

基于大语言模型的中国特色话语翻译语言特征研究 ——以党的二十大报告日译为例

上海对外经贸大学 国际商务外语学院 白 杨
吉林外国语大学 人工智能学院 金明哲

[摘 要] 以党的二十大报告日译本为参照,选取5个主流神经机器系统(NMT)与5个大语言模型(LLMs)生成的译文为分析对象,运用文本挖掘工具与编程语言对其进行计量分析、可视化呈现及自动质量评估。研究发现,NMT与LLMs译文在文本特征上与人工译文存在显著差异,表现为动词使用偏低、名词化倾向显著、汉字词及句式存在机械对应现象。基于BLEU、TER与BERTScore的自动评估进一步显示,机器译文虽在基本语义传递上达到一定水平,但在处理文化负载词、四字格及对仗句式时易出现机械直译、表达僵化等问题;在修辞识别与情感色彩传达方面亦表现欠佳,导致译文在交际效果上存在明显局限。针对中国特色话语中译日任务,当前NMT与LLMs译文质量难以达到专业翻译所要求的标准。

[关键词] 二十大报告 日译本 大语言模型 文本挖掘 翻译质量评估

引 言

中国共产党第二十次全国代表大会上的报告蕴含丰富的中国特色话语,其外译工作对于在国际语境中信息的准确传递与文化内涵的有效沟通至关重要,是中译日对外宣讲、讲好中国故事,构建国际话语体系的最紧要、最迫切的课题(李运博、卿学民,2025)。实现二十大报告的高质量翻译不仅要求准确传递原文信息,更需有效转化中国式的文化内涵与价值观,以促进深层次的跨文化沟通与认同。

随着神经机器翻译技术(Neural Machine Translation, NMT)的成熟与大规模应用,以及大语言模型(Large Language Models, LLMs)在自然语言处理领域的快速发展,各类翻译工具在语言服务中的作用日益凸显。然而,在面对党政文献这类兼具政治严肃性、文化独特性与修辞复杂性的高要求翻译任务时,NMT与LLMs能否生成符合专业标准的译文,目前尚缺乏基于系统量化的实证研究。

为探究上述问题,本研究采用计量语言学分析与自动评估相结合的研究路径,以党的二十大报告中译日任务为案例,综合运用文本挖掘工具(MTMineR_2023)、日语分词工具(MeCab),以及编程语言(R、Py-

[收稿日期]2025-12-26

[基金项目]本文为2023年度教育部人文社会科学研究一般项目(青年基金项目)“《伤寒论》在日本的翻译、接受与影响研究”(项目批准号:23YJC-ZH002)阶段研究成果

[作者信息]白杨,女,上海对外经贸大学国际商务外语学院讲师,研究方向:中日语言文化交流、中国特色话语翻译,电子邮箱:bai@suibe.edu.cn;
金明哲(通讯作者),男,朝鲜族,吉林外国语大学人工智能学院教授、日本先端科学大学综合研究所特任教授,研究方向:数据科学、文本挖掘及分析、人工智能,电子邮箱:singane@gmail.com

thon),对人工参考译文、主流NMT系统及代表性LLMs生成的译文进行深度文本挖掘、计量统计与可视化分析。同时,引入BLEU、TER、BERTScore等多元自评估指标,从形式对应、编辑代价及语义保持等多个维度量化评估译文质量。本研究旨在通过系统的数据挖掘与对比分析,揭示NMT与LLMs在翻译中国特色话语时呈现的语言特征差异与质量表现,以期对相关领域的机器翻译质量评测、译后编辑实践、翻译教学及技术优化路径提供基于实证的参考依据。

1 文献综述

1.1 中央文献外译研究

中央文献饱含中国特色话语,承载着中国特色、中国风格、中国气派,是构建党和国家话语体系及传播中国声音的重要载体。其外译是向世界“讲好中国故事,传播好中国声音,展示真实、立体、全面的中国”的最后一公里(黄友义、李晶,2022)。鉴于此,学界围绕中央文献外译展开了广泛探讨,研究已覆盖英、日、俄、法、德、西等多语种及少数民族语言。同时,学界借助功能对等、生态翻译学、目的论、语料库、对外翻译等多元视角,在翻译路径、翻译策略选择以及翻译规范等方面取得了丰硕成果。

在众多语种中,相较于成果相对富集的英译研究,日译方向亦受关注且进展显著。现有研究主要聚焦于四个维度:其一为语言本体与修辞策略层面,包括系统总结了政经术语及文化固有词的翻译方法(修刚、米原千秋,2016;李钰婧,2019;董海涛,2022),积累了丰富的四字格、熟语等中国特色话语的日译经验(李玥,2019;刘健,2019;张桂丽等,2023),并对隐喻等修辞翻译策略进行了探讨(荆瑶,2021;邓文肫,2025);其二,在翻译规范与质量优化层面,既有关于提升译文可读性的句法优化策略的考察(陈岩、孟海霞,2018),又有基于翻译规范理论的词汇修订效果的评估(于栋楠,2024);其三,在受众认知与传播效果层面,研究视域由“译者中心”转向“受众关照”,日本受众对政治著作日译本的接受心理及认知语境特点得到系统考察(薛悦、修刚,2022;薛悦,2023);其四,在宏观理论建构层面,李运博、卿学民(2021、2025)强调译者应在政治文献翻译中构建叙事意识实现政治性与文学性的有机结合,修刚等(2018)探讨了逻辑关系梳理在跨文化交际中的战略价值。此外,在方法论上,语料库驱动的定量分析(童富智、修刚,2020)与可视化工具的应用(王俭、姜龙范,2021)为翻译共性及语言特征研究提供了实证支撑。综上所述,当前中央文献日译研究呈现出理论视角多元化、研究对象精细化、研究方法实证化的特点,这不仅为构建融通中外的中国特色对外话语体系提供了重要参考,也为今后多语种对比及智能化翻译研究奠定了学理基础。

1.2 人工智能大语言模型翻译质量评估研究

随着大语言模型(LLMs)的迭代升级,生成式AI为翻译实践提供了自动化与语境优化的新路径,极大地提升了翻译效率。基于此,学界已开始聚焦于LLMs与NMT,以及人类译文的对比研究与质量评估。

在汉译日机器翻译研究方面,毛文伟、朱海莹(2024)基于计量文体学方法,以《阿Q正传》日译本为参照,系统考察了Google翻译与GPT-3.5/4o译文的语言特征。研究发现,二者在词汇多样性、词语搭配、修饰表现等方面均存在不足,尤其是精确度不足、语气平淡、感情色彩匮乏,难以再现人工译本的生动与自然(毛文伟、朱海莹,2024)。研究进一步指出,这些问题正是译后编辑过程中亟需关注和改进的环节。在日译汉机器翻译质量评估方面,王羽迪(2025)以GPT-4o模型为对象,对其在不同类型文本日译汉任务中的翻译表现进行了评估,并认为GPT-4o在信息型文本与正式性文本的处理中表现突出,然而在表情型文本与感染型文本中表现欠佳,尤其是涉及修辞、情感色彩及文化隐喻的内容,译文质量明显不足。该研究不仅揭示了GPT-4o在不同文本类型中的差异化特征,也为理解其在日译汉任务中的优势与局限提供了实证依据。

尽管机器翻译质量研究已成为学界热点之一,但当前NMT与LLMs应用于中国特色话语翻译的质量评估仍处于起步阶段。既有研究多聚焦于汉译英任务,如文旭、田亚灵(2024)选取党的二十大报告中10个内含政治隐喻的例句,对GPT-4o与NMT(Google Translate、Youdao Translate、DeepL Translate)英译进行了对比分析,结果显示,尽管GPT-4o在整体表现上优于NMT,但其在处理意识形态表述、复杂句式、文化负载词、以及比喻和转喻的识别与翻译方面仍存在明显不足。

