

# 基于机器学习的政经语篇人机翻译风格研究\*

——以《国富论》中译本为例

孔德璐

(同济大学外国语学院, 上海 200092 / 比利时根特大学艺术与哲学学院, 根特 9000)

**摘要:** 本文利用机器学习算法, 基于自建《国富论》人机多译本平行语料库, 通过分类、聚类、特征选择实验, 考察人类专家译本与大语言模型机器译本在组间和组内的翻译风格差异。结果表明, 人机译本组间分类效果最好, N 元语法特征突显出机器译本受原文影响较大、直译较多, 而人类译者能够发挥主体性; 人工译本组内分类效果较好, 助词、逗号比例、平均词长等特征表明郭大力译本行文简洁明了, 语法显化及欧化程度相对较低; 机器译本组内分类效果最差, 机器译本之间的翻译风格差异并不明显。

**关键词:** 国富论; 机器学习; 政经语篇; 人机译本; 翻译风格

**中图分类号:** H315.9 **文献标识码:** A **文章编号:** 2097-4272 (2025) 01-0120-11

## 0. 引言

政经语篇是一种特殊的文本体裁, 重点关注政治和经济关系在话语上的呈现模式, 强调经济、政治和学术知识的解释性和语境性 (Maesse 2015: 285)。《国民财富的性质和原因的研究》(*The Wealth of Nations*, 简称《国富论》) 出版于 1776 年, 是古典经济学之父亚当·斯密 (Adam Smith) 的政经语篇代表作。如今, 人工智能技术发展迅猛, 以 ChatGPT 为代表的大语言模型 (Large Language Model, LLM) 在翻译任务上也初露锋芒, 斩获不俗的成绩 (李梅、孔德璐 2024; 于蕾 2024)。尽管机器翻译在速度和成本方面具有显著优势, 但在准确性、语义理解和文化传递等方面和人工翻译相比仍存在差距 (梁君英、刘益光 2023; 沈梦菲、黄伟 2024)。人机翻译在涉及高度专业化和文化特异性的政经语篇中呈现何种差异仍有待讨论。本文基于自建语料库, 结合机器学习方法, 对政经语篇人机译本的翻译风格进行量化对比研究, 为进一步提升翻译质量提供实证和理论参考。

## 1. 研究基础

翻译风格研究主要涉及翻译语言模式识别和特征量化阐释过程, 是基于语料库的翻译研究中的重要课题之一 (胡开宝、田绪军 2023)。Baker (2000: 260) 提出译文中会不可避免

收稿日期: 2021-09-06; 作者修订: 2024-06-03; 本刊修订: 2024-12-07

**基金项目:** 国家社会科学基金项目“神经网络机器翻译的译后编辑量化系统模型研究” (19BYY128); 国家留学基金委员会 2024 年国家建设高水平大学公派研究生项目 (202406260211)

**作者信息:** 孔德璐, 博士研究生, 研究方向为语料库翻译学、机器翻译译后编辑、翻译技术, E-mail: kongdelu2009@hotmail.com。

地展现“译者的指纹 (Thumbprint)”, 率先利用语料库方法加以观察。国内相关研究也逐步推进, 研究层次多样、成果丰富 (黄立波、石欣玉 2018; 乐明、余潇慧 2022; 王梦瑶 2024)。然而, 语料库翻译风格研究范式仍待完善。研究者往往会预先选定一些特征, 以此来映射译作风格, 却忽略可能存在的更突显风格的隐性特征, 进而削弱研究可信度。例如, 类/形符比 (TTR) 并不一定能区分翻译风格 (Mikhailov & Villikka 2001)。同时, 预先设定计量语言特征的做法也被批评为“总是基于学者的自觉……显然这种方法是‘人工设计’的” (Ilisei & Inkpen 2011)。

