闻文本，汇总利用代表

性神经网络翻译机器——谷歌翻译引擎进行英译。同时从 Jurist，Law & Crime，The Law Society Gazette，JDSURPRA，THE NATIONAL LAW REVIEW 等 5 个英文法律类新闻网站的搜集英语新闻语料，使用文本整理器进行语料清洗，分别搭建起库容为 57,803 个形符的汉语法治新闻语料库(Corpus of Chinese Legal News，简称为 CCLN)，库容为 66,785 个形符的英译法治新闻语料库(Corpus of Translated Legal News，简称为 CTLN)及库容约为 41,666 个形符的英语法治新闻语料库(Corpus of English Legal News，简称为 CELN)。

语料库创建后，为进行中文语料处理，笔者采用 CorpusWordPaser 进行中文分词。三个语料使用 CUC_Paraconc，AntConc 进行语料检索，语料分析工具主要使用 WordSmith 4.0。

4. 研究发现

4.1. 词汇丰富度

在语料库语言学中，形符(token)是指语料中出现的所有词，包括重复出现的词次；类符(type)则是语料库文本中出现的任何一种独特的词性。词汇密度最早由 Ure 于 1971 年提出，而类符/形符比(type/token ratio, TTR)常被用做判断词汇密度的方法[5]，但由于随着采集文本数量的增多，大量功能词在文中反复出现，形符不断增加，但类符增长速度相较于形符往往更慢。以此推理，文本长度越长，形符越多，类符/形符比越低，得出的数据就有失准确性。因此，标准化类符/形符比(standardtype/tokenratio, STTR)，即计算每一千词的 TTR，最后对所有结果进行均值处理。把两个语料文本导入 wordsmith 4.0 后，得到以下数据：

Table 1. Comparison of the overall characteristics and lexical richness of the corpora of CCLN, CTLN, CELN
表 1. CCLN，CTLN，CELN 三个语料库总体特征和词汇丰富度对比

项目文本	汉语法治新闻	英译法治新闻	英语法治新闻
形符数	57,803	66,785	42,994
类符数	729	5849	6297
类符/形符比	13	9	15
标准化类符/形符比(STTR)	45.96	37.30	43.48
STTR 标准差	50.29	59.07	53.99

我们可以看出，CCLN 原文语料库标准化类符/形符比略高于 CELN 可比语料库(见表 1)，表示汉语的词汇丰富度更高。在这一前提下，反观机器英译的中文法治新闻，虽然形符较 CCLN 原文语料库更多，即翻译后整体篇幅更长，但类符数却低于中文文本；同时，CTLN 译文语料库的 STTR 为 37.30，远低于 CCLN 原文语料库的和 CELN 可比语料库。为进一步探析语料库的 STTR 存在较大差异原因，选取 serious(以及其副词形式)和 special 进行举例说明(为减少歧义，例文主要改译待分析词汇以及明显误译，例句他处可能存在一些错误)，其中译后编辑的译文表达主要来自于 China daily 新闻网站：

Serious(ly)

1) 严重

原文：严重危害公众身心健康和生命安全

机器译文：Seriously endangering the physical and mental health and life safety of the public.

2) 严峻

原文：要高度重视黑恶犯罪向网络发展蔓延的严峻态势

机器译文：We should attach great importance to the serious situation of the development and spread of gangster crimes to the Internet.

谷歌翻译引擎将严重、严峻两个词都处理成了 serious 或 seriously，但如果为避免新闻用语的重复，“严峻态势”可改译为 grim trend，“严重危害”可改译为 critically endangering。

Special

1) 专门

原文：目前我国并没有一部针对虚拟财产保护的专门立法。

机器译文：At present, there is no special legislation on the protection of virtual property in my country.

译后编辑的译文：At present, there is no specialized on protecting property in China.

机器翻译大多采用直译，很难根据逻辑调整译文。原文中“一部”所修饰的应是某一部专门的法律而不是“立法”这一动词。在 China daily 中，“为……专门出台的法”表述为“the specialized law on something”。

2) 专项

原文：“扫黄打非”部门开展系列专项整治。

机器译文：The “anti-pornography and anti-illegal” department launched a series of special rectifications.

译后编辑的译文：The Department Against Pornographic and Illegal Publications launched a series of targeted crackdown operations。

将“专项整治”译为 Special rectifications，虽然在词汇意义上没有问题，但 targeted 更强调目标性；同时在“扫黄打非”这个语境下，这项整治行动更偏向于打击、制裁(crackdown)，而机器译文中的 rectification 更强调改正、矫正之意，程度较轻。

从以上两个例子的分析，我们可以看出：在翻译过程中，计算机较为缺乏变通思维，难以做到为避免重复而换词使用，并且因为在定向思维系统的局限下，多采用直译这一翻译策略，其中可能出现一些词语的错译。就 CTLN 译文库与 CELN 可比语料库在 STTR 的差异与新闻写作特点而言，大部分新闻记者为提高读者阅读兴趣在新闻话语中可能会从自身单词储备中提取近义词并尝试换词表达，词汇丰富度也相应更高。