总体来看,现有中国特色话语机器翻译质量评估研究主要集中于汉译英方向,针对汉译日的研究仍较为匮乏。相较于汉译英任务,中国特色话语的汉译日在语言结构与语用表达层面呈现出更为突出的复杂性。具体而言,日语作为黏着语,具有主语频繁省略与高度语境依赖的特征,其语体系统更为精细,常通过委婉表达与语气缓和实现语用调节。这些特征使得机器翻译在信息捕捉与传递、逻辑关系理解与重构以及译文可接受性等方面面临更大挑战。中国特色话语的日译是“讲好中国故事、传播好中国声音”的关键环节,为此,有必要立足汉日语言差异,对机器与大语言模型下中国特色话语日译质量开展系统性评估研究,从而为提升中国特色话语日译的忠实度与地道性提供实证依据,并进一步推动中国特色话语外译多语种评价体系的细化与发展。

2 研究设计

2.1 理论基础

党的二十大报告是我国重要的政治传播文本,其外译超越了单纯的语言形式转换,本质上是一种政策理念、国家形象与意识形态的跨文化传播行为。因此,译文分析不仅要关注语言忠实度,更需兼顾其在目标语文化语境下的交际效果,以实现讲好中国故事的目标。本研究以纽马克(Peter Newmark, 1981、1988)语义翻译与交际翻译理论为基础框架。纽马克指出,所有翻译策略均可置于语义翻译和交际翻译这两个基本取向所构成的光谱上,并提出了从最侧重源语言的一端到最侧重目标语言一端的八种翻译方法:逐字翻译(Word-for-word translation)、直译(Literal translation)、忠实翻译(Faithful translation)、语义翻译(Semantic translation)、交际翻译(Communicative translation)、地道翻译(Idiomatic translation)、意译(Free translation)和归化(Adaptation)(彼得·纽马克,2001:45-47)。通过将这八种翻译方法映射至语义翻译和交际翻译维度,能够有效辨识译者或机器翻译在不同策略维度上的倾向,从而为译文特征分析与质量评估提供理论依据。

2.2 研究问题

本研究旨在讨论以下三个问题:第一,在中国特色话语汉译日任务中,LLMs生成译文的翻译质量如何?第二,NMT和LLMs译文与人工译文相比表现出哪些差异,LLMs是否已具备替代人工译者的可能性与条件?第三,在对NMT和LLMs生成的译文进行译后编辑时,需要注意哪些方面?

2.3 文本准备及语料采集

本研究以党的二十大政治报告及其官方日译本(中央党史和文献研究院,2023)作为源文本与参考文本(HT)。该译本由中央党史和文献研究院翻译并审定,语言准确、术语统一、风格严谨,具有高度的政治权威性与学术规范性,是本领域重要的规范译本,用作机器翻译质量评估的参照标准具有充分的合理性与可靠性。候选文本包括两类:其一为5个大语言模型生成的译本样本,分别取自 Gemini、ChatGPT-5、ChatGPT-4o、Qwen3、Wenxin4.5^[1];其二为5个神经机器翻译工具的译文,来源于DeepL Translate、Google Translate、Baidu Translate、Youdao Translate、Niu Translate。

为保证数据可比性,LLMs译本的生成采用了统一的翻译提示语(prompt,如下),并要求输出文体与官方译本保持一致:

你是一位资深翻译工作者,擅长中文和日文的互译。请将下面的中文翻译成日文。要求:用词准确,不丢失原文信息;在保留原文风格的基础上,选择自然的日语表达;文体用日语简体。

NMT译文则依据平台输入限制,以段落为单位切分源文本后逐段获取。所有译文均以文本格式保存,并按预定规则进行编号。获取的全部文本(参考文本与10份候选文本)均使用文本分析工具MTMineR_2023进行预处理,为后续的量化统计与可视化分析提供基础。

2.4 研究方法

本研究对译文的量化分析主要包含文本分析统计与翻译质量评估两个环节。首先,使用日语分词工具MeCab对所有文本进行分词;随后借助文本分析工具MTMineR_2023完成数据清洗、特征提取与基础统计分析;最后基于R、Python等编程语言进行进一步的量化计算与可视化呈现。

翻译质量评估遵循“参考译文最佳相似性”原则,通过度量候选译文与参考译文之间的相似度进行评估。当前主流自动评估指标可归为三类:基于N-gram相似性、基于编辑距离以及基于预训练语言模型(如BERT)的语义向量相似性(管新潮、陆晓蕾,2022:146)。在基于N-gram相似性的指标中,BLEU(Bilingual Evaluation Understudy)应用最为广泛。它通过计算候选译文与参考译文在n元语法层面的重合度,评估译文的忠实度与流畅性,BLEU值越高通常表明候选译文在字词形式上与参考译文越接近。基于编辑距离的代表性指标是TER(Translation Edit Rate)。其核心是计算将候选译文转化为参考译文所需的最小编辑操作次数,并将结果标准化为“每参考词的平均编辑率”。TER值越低意味着译后编辑成本越小,通常对应更高的翻译质量。

国内相关研究多以BLEU为主要评估依据,或辅以TER进行验证。例如,毛文伟、朱海莹(2024)及王羽迪(2025)均以BLEU值为核心评估标准,文旭、田亚灵(2024)则同时采用了BLEU和TER两项指标。然而,BLEU主要关注字面重合,难以捕捉深层语义等值与句法连贯;TER虽能反映译后编辑成本,但同样无法有效衡量语义一致性。BERTScore通过BERT模型获取参考译文与候选译文中各token的语义向量,并基于余弦相似度进行贪心匹配,最终以F1值综合反映两者的语义相似程度。该方法能够识别同义表达、语序调整及语义重述,在语义层面提供了有效的评估维度,能够有效弥补BLEU和TER的不足。BLEU侧重表层形式一致性,TER反映译后编辑成本,BERTScore则衡量深层语义相似性,三者各具侧重,互为补充。基于此,为避免单一指标的局限性,本研究整合BLEU、TER与BERTScore,从“表层一致性—编辑成本—语义保持”三个维度,对神经机器翻译工具及大语言模型的译文质量进行全面、系统的评估。

3 译本基础特征的描述性统计分析

3.1 基础描述统计

本研究借助MTMineR_2023文本分析工具,对人工(HT)、大语言模型(LLMs)及神经机器翻译(NMT)三组文本进行了基础描述统计,结果如表1所示。

为深入考察各文本在词汇、句长及字符使用层面的特征,根据表1的基础信息,进一步计算了词长、句长、汉字密度、假名密度等核心指标汇总于表2。

基于表2数据,各译本在主要指标上呈现以下规律:词长:HT>NMT>LLMs;句长:HT>LLMs>NMT;

汉字密度:LLMs>NMT>HT;平假名密度:HT>NMT>LLMs;外来语(片假名)密度:NMT>HT>LLMs。综合分析上述指标可知,LLMs译文在句长等部分指标上较NMT更接近人工译文。具体而言,人工译文词长与句长均为最长,同时具备最低的汉字密度与最高的平假名密度,这些特征共同指向人工译本和语成分较高、句式较舒展、具有相对明显的口语化倾向。LLMs译文的句长较NMT更接近人工译文,但其汉字密度显著偏高、平假名密度偏低,这一组合特征暗示其语言风格更偏向书面语。NMT译文句长最短,结构最为简洁,同时其外来语密度最高,体现出较强的技术文档风格。此外,在数字符号使用上,LLMs与NMT比人工更偏好使用阿拉伯数字,NMT的数字比例最高,进一步印证了其技术文档风格。

表1 各译本基础信息一览

译本	字节数	句子数	字符数	汉字数	平假名数	片假名数	罗马字数	数字
HT_JP	99002	680	45496	26624	16910	1931	31	2
LLMs01_Gemini	96614	648	44064	27197	14882	1960	25	86
LLMs02_GPT5	93153	691	42648	27545	13450	1636	17	123
LLMs03_GPT4o	91598	704	42106	25864	14598	1626	18	157
LLMs04_Qwen	95514	744	44200	26989	15408	1782	21	9
LLMs05_Wenxin	95390	658	43967	26491	15751	1714	11	194
NMT01_DeepL	93416	737	43051	27149	14325	1556	21	165
NMT02_Google	97739	842	44901	25985	16582	2280	54	230
NMT03_Baidu	90928	630	42016	25305	14609	2091	11	150
NMT04_Youdao	90389	693	41736	24357	15228	2130	21	167
NMT05_Niu	94277	628	43444	26353	15220	1862	9	127