通过机器学习方法, 构建特征集并将所研究的译本进行多维向量化, 再利用算法提取高贡献度语言特征, 如此便能部分弥补语料库翻译风格研究中预设计量语言特征的局限。该方法源于计量风格学, 最初采用机器学习方法进行作者身份判定, 逐步沿用至翻译研究。Lynch & Vogel (2018) 运用支持向量机模型来识别译者风格。国内学者如詹菊红、蒋跃 (2017) 使用支持向量机模型区分《傲慢与偏见》 (*Pride and Prejudice*, 1813) 的两个中文译本, 并对部分显著特征进行阐释。

目前翻译风格研究多见于文学翻译领域, 较少关注非文学体裁。文学译者为了追求“文学性”, 尚能发挥“译者主体”作用, 对原作进行主观再创作, 而政经语篇具有的严肃性、哲理性和逻辑性则向翻译风格提出了挑战。《国富论》译本众多, 但目前学术界在翻译领域的研究多为译作评析 (刘瑾玉 2015)、译家访谈 (冯婷婷、孙智帅 2018) 或汉译史研究 (刘瑾玉、王克非 2020), 针对《国富论》再译本的讨论略显不足, 多译本翻译风格的计量研究相对较少, 涉及人机译本对比的研究更是鲜见。

基于此, 本研究结合机器学习算法, 通过自建《国富论》政经语篇平行语料库, 对 4 个专家译本和 3 个 LLM 机器译本进行组间和组内的分类聚类实验。在特征解释部分, 结合语料统计数据, 对显著语言特征进行定性探究。本研究拟回答以下问题: 1) 机器学习方法能否区分政经语篇的译本风格? 2) 哪些特征在区分译本风格上发挥显著性作用? 3) 如何针对这些显著特征做进一步阐释和归纳?

## 2. 研究设计

### 2.1 研究语料

本研究选择《国富论》2015 年原文重印本、4 个人工专家译本、3 个 LLM 机器译本, 共计 7 个中译本作为研究语料。在人工译本方面, 参考豆瓣读书网<sup>①</sup>, 以接受度较广、好评度较高的全译本为选择标准, 分别筛选 2011 年中国华侨出版社发行的孙善春、李春长合译本 (简称孙译本, 豆瓣评分 9.3, 评价人数 196)、2010 年中央编译出版社谢宗林、李华夏合译本 (简称谢译本, 豆瓣评分 9.0, 评价人数 694)、2011 年陕西人民出版社杨敬年译本 (简称杨译本, 豆瓣评分 8.7, 评价人数 234)、2009 年上海三联书店郭大力、王亚南合译本 (简称郭译本, 豆瓣评分 8.6, 评价人数 552), 具有一定可比性。

在机器译本方面, 本研究遵循 SOTA 原则, 选择实验时在机器翻译任务上性能保持领先的 LLM 引擎。使用上述《国富论》英文原本作为机译原文, 选取 ChatGPT-4 (简称 ChatGPT), Claude-3-Opus (简称 Claude), Gemini-1.5-Pro (简称 Gemini) 作为引擎, 三者分别对应 OpenAI 公司、Anthropic 公司和谷歌公司推出的 LLM 工具, 进行机器翻译阶段

(2024 年 4 月至 5 月) 均在 LLM 性能天梯榜<sup>②</sup>上位列前茅, 属于时下热门、性能优异的 LLM 引擎。机器翻译流程均使用大语言模型 API 结合 Python 进行, 指令撰写遵循 CRISPE 框架<sup>③</sup>, 最终形成结构化的用户指令, 并在翻译过程中保持统一。

随后进行除噪并保留正文作为研究语料, 使用 Language Technology Platform (LTP)<sup>④</sup> 自然语言处理工具对语料进行分词、词性标注、依存句法标注。样本切分标准设置为 UTF-8 编码下 60 kB 固定文件大小, 即每个中文样本中包含 20 000 字左右, 有助于样本容量的基本统一。最终建成的可比概况如表 1 所示, 实验所用样本共计 210 个。