4.2. 可读性

经过机器英译后，CTLN 译文库形符的增加和类符的减少可以表明译文使用了更多的功能词对实词进行衔接，来达到阐释词汇密度相较于英文更高的中文文本的目的。如从更深层次地探讨，CTLN 译文库的可读性可以从平均句长、平均词长进行考虑。

首先，句子越短，读者需要花费在句法理解的时间相对较少；长度越短的单词，读者可能掌握的概率越大。因此，CTLN 译文库和 CELN 可比语料库的平均句长和平均单词长度可以作为文本理解难度的标准之一。

就平均句长而言，利用 Word 的查找功能检索总句子数，后将将两个文本形符数除以句子数，可大致得出 CTLN 译文库和 CELN 可比语料库平均句长分别为 24.56 和 23.77。因此，CTLN 译文库的平均句长略高于 CELN 可比语料库，可读性相应降低。此外，通过 WordSmith 4.0 进行计算得出，CTLN 译文库和 CELN 可比语料库的平均单词长度均为 5.00，但 CELN 可比语料库平均词长的标准差为 2.99，略高于 CTLN 译文库的 2.95，见表 2。标准差是一组数据分散程度的度量，标准差越大，则大部分数值与平均数值的离散程度越高。即，在平均词长近于相等的情况下，CELN 可比语料库的大部分单词词长上下浮动范围较大。因此，平均词长这一因素参考价值较弱，我们需要从其他角度了解单词难度。

Table 2. The comparison of mean sentence length, mean word length, and their standard deviation between the corpora of CTLN and copora of CELN
表 2. CTLN 译文库与 CELN 可比语料库的平均句长、平均词长及其标准差对比

项目文本	英译法治新闻	英语法治新闻
平均句长	24.56	23.77
平均词长	5.00	5.00
平均词长标准差	2.99	2.95

迷雾指数(The Gunning FOG Index)是由美国罗伯特·冈宁教授提出的一个分析测量文章阅读难度的工具，该指数主要用于测算读懂语料的读者需要接受的正规教育年限。指数越高，读懂需要的教育程度越高，即阅读难度越大[6]。计算公式为：

$$\text{迷雾指数} = 0.4\left[\left(\frac{\text{文本单词总数}}{\text{文本句子总数}}\right)+100\times\left(\frac{\text{文本中长单词数量}}{\text{文本单词总量}}\right)\right]$$

结果显示，CTLN 译文库的多于三个音节的长单词达 15,133 个，迷雾指数为 18.11；CELN 可比语料库的长单词有 8780 个，迷雾指数为 17.79。

综合平均句长、平均词长以及迷雾指数差异来看，CTEN 可比语料库的可读性高于 CTLN 译文库。考虑到机器译文和汉语的标点保持一致的倾向和汉语新闻语篇含有多个分句的长句，进行译前编辑时提前断句、添加主语，增加译文可读性；同时，针对平均词长的差异，可利用检索工具筛选出长音节单词，判断其是否符合目标语习惯用法，如存在差异，可针对某个不符合用法的长单词进行批量换词。

4.2. 受动主语的缺失

英语是主语突出的语言，除祈使句、省略句、命令句外，英语句子都需要一个主语。如果没有特定的人或物作为主语，也需要用形式主语 it 作为替代。然而，汉语是主题突出的语言，造句并不强调句子形式的完整，不要求一定要有主语，所以往往“形散意连”[7]。

受动主语是动作的作用对象，是动作的受影响者。由于法律英语具有正式和注重事实的特点，法律英语中大量使用受动主语[8]。在中文法律条文中，同样也存在大量的受动主语，但是缺少明显的形式标记。机器的翻译又是基于原文本运行的，因此在汉译英法治新闻中经简化的法律条文时，译文可能在源语特点导向下，出现受动主语缺位的情况。收集的语料中出现了多条简化的法律条文，此处以语料库中《治安管理处罚法》一则精简后的条文为例：

原文：偷窥、偷拍、窃听、散布他人隐私的，处五日以下拘留或者五百元以下罚款。

机器译文：Peeping, secretly filming, eavesdropping, or spreading the privacy of others shall be detained for less than five days or fined less than 500 yuan.

译前修改的原文：对于偷窥、偷拍、窃听、散布他人隐私的人，处五日以下拘留或者五百元以下罚款。

经译前编辑后的译文：**A person who** peeps, secretly takes photos, eavesdrops, or spreads the privacy of another person shall be detained for not more than five days or be fined not more than 500 yuan.