资料来源:笔者根据文本分析工具MTMineR_2023的统计结果制作。

表2 各译本词长、句长、汉字假名使用率一览

译本	词长 (单位:字符)	句长 (单位:字符)	汉字占比 (单位:%)	平假名占比 (单位:%)	片假名占比 (单位:%)	数字占比 (单位:‰)
HT_JP	1.66	66.91	58.52	37.17	4.24	0.02
LLMs01_Gemini	1.61	68.00	61.72	33.77	4.45	0.89
LLMs02_GPT5	1.61	61.72	64.59	31.54	3.84	1.29
LLMs03_GPT4o	1.63	59.81	61.43	34.67	3.86	1.78
LLMs04_Qwen	1.63	59.41	61.06	34.86	4.03	0.14
LLMs05_Wenxin	1.61	66.82	60.25	35.82	3.90	2.07
LLMs均值	1.62	63.15	61.80	34.14	4.02	1.23
NMT01_DeepL	1.60	58.41	63.06	33.27	3.61	1.95
NMT02_Google	1.66	53.33	57.87	36.93	5.08	2.76
NMT03_Baidu	1.62	66.69	60.23	34.77	4.98	2.00
NMT04_Youdao	1.63	60.23	58.36	36.49	5.10	1.92
NMT05_Niu	1.63	69.18	60.66	35.03	4.29	1.68
NMT均值	1.63	61.57	60.03	35.31	4.61	2.07

资料来源:笔者根据文本分析工具MTMineR_2023的统计结果制作

3.2 形符类符比(TTR)与词汇多样性

形符(token)与类符(type)是计量语言学中分析文本词汇特征的基本单位。形符指文本中出现的每一

个词项,类符指去除重复后的不同词项数。形符类符比(Type-Token Ratio, TTR)是衡量文本词汇密度与多样性的常用指标。由于日语文本的字符级 TTR 通常低于 0.1,为便于跨文本比较,本研究采用标准形符类符比(STTR)进行分析。各译本的形符数、类符数、TTR 及 STTR 如表 3 所示。

表 3 各译本形符类符比一览

译本	形符数	类符数	TTR	STTR
HT_JP	30161	2727	0.0904	0.4521
LLMs01_Gemini	30168	2802	0.0929	0.4644
LLMs02_GPT5	29216	2800	0.0958	0.4792
LLMs03_GPT4o	28464	2672	0.0939	0.4694
LLMs04_Qwen	29489	2697	0.0915	0.4573
LLMs05_Wenxin	29929	2809	0.0939	0.4693
LLMs 均值	29453	2756	0.0936	0.4679
NMT01_DeepL	29320	2512	0.0857	0.4284
NMT02_Google	29879	2474	0.0828	0.4140
NMT03_Baidu	28352	2549	0.0899	0.4495
NMT04_Youdao	28306	2554	0.0902	0.4511
NMT05_Niu	29054	2527	0.0870	0.4349
NMT 均值	28982	2523	0.0871	0.4356

资料来源:笔者根据文本分析工具 MTMineR_2023 的统计结果制作。

根据表 3 数据,在文本长度(形符数)上,人工译文最长,LLMs 译文总体长于 NMT 且更接近人工译文,其中 Gemini 与人工译文长度最接近。在词汇多样性(STTR)方面,各译本整体数值分布较为接近,但组间差异仍可辨识。总体而言,LLMs 译文的 STTR 普遍高于 NMT 译文,甚至略高于人工译文,这表明 LLMs 在词汇生成层面表现出更强的多样性,能够调用更丰富的词汇组合与表达变体。这一结果反映了 LLMs 在产出过程中更倾向于避免简单重复这一基于大规模预训练的生成特征。然而,需要指出的是,在诸如官方文书等特定文本类型中,过高的词汇变异度可能不利于术语统一与文本风格的严谨性,从而影响核心概念的传播与文本的权威性。因此,在追求译文表达自然流畅的同时,亦需根据文本功能与规范要求对词汇多样性进行适度调控。

3.3 MVR 与词类分布特征

修饰语动词比(Modifier-Verb Ratio, MVR)是日语文体分析的重要指标(毛文伟,2022),其计算公式为 $MVR = (\text{形容词数量} + \text{形容动词数量} + \text{副词数量} + \text{连体词数量}) / \text{动词数量}$ 。该指标常需参照名词率等其他词类分布来解释(桦岛、寿岳,1965)。本文基于 MTMineR_2023 的统计结果,计算了各译本的 MVR 及主要词类分布密度,如表 4 所示。

既有研究表明,在日语文本中高名词密度文本常具有高度概括性,不利于读者对具体情境的联想;高修饰语比例则多与静态情状描写相关;而动词密度的增加能强化动作描写与叙事的动态性(桦岛,1954、1955;桦岛、寿岳,1965)。

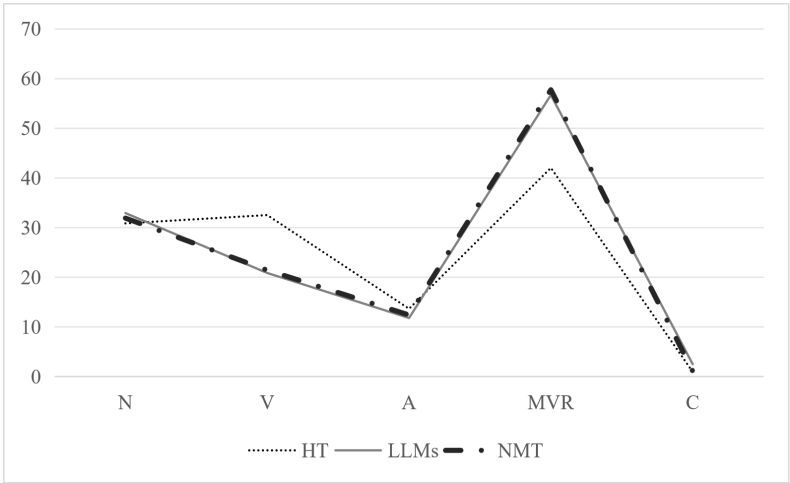
根据表 4 数据,人工译文动词密度显著最高,修饰语比例也较高,但名词密度最低。这一特征表明,人工译文采用动词驱动的叙事结构,从而呈现出更强的动态性与叙事推进感。相较而言,LLMs 译文的名词密度与接续词密度较高,反映出其行文侧重于信息凝练与逻辑衔接,整体风格趋近书面语、偏向静态说明。NMT

表4 各译本词类分布与MVR一览

译本	名词率(%)	动词率(‰)	修饰词率(‰)	MVR(%)	接续词句率(‰)
HT_JP	30.85	32.53	13.67	42.03	0.97
LLMs01_Gemini	33.11	19.38	12.05	62.18	1.97
LLMs02_GPT5	34.40	20.26	10.86	53.59	1.69
LLMs03_GPT4o	32.67	22.32	12.04	53.94	3.18
LLMs04_Qwen	32.20	21.29	11.33	53.24	5.27
LLMs05_Wenxin	32.27	21.11	12.78	60.56	0.50
LLMs均值	32.93	20.87	11.81	56.70	2.52
NMT01_DeepL	33.28	19.23	10.36	53.86	0.65
NMT02_Google	30.85	20.18	11.43	56.62	3.36
NMT03_Baidu	32.27	22.63	12.92	57.10	0.50
NMT04_Youdao	31.28	22.98	13.90	60.48	0.43
NMT05_Niu	31.87	22.19	13.47	60.68	0.64
NMT均值	31.91	21.44	12.41	57.75	1.12

资料来源：笔者根据文本分析工具MTMineR_2023的统计结果制作。

译文的MVR值最高，结合其较高的修饰词密度与较低的动词密度，显示其词类配置偏向于静态描述，动态呈现能力较弱。为直观地呈现组间差异，本文根据表4数据绘制了词性分布与MVR趋势图(图1)。