表 1. 语料详情

类别	作品	样本数	字数	形符数 (词)	类符数 (词)
人工 专家 译本	《国富论》源本	—	—	377595	9483
	郭译本	30	504211	317403	15092
	谢译本	36	621939	374804	15069
	杨译本	30	519996	320906	10637
	孙译本	28	494189	303271	12296
LLM 机器 译本	ChatGPT	29	506239	306076	9698
	Claude	27	473200	284724	12574
	Gemini	30	518546	313914	9802

## 2.2 特征集建立

表 2. 实验用特征集<sup>⑤</sup>

特征类型	特征名称	特征数量
词汇特征	类/形符比、标准类/形符比、平均片段类/形符比、平均词长、平均词频、停用词比例、辛普森指数、词汇多样性度量 (MLTD)、单现词比例、长词比例、高频词比例、词汇重复率、名词比例、动词比例、形容词比例、副词比例、助词比例、助词“过”、助词“的”、助词“得”、助词“地”、助词“着”、助词“了”、介词比例、代词比例、数词比例、连词比例、语气词比例、描述词比例、实词比例、虚词比例……	63
句法特征	平均句长、“把”字句比例、“被”字句比例、陈述句比例、问句比例、感叹句比例、逗号比例、分号比例、顿号比例、括号比例、引号比例、平均句树长 (词)、平均句树长 (字)、平均依存距离 (MDD)、平均句树高度、叶节点比例、平均子节点数、主谓关系比例、动宾关系比例、间宾关系比例、前置宾语比例、兼语比例……	30
可读性特征	词汇丰富度、句法丰富度、名词语义精确度、动词语义精确度、名动语义精确度、实词语义精确度、语义丰富度、语义清晰度、语义噪音	9
N 元语法特征	N=1: (1) 工人 (2) 认为 (3) 这个……(100) 价格 N=2: (1) 贸易_差额 (2) 耕种_土地 (3) 名义_价格……(100) 劳动_工资 N=3: (1) 西班牙_葡萄牙_殖民地 (2) 增加_土地_劳动 (3) 出口_其他_国家……(100) 欧洲_大部分地区 N=4: (1) 收入_财富_唯一_来源 (2) 各种_生活_必需品_便利品 (3) 欧洲_大学_古代_哲学……(100) 劳动_工资_资本_利润	400

特征集是机器学习实验的关键, 一个合适且全面的特征集能够尽可能广泛地将文本样本向量化。遵循 Volansky *et al.* (2015) 提出的风格特征选取原则, 同时参考以往研究 (黄伟、刘海涛 2009; 詹菊红、蒋跃 2017; Lynch & Vogel 2018; 孔德璐 2024), 本文建立的特征集分为

4 个大类：1) 词汇特征，基于词性标注集，考察词汇层面的标准化特征以及具体词性特征的分布情况；2) 句法特征，基于依存句法标注集，考察句法层面的标准化特征以及具体依存句法关系的分布情况；3) 可读性特征，参考雷蕾等（2024）提出的汉语文本可读性分析框架，考察不同译本之间可读性的差异；4) N 元语法<sup>⑥</sup>特征，基于提取出的 N 元词组（N 取 1 至 4），参考现有研究（Zhang *et al.* 2011），通过 TF-IDF 值筛选出高频出现的前 100 个 N 元词组，分析译本中词组的共现情况。实验用特征共 502 个，如表 2 所示。

### 2.3 分类器和参数

借助 Python 自编程序，批量提取出表 2 中列举的所有特征。经过统计，最终将 210 个文本样本转换为对应的基于 502 个特征的数学表达模型，通过机器学习平台 Weka 3.8.4<sup>⑦</sup>验证分类器效果，找到能够区分人机译本组间、组内的显著特征。

本研究使用朴素贝叶斯、支持向量机（SVM）、随机森林作为分类器。SVM 在 1963 年由著名的苏联数学家 Vladimir Vapnik 等人设计，并在广泛应用中展示出良好的性能（Žižka *et al.* 2019: 211）。随机森林结合多个决策树模型来提高分类性能和鲁棒性。多分类器的使用避免了单独使用某一算法导致的对照缺失和结果偶然，增加实验的科学性和客观性。参数调整方面采用网格搜索法，对关键指标遍历择优，其他分类器均进行类似操作。