在原文中，因“偷窥、偷拍……”而被处以罚款或拘留的受动主语在汉语中没有明确形式标注，机器译文也出现了明显的语法错误。针对这种机器翻译容易出现的语法错误，可以在缺失成分的译文句中添加受动主语的对动词形式进行适当的转换。其次，也可以采用译前编辑的方法，即在要翻译的原文上先进行修改后进行翻译。“动词 + 的”这一形式在汉语中往往代指正在做这个动作的人或集体，而的“的”

后的人称却不一定有。因此，在译前翻译时对这类用法加以关注，把句子成分补充完整，以减少机器翻译的障碍，增强准确性。

4.3. 零形回指

零形回指，指“在语言表达中再次提到某个指称实体时，采用零形式进行指代，表面上没有具体的语言符号或语音形式的指称现象[9]。”如，“同时，智能手表的一些智能卖点，也会在无形中淡化孩子的主动思考能力，如(a)遇到不会的字和词，(b)问一句(c)便可呈现答案。”在这一句中，括号里的内容都为空，但存在语义内容，我们可以推断出(a)和(b)都指孩子，(c)指智能手表，三者均属于零形回指。在一些较短的或者逻辑较为简单的中文句子中，零形回指对机器翻译造成的障碍较小。但如果翻译对象分句众多，零形回指使用频繁，机器译文中则可能出现逻辑混乱、指代不明、句子成分缺失的问题[10]。以CCLN 原文库的一段文本为例(a，b，c，d 四处加粗内容均为译前补充，其中 d 为根据语境推测进行的补充)：

原文及补充内容：被告人利用职务便利或职权、地位形成的便利条件，**a. 他**通过其他国家工作人员职务上的行为，为他人谋取利益或不正当利益。**b. 被告人**非法收受他人财物，c. **财物**数额特别巨大，d. **司法机关**依法应当以受贿罪追究其刑事责任。

无译前编辑译文：The defendant took advantage of the convenience of the position or the convenient conditions formed by the authority and status to pass other national staff’s posts. Acts of seeking benefits illegitimate benefits for others, illegally accepting other people’s property, the amount is particularly huge, and should be investigated for criminal responsibility for the crime of accepting bribes according to law.

译前编辑后译文：The defendant took advantage of the convenience of **his** position or the convenient conditions formed by **his** power and status **to seek benefits or illegitimate benefits for others through the behavior of other national staff in his position**. The defendant illegally accepted other people’s property, **and** the amount of property was particularly huge, and t**he judiciary** should investigate his criminal responsibility for the crime of accepting bribes according to law.

没有经过译前编辑的译文，有 3 处主要错误：

1) 介词错译为动词。“通过其他国家工作人员职务上的行为”中的“通过”为介词，表示以某种人或手段作为媒介达成目的；但译文中直接将其译成了 pass。因为机器在这一缺少指代的长句中错误理解了词性，导致句意出现较大偏差。

2) 语法错误。机器一文中“Acts of…others”为名词短语，“illegally…property”为现在分词结构，两者均无法单独构成分句，但机器译文在没有连接词的情况下，用逗号将这二者连接，在语法层面上较为不妥。

3) 主语错误。结合机器译文第二句的结构来看，“should be investigated…to law”的主语为“the amount”，但这一段整个话题链的主语应为被告人，被调查的对象也应是被告人。

但经过译前编辑的译文质量有了明显的提高，在语义和句法层面都没有明显失误。因此，译者需要对分句多、零形回指使用频繁的句子加以格外注意，必要时可以进行译前编辑。从先行词中提取句子成分或结合语境推测缺失成分并补充到句中可以有效减少译后编辑整体工作量、提高翻译效率。

5. 结语

随着全球化趋势的加快，法治新闻的英译传播在展现中国法治文化、国际形象自塑上扮演着十分重要的角色。同时，法治新闻的专业性是法治新闻英译的一大挑战，教育性是影响其对外传播有效性的一

大因素。本文从四个方面分析了法治新闻机器翻译的译前和译后编辑所面临的问题，并提出了相应的解决方案，旨在能够为广大翻译工作者提供一定帮助，推动中国法治新闻传播行稳致远。

参考文献

[1] 高璐璐, 赵雯. 机器翻译研究综述[J]. 中国外语, 2020, 17(6): 97-103.

[2] 胡开宝. 语料库翻译学: 内涵与意义[J]. 外国语(上海外国语大学学报), 2012, 35(5): 59-70.

[3] 戴光荣, 刘思圻. 神经网络机器翻译:进展与挑战[J]. 外语教学, 2023, 44(1): 82-89.

[4] 肖凤华, 殷白恩. 人机共译交互平台在工程翻译中的运用——以传神语联网为例[J]. 中国科技翻译, 2019, 32(3): 38-40+59. http://doi.org/10.16024/j.cnki.issn1002-0489.2019.03.012

[5] 黄智欣, 刘著妍. 基于语料库的英语新闻导语文体特征分析[J]. 武汉冶金管理干部学院学报, 2020, 30(3): 91-93.

[6] 部寒. 中美上市公司年报话语质量对资本市场反应的影响对比研究[D]: [博士学位论文]. 北京: 对外经济贸易大学, 2019.

[7] 朱晓敏. 基于自建语料库的政治文本英译特点研究[J]. 解放军外国语学院学报, 2011, 34(3): 73-77.

[8] 杨宇. 法律文书汉译英的主语选择[D]: [硕士学位论文]. 武汉: 华中师范大学, 2012.

[9] 蒋平. 零形回指现象考察[J]. 汉语学习, 2004(3): 23-28. http://doi.org/10.3969/j.issn.1003-7365.2004.03.005

[10] 侯敏, 孙建军. 汉语中的零形回指及其在汉英机器翻译中的处理对策[J]. 中文信息学报, 2005, 19(1): 14-20.