资料来源：笔者根据统计结果利用绘图软件制作

图1 人工、LLMs、NMT译本词性分布与MVR趋势图

图1显示,LLMs与NMT在各项指标上呈现出较高的相似性,而与人工译本存在显著分野。其中,高动词密度是人工译本的突出特征,也是区分人工与其他译文的核心指标。尽管人工译本的修饰语比例不低,但其极高的动词使用量导致了显著的低MVR。

综合来看,尽管LLMs与NMT在技术原理上存在本质差异,但在当前汉译日任务中,二者在词类选择与句式组织上均表现出高名词化、低动词化的共性倾向,这使得其译文在语篇结构与文体风格上较为趋同,并与以动词使用为特征的人工译文形成了鲜明对比。这一结果揭示了当前NMT与LLMs在生成具有动态感和叙事生命力文本方面存在的共性局限。

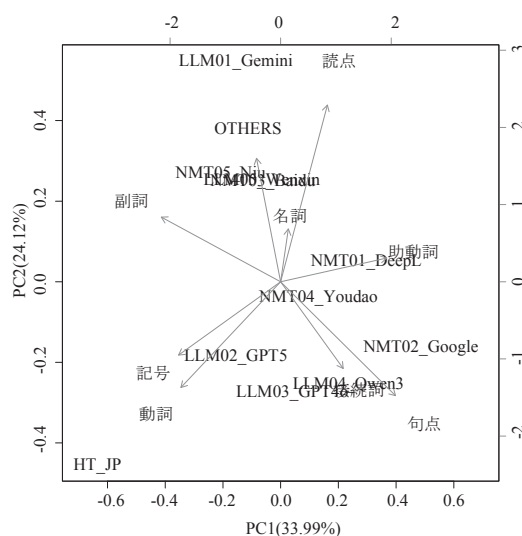
4 译本特征的多维统计分析

4.1 基于N-gram特征的主成分分析(PCA)与热力图

为从局部到整体分析各译文的词类构成特征及文本相似度,本研究采用主成分分析(Principal Component Analysis, PCA)与热力图(HeatMap)两种可视化方法。首先,基于译文词性标注序列的N-gram(N=1, 2, 3)相对频率矩阵,以协方差矩阵进行PCA分解,提取主要成分以降低维度并揭示系统性差异。其次,基于文本间相关系数矩阵,采用基于最短距离的single聚类法,并通过热力图(色阶由左至右表示数值由低至高)直观呈现文本相似梯度与聚类模式。

4.1.1 基于词类Unigram的分析

词类Unigram反映了各词类在文本中的分布频率,是识别文体特征的关键指标(桦岛、寿岳,1965;金,1994、2003;浅石,2017;井关等,2023)。基于词类Unigram数据生成的PCA双标图(图2)显示,人工译本在动词、符号及副词等类别的使用率相对较高,而在读点、名词及其他(OTHERS)上的使用率较低。图中,人工译文孤立分布于左下区域,与所有机器译文呈现显著离散。在LLMs译文中,Gemini最接近读点与OTHERS,表明其文本在断句习惯与非核心词类使用上表现出松散与冗余。GPT-4o与Qwen3则靠近句点与接续词,提示其倾向于生成由短句或高频分句构成的文本结构。



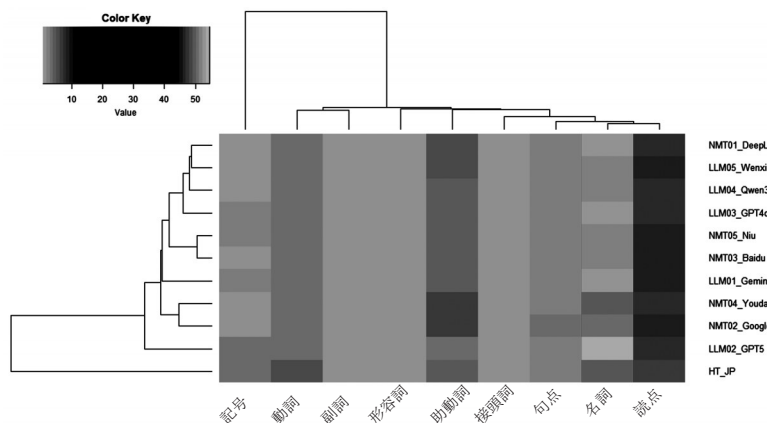
资料来源:笔者根据文本统计结果利用R语言制作

图2 基于Unigram的主成分分析图

进一步观察基于词类Unigram频率的热力图(图3)可见,人工译文在动词项上的颜色显著深于其它,而在名词与读点上的颜色则明显偏浅,这直观印证了前述人工译文动态性更强、名词化程度更低的特征。同时,热力图的聚类结构清晰显示,LLMs与NMT在各词类上的分布模式具有较高一致性,且彼此之间的距离明显小于其与人工译文的距离,进一步支持了LLMs与NMT译文在词类使用模式上趋同并与人类译文存在系统性差异的结论。

4.1.2 基于词类Bigram的分析

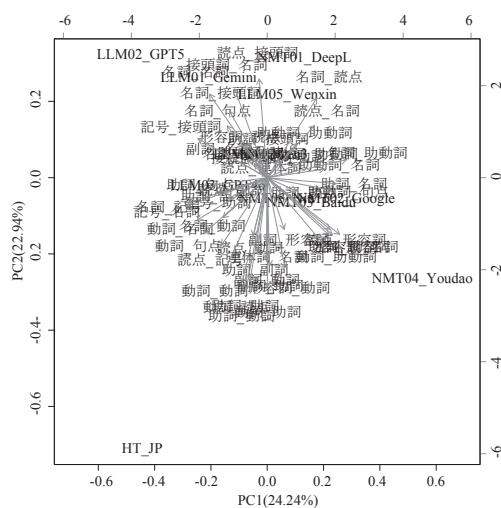
词类Bigram反映了文本中的词类搭配模式与句法结构倾向。本研究基于词类Bigram对文本进行了统