## 3. 研究结果

### 3.1 分类实验结果

分类模型精度评估采用十折交叉验证法，该方法将总数据集随机切分 10 份，每次运行时都使用其中一份作为测试集，剩下 9 份作为训练集，并重复验证 10 次，最后的精度验证结果是 10 次重复验证结果的平均值。最终的实验结果如表 3 所示。

表 3. 分类实验结果

实验对象	分类器	实验结果		
		准确率	平均召回率	AUC 值
组间分类 (人工 vs. 机器)	SVM	97.62%	0.976	0.974
	朴素贝叶斯	94.28%	0.943	0.962
	随机森林	91.43%	0.912	0.999
组内分类 (人工译本组)	SVM	91.13%	0.911	0.941
	朴素贝叶斯	58.87%	0.589	0.799
	随机森林	95.97%	0.960	0.996
组内分类 (机器译本组)	SVM	81.40%	0.814	0.859
	朴素贝叶斯	40.69%	0.407	0.550
	随机森林	82.55%	0.826	0.895

注：实验对象指实验中用于分类和分析的具体译本或译本组，如组间分类将 4 个人工译本和 3 个机器译本分别视为人工译本和机器译本两个大类；组内分类则单独将人工或机器译本组内的各译本视为不同的类别。AUC（Area Under Curve）为 ROC 曲线下与坐标轴围成的面积。

实验结果包含 3 个指标：准确率表示模型正确预测样本数占总样本数的比例；平均召回率表示样本中的正例有多少被预测正确；AUC 值用来反映模型预测能力。其中，平均召回率和 AUC 值越接近 1，模型分类结果和预测能力越好。从表 3 中可以得到以下发现：1)



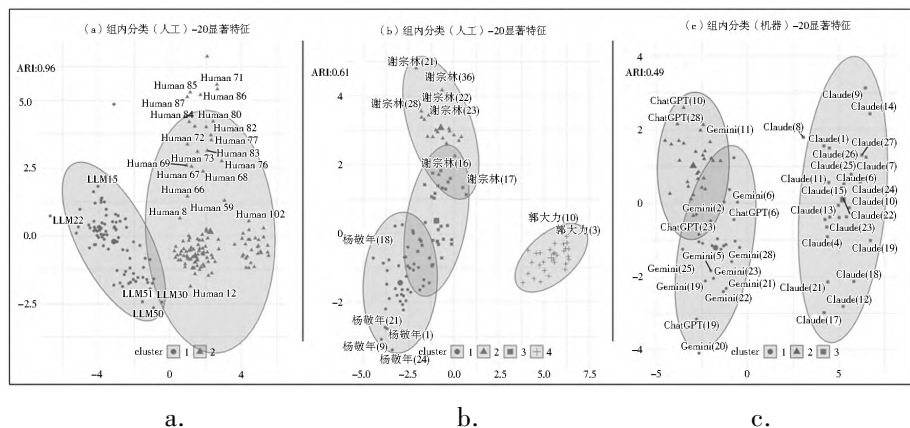
准确率方面,组间分类的 SVM 分类器最高,达到 97.62%,平均召回率和 AUC 值均超过 0.97,证明其分类结果和预测能力较好;组内分类(机器译本组)的朴素贝叶斯分类器准确率最低,仅为 40.69%,表明模型分类能力接近随机猜测。2) 实验对象方面,组间分类实验效果最好,3 种分类器平均准确率为 94.44%;其次是组内分类(人工译本组),平均准确率为 81.99%;效果最差的是组内分类(机器译本组),平均准确率仅为 68.21%。说明人机译本之间语言特征较为明显,特征差异较大,而机器译文内部差异较小,语言使用较为相似。