资料来源:笔者根据文本统计结果利用R语言制作

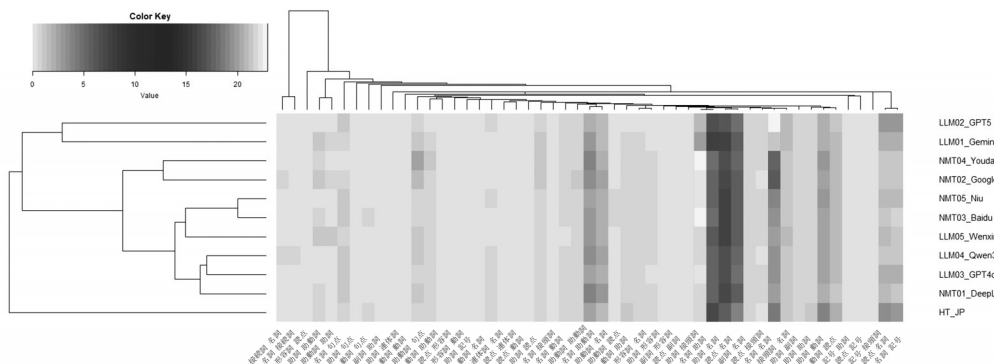
图3 基于Unigram的热力图

计,并取频率最高的50个项目进行分析,得到PCA图(图4)与热力图(图5)。



资料来源:笔者根据文本统计结果利用R语言制作。

图4 基于Bigram的主成分分析图



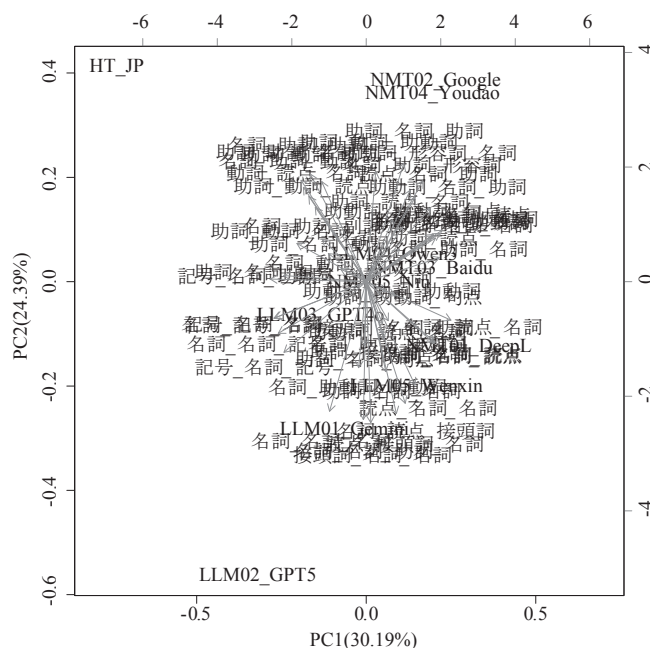
资料来源:笔者根据文本统计结果利用R语言制作。

图5 基于Bigram的热力图

图4显示,在词类搭配与句法结构层面,LLMs和NMT译文与人工译文仍存在根本差异。人工译文呈现出以动词为核心的句法组织模式,高频搭配为“动词-助词”“助词-动词”及“动词-动词”,这支撑了其叙事密度高、节奏紧凑的语篇特征。与之相对,LLMs译文则在“名词-名词”“名词-读点”“接头词-名词”及“名词-接头词”等模式上表现高频,显示出名词短语连续使用、读点分隔意群依赖及修饰结构繁复的显著倾向。热力图5亦印证了此模式。NMT译文的高频搭配集中于“名词-助词”与“助词-名词”,表明其倾向于采用结构清晰、成分完整(如较少省略主宾语)的简单句式。这一特征或反映出NMT系统为保证基础句法的正确性而采取的保守生成策略。

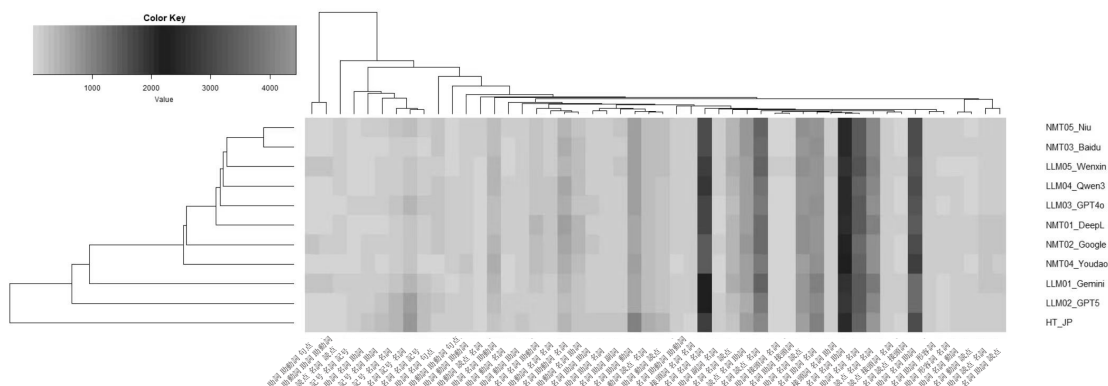
4.1.3 基于词类 Trigram 的分析

词类 Trigram 有助于在更复杂的结构层面揭示文本的句法特征与语篇组织方式。本研究基于词类 Trigram 统计,选取高频50个项目进行分析,得到PCA图(图6)与热力图(图7)。



资料来源:笔者根据文本统计结果利用R语言制作。

图6 基于Trigram的主成分分析图



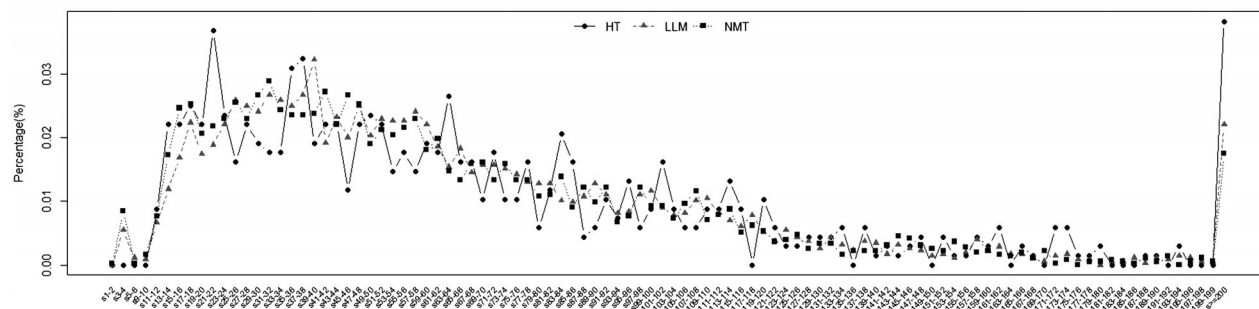
资料来源:笔者根据文本统计结果利用R语言制作。

图7 基于Trigram的热力图

在 Trigram 层面, NMT 译文集中分布于 PC1 与 PC2 的正值区域, 其典型模式为“助词-名词-助词”与“名词-助词-名词”, 呈现出以名词为中心、修饰成分密集衔接的“高修饰链”结构。LLMs 译文的分布相对分散, 但核心仍具名词性特征, 尤以 Gemini 与 GPT5 为著, 其高频模式包括“名词-名词-名词”“读点-名词-名词”及“名词-名词-助词”, 这体现了其句子结构的高度凝缩、复合术语密集及谓语成分时有省略的特点, 整体构成信息高度压缩的文本。人工译文则保持其独特性, 高频“Trigram 子名词-助词-动词”体现了完整的句子结构及侧重动作描述的叙事倾向, “助词-名词-助词”则可能与话题引入或信息聚焦相关。两者共同构成了人工译文句法完整、富于变化、兼具叙事与论述的语篇特质。

4.2 文本的句长变化

句长分布特征是区分人类文本与 AI 生成文本的关键指标之一。已有研究指出, 人类文本的句长通常呈现更大的变异度与离散性, 而 AI 生成文本的句长分布则往往更为集中, 可预测性更强 (Muñoz-Ortiz, 2024)。为考察不同译文在此维度上的差异, 本研究对各组译文的句长分布进行了统计与可视化分析 (图 8)。



资料来源:笔者根据文本统计结果利用 R 语言制作。

图 8 HT\LLMs\NMT 的句长变化图

如图 8 所示, 人工译文的句长波动范围显著大于 LLMs 与 NMT, 且分布更为分散; 相比之下, LLMs 与 NMT 译文的句长则呈现出较高的集中趋势, 波动范围较小。这一结果表明, 人工译文在句长上具有更高的灵活性与多样性、译文句式在复杂度与变化性上具有明显优势, 这体现出人类译者在句式调配与语篇组织中的灵活性与创造性; 而机器译文 (包括 LLMs) 则表现出更强的规律性与一致性, 反映出其基于概率模型的生成机制对常规、稳妥句式的偏好。本研究结果进一步支持了句长分布特征可作为有效区分人工与 AI 生成文本的依据。

5 各译本翻译质量评估

为系统评估 LLMs 与 NMT 在中国特色话语中译日任务中的表现, 本研究采用自动评价方法对译文质量进行量化分析。基于 Python 环境, 分别计算各 LLMs 与 NMT 译文相对于人工参考译文的 BLEU、TER、BERTScore 三项指标。上述指标分别从形式对齐、译后编辑成本及语义保持三个层面评估译文质量, 具体计算结果汇总于表 5。

表 5 的评估结果显示, 各机器译本的 BLEU 值与 BERTScore 普遍处于较高区间, 表明其在词汇层面的一致性及基本语义的传达上均取得了较好效果。然而, 与此形成鲜明对比的是, 各译本的翻译编辑率 (TER) 得

表5 各译本BLEU\TER\BERTScore数值一览

译本	BLEU	TER	BERTScore
LLMs01_Gemini	46.2488	1.0309	0.7622
LLMs02_GPT5	45.8199	1.0154	0.7303
LLMs03_GPT4o	47.5484	1.0000	0.7669
LLMs04_Qwen	45.7999	1.0000	0.7651
LLMs05_Wenxin	46.7659	1.0014	0.7460
NMT01_DeepL	44.5286	1.0252	0.7482
NMT02_Google	41.6066	1.0449	0.7942
NMT03_Baidu	48.9022	1.0014	0.7165
NMT04_Youdao	44.3420	1.0000	0.7118
NMT05_Niu	53.4407	1.0575	0.7309

资料来源:笔者根据Python编程语言计算获得数据并制表。

分普遍高于1.0,这揭示了其在句子结构、语篇组织及整体风格上与人工译本存在本质差异,导致译后编辑成本高昂。

这一矛盾现象在纽马克语义翻译与交际翻译理论框架下可得到统一阐释:较高的BLEU与BERTscore值反映了LLMs与NMT译文在实现语义翻译,即关注源文本的语言形式与语义内容方面达到了一定水平;而极高的TER值则暴露出其在交际翻译,即关注读者的接受效果与译语文化的适应性维度存在严重缺陷。具体而言,这种失衡体现在以下四个方面:

①源语结构过度移植导致的冗余与扩展。在处理形式规整、对仗性强的语句时,LLMs和NMT为追求与原文形式“表面对齐”,常机械复制源语格式,导致译文句式冗长、功能词冗余。例(1)和例(2)显示,相较于精炼规整的人工译文,机器译本普遍存在结构扩展与冗余的倾向,要将其修正为符合目标语习惯的简洁表达,需进行大量删除、合并等编辑操作。

例(1)

原文:敢于说前人没有说过的新话,敢于干前人没有干过的事情。

人工译文:誰も語ったことのない新たな言葉を言い、誰もしたことのないことをする。

译文1:前人が言わなかった新しき言葉を敢えて言い、前人が為さなかったことを敢えて為す。

译文2:前人が語らなかった新しい言葉を敢えて語り、前人が行わなかった新しい事業を敢えて行う。

译文3:先人が語らなかった新たな言葉を語る勇気をもち、先人が行わなかった新たな事業に挑戦する勇気をもつ。

译文4:先人が語らなかった新たな言葉を敢えて語り、先人が行わなかったことを敢えて行う。

译文5:先人が言ったことのない新しい話を敢えて言い、先人がやったことのないことを敢えて行う。

译文6:先人が言わなかった新しい言葉を敢えて語り、先人が行わなかったことを敢えて行う。

译文7:先人が語らなかった新しいことを敢えて語り、先人が行わなかったことを行う。

译文8:前人が言ったことのない新しい話を思い切って言い、前人がやったことのないことを思い切って行う。

译文9:前人の語ったことのない新しい言葉を述べなければなりません。今までやったことのない

ことを思い切ってやる。

译文10:先人が言ったことのない新しい話を敢えて言い、先人がやったことのないことを敢えてやる。

例(2)

原文:为党育人、为国育才

人工译文:党と国のための人材育成

译文1:党のために人材を育成し、国家のために人材を育成すること

译文2:党のために人を育て、国家のために才を養成すること

译文3:党のために人材を育て、国家のために人材を養成し

译文4:党のために人材を育て、国のために才能を養うことを堅持し

译文5:党のために人を育て、国家のために人材を育成すること

译文6:党のための人材育成、国家のための人材育成

译文7:党のために人材を育成し、国のために人材を育成すること

译文8:党のために人を育て、国のために人材を育てること

译文9:党のために人を育て、国のために人材を育成すること

译文10:党のために人を育成し、国のために人材を育成する

②同形汉字机械对应导致的表达生硬。面对富含四字格、成语等文化负载词的文本,LLMs与NMT倾向于采取“同形汉字直接替换”策略,即极端的“逐字翻译”。这虽在字面上实现了对等,却产生了大量不符合日语表达的生硬译文。例(3)、例(4)表明,优化此类译文需进行“交际翻译”导向的大规模风格化改写、词汇替换与结构调整等操作。

例(3)

原文:长期共存、互相监督、肝胆相照、荣辱与共

人工译文:「長期にわたって共存し、互いに監督し合い、肝胆相照らし、荣辱を共にする」という方針

译文1:長期的な共存、相互監督、肝胆相照、荣辱共有を堅持し

译文2:長期共存・相互監督・誠実な協力・荣辱を共にする原則

译文3:長期共存、相互監督、肝胆相照らし、荣辱を共にすること

译文4:長期的な共存、相互監督、肝胆相照らす精神、荣辱を共にする態度

译文5:長期的に共存し、相互に監督し、肝脾相照らし、荣辱を共にすること

译文6:長期共存、相互監督、肝胆相照、荣辱を共にし

译文7:長期共存、相互監督、相互信頼、名誉と不名誉の共有

译文8:長期的な共存、相互監督、肝胆相照、荣辱と共に堅持し

译文9:長期的共存、相互監督、肝胆相照らし、荣辱を共にし

译文10:長期共存、相互監督、肝胆相照らし、荣辱共にし

例(4)

原文:自信自强、守正创新,踔厉奋发、勇毅前行

人工译文:自信をもって自強をはかり、根本を貫いて革新を起こし、意気盛んに、勇気凛々と、
译文1:自信自強、守正創新、踔厲奮發、勇毅前行し、
译文2:自信自強・守正創新し、奮励努力・勇往邁進し、
译文3:自信と自強をもって、正道を堅持しつつ革新を図り、勇ましく奮い立ち、果敢に前進し、
译文4:自信自強・守正創新、踔厲奮發・勇毅前行し
译文5:自信と自立を持ち、正道を守り革新し、勇猛果敢に努力し、勇毅に前進し、
译文6:自信と自強の精神で正道を堅持しつつ革新を推進し、奮闘努力し勇気を持って前進し、
译文7:自信と自力更生を堅持し、誠実を堅持し革新を求め、勇気と決意をもって前進し、
译文8:自信を持ち、正の革新を守り、勇敢に前進し、
译文9:自信を、守が革新踔厲安き、勇毅進、
译文10:自信を持って自強し、正を守り革新し、奮發し、勇敢に前進し、

③在文化修辞层面,对隐喻等修辞手法理解肤浅,多停留于字面直译(如例(5)),无法有效传述修辞效果。

例(5)

原文:让权力在阳光下运行。

人工译文:監督の下で権力を正しく行使する。

译文1:権力が太陽の下で運用されるようにしなければならない。
译文2:権力を陽の下で運営させる。
译文3:権力の運用を太陽の下で行わせる。
译文4:権力が公開の下で運営されるようにする。
译文5:権力が陽の下で運営される。
译文6:権力が陽光の下で運営されるようにする。
译文7:権力の公開行使を確保する。
译文8:権力を日光の下で運行させる。
译文9:権力が太陽の下で運営されるようにする。
译文10:権力が太陽の下で運行されるようにする。

在例(5)中,多数LLMs和NMT将“阳光”直译为「太陽/陽光」,尽管有2例捕捉到了部分引申义「公開」,但仍无法完全准确传达原文中“监督”“透明化”的深层语义。

④在情感语用层面,几乎无法捕捉和再现原文的语气与情感色彩(如例(6)),译文整体缺乏应有的交际活力与情感共鸣。

例(6)

原文:坚定信心、同心同德,埋头苦干、奋勇前进,为全面建设社会主义现代化国家、全面推进中华民族伟大复兴而团结奋斗!

人工译文:自信を固め、心を一つにし、地道に努力し、勇往まい進し、社会主義現代化国家を全面的に建設するために、中華民族の偉大な復興を全面的に推進するために、団結奮闘しようではないか。

译文1:自信を固め、心を一つにして進み、黙々と努力し、奮勇前進し、社会主義現代化強国の全面的建設、中華民族の偉大な復興の全面的推進のために団結奮闘せねばならない!

译文2:信念を堅く持ち、心を一つにし、懸命に働き、勇敢に前進し、社会主義現代化国家の全面的建設と中華民族偉大なる復興の全面的推進のために団結して奮闘するのである。

译文3:確固たる信念を持ち、心を一つにして、懸命に働き、勇敢に前進し、社会主義現代化国家の全面的建設と中華民族の偉大な復興の全面的推進のために団結して奮闘しよう!