### 3.2 聚类实验结果

聚类分析是一种无监督的机器学习方法,将数据对象划分成多个类或簇,使同一类或簇中的对象具有较高的相似度(刘颖 2014: 123)。本部分运用 K-means 聚类分析进一步检验分类结果的信度和效度。根据实验对象不同,聚类实验同样进行 3 次。在进行每次实验之前,本研究利用特征选择分类器,使用卡方评估法,通过计算各类别中样本的卡方值来评估特征贡献度。从结果中挑选出 20 个卡方值最高的显著特征进行聚类实验。这种降维的方法可以在确保准确率的同时减少实验复杂度,节省实验时间,同时可以提取出显著特征以便进行特征阐释。调整兰德系数(Adjusted Rand Index, ARI)是一种衡量聚类效果的外部指标,取值范围为  $[-1, 1]$ ,值越大意味着聚类结果与真实情况越吻合。

经过观察,得到以下发现:1) 聚类效果最好的是组间分类(图 1a),ARI 为 0.96,最接近 1,且人工译本和机器译本置信椭圆的重合度很小,说明二者在 20 个显著特征上的语言风格差异较大。2) 组内聚类实验中,人工译本组(图 1b) ARI 为 0.61,相比组间分类稍低,但整体效果仍较好。不同译者之间的语言风格差异相对明显,内部区分度较高,特别是郭译本可谓独树一帜。而效果最差的是机器译本组(图 1c),ARI 仅为 0.49,且 Gemini 和 ChatGPT 译本重合度高,样本之间的混淆度高。

图 1. 聚类实验结果



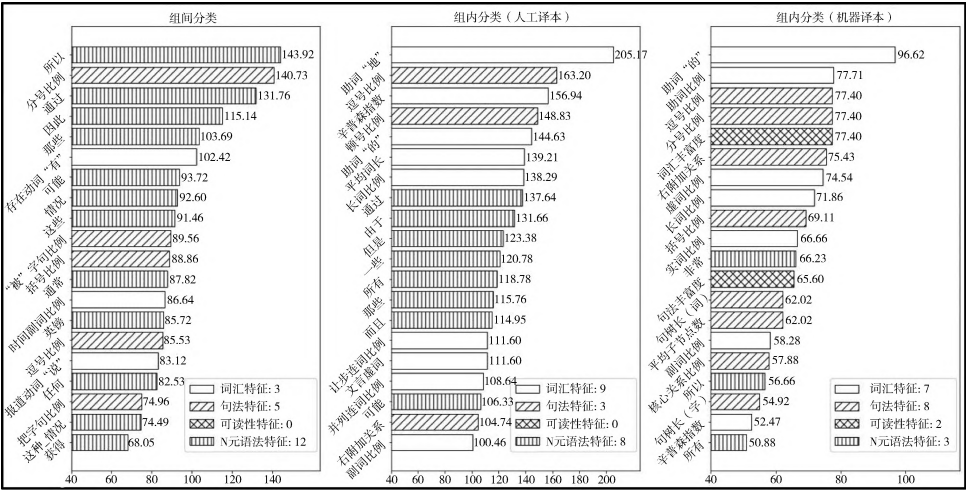
聚类分析验证了分类实验结果,说明每个聚类实验中的 20 个显著特征能够识别不同译本或译本组的翻译风格。

### 3.3 特征贡献度排名

本小节展示语言特征在不同实验中的表现。结合下页图 2 能够发现:1) 在组间分类

中，N 元语法特征整体占比较大、贡献度较高。2) 在人工译本的组内分类中，前 20 个显著特征的平均贡献度相较其他两组最大，末尾特征贡献度（100.46）甚至大于机器译本组内分类的首个特征贡献度（96.62），且词汇特征和句法特征占比较大。3) 在机器译本的组内分类中，前 20 个显著特征的平均贡献度最低，反映出该组译文区分度较小，且同样是词汇和句法特征占据较大比例。

图 2. 特征贡献度排名



总体来看，N 元语法特征在组间分类中发挥较大作用，词汇特征、句法特征在组内分类中平均贡献度较高。而可读性特征类型仅在机器译本的组内分类中出现两次，整体分类能力有限。