译文4:確固たる信念を持ち、心を一つにし、ひたむきに努力し、勇気を出して前進し、社会主義現代化国家の全面的建設および中華民族の偉大な復興のため、団結して奮闘しよう!

译文5:信心を固め、同心同徳で、ひたすら実務に励み、勇猛果敢に前進し、社会主義現代化国家を全面的に建設し、中華民族の偉大な復興を全面的に推進するために団結して奮闘しなければならぬ!

译文6:自信を固め、心を一つにし、埋もれて努力し、勇敢に進み、社会主義現代化国家の全面的な建設と中華民族の偉大な復興の全面的な推進のために団結して闘おう!

译文7:揺るぎない自信と団結の精神をもって、我々は奮闘努力を惜しまず、勇敢に前進し、団結して現代社会主義国家の全面的建設と中華民族の偉大な復興の推進に努めなければならない。

译文8:自信を固め、心を一つにし、全力を尽くして努力し、勇敢に前進し、社会主義現代化国家を全面的に建設し、中華民族の偉大な復興を全面的に推進するために団結して奮闘しなければならない。

译文9:信念を固め、一心同体となり、懸命に努力し、勇気を奮い立たせ、社会主義現代化国家の全面的建設と中華民族の偉大な復興の全面的推進のために団結奮闘しなければならない。

译文10:自信を固め、心を一つにし、努力に没頭し、勇敢に前進し、社会主義現代化国家を全面的に建設し、中華民族の偉大な復興を全面的に推進するために団結奮闘しなければならない!

在例(6)中,人工译文中的「勇往まい進」是日语中地道的口号表达,句末的「しようではないか」准确传递了极具热情的鼓舞与劝诱语气,二者共同再现了源文本中浓厚的热血色彩与感染力。然而,LLMs与NMT译文均未能还原这两个关键语言成分所表达的感召力与情感共鸣。本文统计了政治文本中频率较高的句末情态表达在各文本中的使用情况,包括表示“强烈断定”的「ほかならない」、表示“劝诱・号召”的「(よ)うではないか」、表示“意志・劝诱”的「よう・う」、表示“义务及强必要性”的「なければならない」、以及表示“中等必要性”的「必要がある／である」(详见表6)。

统计结果显示,人工译者在情态表达的运用上具有更高的多样性与均衡性。相比之下,LLMs与NMT在一定程度上能够再现情感较为客观、译词相对固定的表达,如表示“有必要”的「必要がある／である」。在「なければならない」的处理上,各类工具虽普遍能够生成,但频率存在明显失衡,其中Gemini与Wenxin、Google Translate表现出过度使用的倾向。然而,在处理具有较强主观色彩或承载细腻情感的「ほかならない」「(よ)うではないか」「よう・う」等表达时,机器翻译整体上仍存在明显不足,难以有效再现原文中蕴含的逻辑强调感与呼吁感等语气特征。

综上所述,LLMs与NMT译文的高TER值从根本上揭示了其在实现交际翻译目标上的瓶颈。未来的技术优化与译后编辑实践需着力引导模型超越对源语形式的机械忠实,加强对目标语社会文化语境、文体惯例及读者接受心理的考量,在语义翻译与交际翻译之间寻求更合理的动态平衡。

表6 各译本句末情态表达统计

译本	ほかならない	(よ)うではないか	よう・う	なければならない	必要がある / である
HT_JP	3	3	5	59	15
LLMs01_Gemini	0	0	0	166	15
LLMs02_GPT5	0	0	0	29	9
LLMs03_GPT4o	0	0	1	47	16
LLMs04_Qwen	0	0	1	43	11
LLMs05_Wenxin	0	0	0	222	1
NMT01_DeepL	0	0	1	24	11
NMT02_Google	0	0	2	144	5
NMT03_Baidu	0	0	1	64	4
NMT04_Youdao	0	0	0	8	2
NMT05_Niu	0	0	0	67	0

资料来源:笔者根据文本数据统计结果制作。

6 结 语

本研究通过计量分析与自动评估发现,当前神经机器翻译(NMT)与大语言模型(LLMs)在处理中国特色话语中译日任务时,虽能实现基本语义传递,但在译文质量的关键维度与人工翻译存在显著差距,尚难以独立胜任对文化准确性与交际效果有较高要求的翻译任务。

具体而言,在句法与文体层面,LLMs及NMT译文与人工译文最显著的差异体现在动词使用上,人工译文倾向更高频的动词驱动叙事,而LLMs与NMT译文普遍呈现高名词化、低动词驱动的特征,句法结构受源语束缚明显,缺乏对目标语表达习惯与叙事节奏的灵活顺应。在文化适应性层面,两类机器译文过度依赖同形汉字词,多采取机械对应策略,导致表达生硬、文化负载词转换不到位,难以实现地道的文化调试。在修辞与情感再现层面,两类机器译文对原文修辞手法、语气色彩及情感内涵的识别与转换能力明显不足,译文缺乏应有的交际活力与感染力。上述局限共同制约了两类机器译文在跨文化交际中的传播效力,使其难以充分实现“讲好中国故事”所需的深度跨文化沟通目标。

基于此,在翻译实践与教学中应作出相应调整。就翻译实践而言,译后编辑不应仅停留于表层纠错,而应从整体上提升译文的动态性与叙事感。具体而言,在句法层面,应对机器译文中常见的冗长名词句进行适当拆分,同时增加谓语动词的使用,以增强句子的推进力与可读性;在文化调试层面,应警惕同形汉字词的机械对应,优先选用符合日语表达习惯的地道说法,以实现文化负载项的在地化表达;在情感补偿层面,则需结合语境,对原文中的语气与情态意义进行适度重构,例如通过补充契合语境的终助词或调整句末表达形式,以弥补机器译文在主观色彩与语用功能方面的不足。就翻译教学而言,应引导学习者充分认识当前机器翻译在上述维度的能力边界,并有针对性地提升其进行文本矫治、文化调适与情感补偿等高级译后编辑能力,以更好地适应人机协作背景下的翻译实践新常态。

同时,本研究亦存在一定局限性。本文未考虑LLMs不同参数设置(如temperature值)对译文生成结果的影响,且研究主要依赖文本统计与对比分析,尚未引入日语母语者评价,因而在译文自然度与可读性等方面仍有待进一步验证。上述问题尚需后续研究通过多参数实验设计与感知评价方法加以拓展与深化。

注释

[1] 本文对于若干主流大语言模型开展实验。在初步测试中发现,DeepSeek、Kimi、豆包等通用大语言模型在涉及政治话语的翻译任务中显示话题敏感,未能生成有效译文,Qianwen、Wenxin4.5表现相对稳定。因此本文选取能够完成《党的二十大报告》汉译日任务的5个大语言模型(其中国外3个,国内2个)所生成的译文作为LLMs译文样本,并对译文生成时的模型版本信息进行了标注。

参考文献

- 于栋楠. 翻译规范理论视域下的政治文献日译本词汇修订效果评估研究[D]. 天津: 天津外国语大学, 2024.
- 习近平. 高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗: 在中国共产党第二十次全国代表大会上的报告[M]. 北京: 人民出版社, 2022.
- 习近平. 高举中国特色社会主义伟大旗帜为全面建设社会主义现代化国家而团结奋斗: 在中国共产党第二十次全国代表大会上的报告(日文版)[M]. 中央党史和文献研究院, 译. 北京: 外文出版社, 2023.
- 王羽迪. 生成式AI在日译汉翻译中的质量评估研究——以大语言模型GPT-4o为例[J]. 日语学习与研究, 2025(2).
- 王丽丽. 党政文献翻译的忠实原则——以党的二十大报告英译为例[J]. 中国翻译, 2023, 44(6).
- 王俭, 姜龙范. 中央文献翻译研究的文献计量及可视化分析[J]. 天津外国语大学学报, 2021, 28(5).
- 毛文伟, 朱海莹. 计量文体学视角下的汉日机器翻译语言特征研究——以《阿Q正传》日译本为例[J]. 日语学习与研究, 2024(6).
- 毛文伟. 日本的语料库文体学研究: 进展、问题及展望[J]. 外国语(上海外国语大学学报), 2021, 44(3).
- 毛文伟. 数据挖掘技术在文本特征分析中的应用研究——以夏目漱石中篇小说为例[J]. 外语电化教学, 2018(6).
- 毛文伟. 数据挖掘技术在学习者作文特征分析中的应用研究[J]. 日语学习与研究, 2022(2).
- 文旭, 田亚灵. ChatGPT应用于中国特色话语翻译的有效性研究[J]. 上海翻译, 2024(2).
- 邓文舛. 概念隐喻视角下中国特色话语的日译研究[D]. 南宁: 广西大学, 2025.
- 付晶, 王跃洪. 技术文献翻译研究——纽马克交际翻译理论的应用[J]. 中国科技翻译, 2014, 27(4).
- 达哇才让. 党政文献民族语文翻译难点及策略研究[J]. 民族翻译, 2023(2).
- 刘健. 中国共产党文献日译本中的汉日同形词——以《习近平谈治国理政》日译本中的“四字格”词为例[J]. 日语学习与研究, 2019(6).
- 许钧, 穆雷主编. 翻译学概论[M]. 南京: 译林出版社, 2021.
- 李运博, 卿学民. 时政文献的日文翻译方法与策略——以中国共产党二十大报告的日译本为例[J]. 日语学习与研究, 2025(1).
- 李运博, 卿学民. 叙事学视域下的政治文献翻译研究——以习近平总书记“七一”重要讲话日译本为例[J]. 日语学习与研究, 2021(6).
- 李玥. 从功能对等理论看十九大报告中熟语的日译[J]. 日语学习与研究, 2019(1).
- 李钰婧. 中央文献外译的跨文化交际策略——以十九大报告中“实践”一词的日译为例[J]. 日语学习与研究, 2019(6).
- 张威, 白一博. 中国特色对外话语在英语世界的译介与传播——背景、主题与价值[J]. 中国外语, 2022, 19(6).
- 张桂丽, 黄忠, 李念念, 等. 对外话语建设背景下的中国特色词四字格的日译考察——基于《政府工作报告》语料库[J]. 北京联合大学学报(人文社会科学版), 2023, 21(5).
- 陈岩, 孟海霞. 论中国共产党第十九次全国代表大会报告中重要表述的日译[J]. 日语学习与研究, 2018(1).
- 彼得·纽马克. 翻译教程[M]. 上海: 上海外语教育出版社, 2001.
- 荆瑶. 《共同构建人类命运共同体》日译本隐喻比较研究[J]. 大学, 2021, (37).
- 修刚, 米原千秋. 中日政治文献“同形词”的翻译——以2015年《政府工作报告》日译为例[J]. 天津外国语大学学报, 2016, 23(4).
- 修刚, 李运博, 花超. 十九大报告翻译实践与中央文献的日文翻译策略[J]. 日语学习与研究, 2018(2).
- 高冲. 基于核心期刊的中央文献翻译研究学术图景分析——基于CiteSpace的文献计量可视化分析(1993-2022)[J]. 中州大学学报, 2024, 41(1).
- 黄友义, 李晶. 做好中央文献翻译打通国际传播的最后一公里[J]. 天津外国语大学学报, 2022, 29(2).
- 董海涛. 党政文献日译中的同形词翻译探析[J]. 民族翻译, 2022(5).
- 韩子满, 柴通达. 人工智能知识翻译能力探析——以文学掌故为例[J]. 外语电化教学, 2024, (5).
- 童富智, 修刚. 基于语料库的中日对应词语义韵对比研究——以中央文献日译「坚持(する)」为例[J]. 日语学习与研究, 2020(2).
- 赖斯(Reiss, K). 翻译批评: 潜力与制约[M]. 罗德斯(Rhodes, E.F.)译. 上海: 上海外语教育出版社, 2004.
- 管新潮, 陆晓蕾. 基于Python的语料库翻译: 数据分析与理论探索[M]. 上海: 上海交通大学出版社, 2022.
- 薛悦, 修刚. 融通中外: 时政文献外译受众认知语境的强化与调节[J]. 上海翻译, 2022(6).
- 薛悦. 弗米尔目的论视角下的复句翻译研究——以《习近平谈治国理政》等著作的日译本对比为例[J]. 日语学习与研究, 2023(2).
- 井関龍太, 菊池理沙, 望月正哉, 等. 品詞構成に基づく文体指標は読者の印象とどのように関わるか—MVRと品詞構成率の心理学的検

- 討一[J]. 計量国語学, 2022, 33(7).
- 中尾桂子. 品詞構成率に基づくテキスト分析の可能性—メール自己紹介文, 小説, 作文, 名大コーパスの比較から—[J]. 大妻女子大学紀要: 文系, 2010(42).
- 金明哲, 鄭穹穹. テキストコーパスマイニングツールMTMineR[J]. 計量国語学, 2020, 32(5).
- 金明哲. 文節パターンに基づいた文章の書き手の識別[J]. 行動計量学, 2013, 40(1).
- 金明哲. 助詞のn-gramモデルに基づいた書き手の識別[J]. 計量国語学, 2003, 23(8).
- 金明哲. 読点の打ち方と文章の分類[J]. 計量国語学, 1994, 19(7).
- 浅石卓真. テキストの特徴を計量する指標の概観[J]. 日本図書館情報学会誌, 2017, 63(3).
- 高橋寛治, 竹野峻輔, 山本和英. 機械翻訳システムの出力安定性の評価[J]. 人工知能学会論文誌, 2017, 32(5).
- 樺島忠雄, 寿岳章子. 文体の科学[M]. 京都: 綜芸舎, 1965.
- 樺島忠雄. 現代文における品詞の比率とその増減の要因について[J]. 国語学, 1954, 18.
- 樺島忠雄. 類別した品詞の比率に見られる規則性[J]. 国語国文, 1955, 250.
- 劉雪琴, 金明哲. 宇野浩二の病氣前後の文体変化に関する計量的分析[J]. 計量国語学, 2017, 31(2).
- Muñoz-Ortiz A, Gómez-Rodríguez C, Vilares D. Contrasting Linguistic Patterns in Human and LLM-Generated News Text[J]. Artif Intell Rev, 2024; 57(10).
- Newmark, Peter. Approaches to Translation[M]. Oxford: Pergamon Press, 1981.
- Newmark, Peter. A Textbook of Translation[M]. London: Prentice Hall, 1988.

Linguistic Features and Quality Assessment of Chinese-Specific Discourse Translation via Large Language Models: A Case Study of the Japanese Version of the 20th CPC National Congress Report

BAI Yang (*Shanghai University of International Business and Economics*)

JIN Mingzhe (*Jilin International Studies University*)

Abstract: Drawing upon the Japanese version of the Report to the 20th National Congress of the Communist Party of China as a reference, this study examines translations generated by five mainstream Neural Machine Translation (NMT) systems and five Large Language Models (LLMs). Utilizing text-mining tools and programming languages, a quantitative analysis, visualization, and automated quality evaluation were conducted. The findings reveal significant stylistic discrepancies between machine-generated translations and human translations (HT). Specifically, NMT and LLM outputs exhibit lower verb frequencies and a pronounced tendency toward nominalization, alongside a mechanical correspondence in Kanji vocabulary and syntactic structures. Automated evaluations based on BLEU, TER, and BERTScore further indicate that while machine translations achieve a functional level of basic semantic transfer, they frequently suffer from rigid literalism and inflexible expression when processing culture-loaded words, four-character idioms, and antithetical structures. Furthermore, these systems underperform in rhetorical recognition and the conveyance of affective nuances, leading to clear limitations in communicative effect. Consequently, for the Japanese translation of discourse with Chinese characteristics, the quality of current NMT and LLM outputs has yet to meet the rigorous standards required for professional translation.

Keywords: The 20th CPC National Congress Report; Japanese version; Large Language Models; Text mining; Translation Quality Assessment

(责任编辑:王晨萌)