4. 分析与讨论

4.1 组间分类：人机译本

本小节通过统计分析人机译本组间分类的显著语言特征，探究人类专家译文和 LLM 机器译文的风格差异。囿于篇幅，表 4 仅展示排名靠前的 6 个特征数据统计。

表 4. 组间分类显著特征数据统计

	所以	通过	因此	那些	分号比例	“被”字句
机器译本	2.00	26.36	38.09	15.40	421.81	43.43
人工译本	17.67	4.87	17.02	12.58	299.02	24.03
p 值	<0.00	<0.00	<0.00	<0.00	<0.00	<0.00

注：表格中数据经过标准化，均为每万词出现频次。p 值为卡方检验，下同。

首先，“所以”和“因此”承接原因和结果的逻辑关系，是政经语篇中常见的词语。机器译本对“所以”的使用十分匮乏，却大量使用了“因此”来衔接因果，且均有显著差异。究其原因，第一，这两个词遵循互补原则。原文中表达因果语义转折次数有限，而政经语篇内在逻辑性和严谨性要求忠实传达原意，所以“因此”一词在机器译文中频繁出现就会直接导致“所以”的使用频次下降。第二，LLM 的翻译策略更偏直译。原文中频繁出现 therefore（万词率为 23.09，多译为“因此”），相反 thus 和 so 使用较少（万词率为

4.29, 多译为“所以”), 和机器译本趋势十分相似。这说明在人工译本中, 这两个词的分布较为平均, 人类译者更能发挥主体性来平衡“因此”和“所以”的使用, 也更善于建构因果和意图的衔接关系(梁君英、刘益光 2023)。相比之下, 机器译文中中规中矩, 将 *therefore* 翻译为“因此”的频次更高, 也即 LLM 应用于机器翻译仍存在以往机翻系统出现的“机械性”现象(胡开宝、李翼 2016)。

其次, “所以”之外的语言特征在机器译本中均被“过量使用”, 如表 4 中的“通过”直译了英文的 *by*, “那些”(以及图 2 中出现的“这些”)直译了原文中频繁使用的指示代词 *that* 和 *this*; 两个句法特征同样如此, 分号特征差异在以往研究中也有出现, 本质上是因为机器无法做到人类译者那样, 能够根据英语中分号的作用转换为汉语语境下的分号、逗号、顿号, 或直接另起分句(詹菊红、蒋跃 2017; 孔德璐 2024), 机译“被”字句比例高是受到原文多被动表述的影响, 人类译者能够灵活变通, “被被不停”的情况也就较少出现(蒋跃, 等 2016; 沈梦菲、黄伟 2024)。由此看来, LLM 机器译本在语言使用上无法彻底避开“源语影响”和“语法显化”这两座大山(贺阳 2008; 胡显耀 2010)。

最后, 机器翻译的成本和速度是其优势, 但和人工译本相比, 图 2 中占多数的 N 元语法特征已成为一种“机器翻译腔”(Daems *et al.* 2017), 这在影响读者阅读体验的同时, 也会降低译本接受度。随着 LLM 工具在翻译实践中的普及, 译者和学者应当更加关注机器译文风格特征, 培养针对 LLM 译文的译后编辑能力, 尽量减少“机器翻译腔”现象。

#### 4.2 组内分类: 人工译本

接下来观察区分人工译本的显著特征, 如表 5 所示。

表 5. 人工译本显著特征数据统计

	助词“地”	逗号比例	辛普森指数	顿号比例	助词“的”	平均词长
郭译本	0.58	7142.50	0.36	5.775	705.51	1.50
谢译本	9.51	6319.55	0.40	256.642	876.95	1.60
杨译本	29.00	6031.04	0.44	394.385	980.63	1.57
孙译本	21.25	5922.16	0.41	430.057	885.30	1.57

首先, 助词“地”比例是区分人工译本贡献度最高的特征, 其中郭译本使用最少, 通篇出现频次不超过 20 例, 谢译本次之(共 340 余例), 孙译本较多(共 650 余例), 杨译本最多(共 940 余例); 助词“的”比例表现出类似趋势, 郭译本使用最少, 杨译本最多。在现代汉语书面语中, 结构助词“地”用在状语后面, 连接状语和中心语, “的”使用频次最高, 连接定语和中心语(刘雪梅 2006)。检索语料后发现, 谢、杨、孙 3 个译本在助词“地”左侧最高频共现的状语分别是“不断”(22 次)、“部分”(83 次)、“或多或少”(38 次), 郭译本均以零助词形式搭配这些状语, 同时使用虚词来替代助词“的”。如郭译本第 5 篇文章的标题“论君主或国家之收入”, 而谢、杨、孙译本均译为“论君主或国家的收入”, 另在谈及义务时, 郭译本多用“义务之实行”。翻译文本中“的”的高频出现是翻译汉语欧化的标志之一(谢耀基 2001: 18), 郭译本中助词“的”的出现频次相对较低, 语言特征偏向原创汉语, 即欧化程度较低。

其次,郭译本在逗号、顿号使用方面有着显著的差异,其逗号使用比例最高(42 602例),顿号使用比例最低(35例);相反孙译本逗号使用比例相对最低(22 705例),顿号使用比例最高(1651例)。例(1)截取自郭译本中的实际语料,译者直接借鉴英文逗号所发挥的连接并列作用,将诸多职业用逗号连接起来(其他3个译本均使用顿号)。

(1) 郭译本:为了制造这种剪刀也就须有许多种工人把他们各色各样的手艺结合起来,例如矿工,熔炉的建造者,木材的采伐者,烧炭工人,制砖工人,泥水匠,熔炉的工人,锻工,铁匠,以及其他等等。

究其原因,除了谢、杨、孙3位译者不拘泥于原文,会根据原文的具体含义和词汇间的关系调整标点符号的使用之外,郭译本多用逗号具有一定的时代印记。五四运动之后,随着《新式标点符号议案》的颁布,顿号逐渐从逗号中独立出来,并形成固定的使用方法,成为汉语独有的标点符号(韩秋红2011)。而郭译本创作于20世纪20至30年代,顿号的使用尚未普及,人们仍多用逗号来表示并列词组之间的停顿。

最后,辛普森指数用来测量语言多样性(Gazzola *et al.* 2020),数值越小,表示多样性越高。结合表1和表5,可以发现郭译本辛普森指数最低,约为0.36,说明该译本词汇丰富度更高,词语运用更加多样和广泛;杨译本最高,约为0.44,也即随机抽取的两个词项很大程度归为同一类别,词汇丰富度较低。在平均词长上,郭译本明显更短,原因是该译本一字词的使用比例较大。在去除停用词后,郭译本一字词在前100个高频词中类符占36个,形符占44%。而谢、杨、孙译本一字词类符均只占24个左右,且谢译本形符比最低,仅为33%。总地说来,平均词长差别能够反映出郭译本行文简洁明了,善于应用一字词来准确表达原文思想。

#### 4.3 组内分类:机器译本

最后观察区分机器译本的显著特征,如表6所示。

表6. 机器译本显著特征数据统计

	助词“的”	助词比例	逗号比例	分号比例	词汇丰富度	右附加关系
ChatGPT	939.93	1112.94	5854.53	595.36	5.90	11.04%
Claude	750.29	932.45	2.77	0.26	6.35	9.09%
Gemini	877.98	1035.52	5870.29	633.43	5.93	10.23%

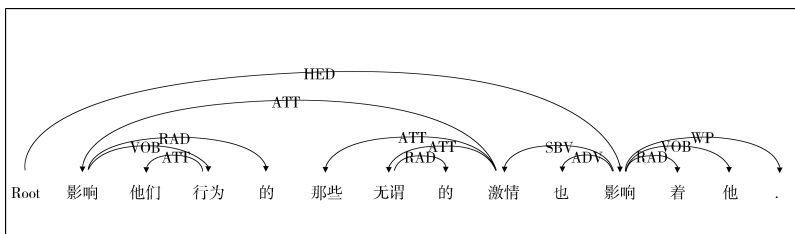
首先,助词“的”的使用比例及其上层特征助词比例是区分3个机器译本的显著特征。Claude译本在助词使用上明显低于其他两个译本,且参考胡显耀(2010)对汉语语体特征统计,在Claude译本中,助词“的”(万词率750.29)的使用频次甚至低于人工译者非文学翻译文体(万词率785.8)。加之助词分布频次高是欧化标志之一(谢耀基2001),由此可说,相较于其他两个译本,Claude译本在助词使用上的欧化程度较低。

其次,Claude译本中逗号和分号的使用频次较低。检索发现,Claude译本中绝大多数使用的是英文半角符号,而其他两个译本使用了中文全角符号,后者属于本研究考察的中文翻译风格符号特征范畴。中文译文中不应出现英文半角标点,ChatGPT和Gemini译本不需要调整标点,节省译后编辑时间,略优于Claude译本。



最后,根据雷蕾等(2024),词汇丰富度计算的是文本中所有单词的熵值,值越大,说明文本所使用的单词越不确定,文本词汇越富于变化。Claude 译本的词汇丰富度稍高,而 ChatGPT 和 Gemini 之间差距不大。右附加关系(Right Adjunct, RAD)是一种依存语法结构,指句子中修饰语出现在被修饰成分的右侧,而助词是重要的判断标准。如图3所示, RAD 结构多指向附加在“影响”右边的助词“的”和“着”,所以助词分布同样影响着依存语法中 RAD 结构的比例。Claude 译本中助词分布相对较少,其 RAD 结构的数量也明显下降。

图3. 依存图结构(Gemini 译本例句)



总结来看,机器译本之间翻译风格差异较小。这或许是出于 LLM 训练算法和数据上的近似,又或是 LLM 面对一些语言外部知识等“人文内核”的捉襟见肘(冯志伟、张灯柯 2024)。

## 5. 结语

本文借助机器学习算法,考察经典政经语篇《国富论》的人机中文译本翻译风格,在人工译本和机器译本组间和组内进行分阶段实验,并从 502 个特征中实现贡献度排名,结合语篇对显著特征进行统计、分析和阐释。

研究得出以下结论:1)表3和图1的结果表明,机器学习算法能够有效区分各译本,其中人机译本组间分类效果最好(平均准确率 94.44%),其次是人工译本组内分类(81.99%),而机器译本组内分类效果相对最差(68.21%)。2)通过卡方评估对特征贡献度进行排序,揭示在聚类过程中发挥重要作用的特征。图2展示出人机译本在词汇、标点、N 元语法层面的高贡献度特征呈现出了不同的分布情况。3)人机译本组间对比中,“所以”“因此”等 N 元语法特征突显出机器译本受原文影响较大、直译较多,而人类译者更加懂得变通;人工译本组内对比中,助词“地”、逗号比例等特征突显出郭译本和现代汉语行文方式有所区别,体现出一定的时代特点,同时,郭译本行文简洁明了,在欧化程度方面低于其他 3 个译本;机器译本组内对比中,各译本的翻译风格并不十分鲜明,仅能凭借助词、逗号与分号、词汇丰富度等特征将 Claude 译本突显出来,ChatGPT 译本与 Gemini 译本的风格差异不大。

本文尚存在以下局限:1)选取人机译本数量有限,未来可将学界讨论较多的其他译本纳入考察范围;2)LLM 更新迭代速度较快,因此译文复现性不强,可考虑建立模型以实时考察 LLM 机器译本同人工译本之间的差异;3)特征之间有交叉的空间,如助词比例和 RAD 结构之间存在一定相关性,这种特征间的交互作用尚待进一步挖掘。

\*衷心感谢匿名审稿专家提出的宝贵修改意见。

## 注释:

- ① 豆瓣读书网站: <https://book.douban.com/>。(访问时间: 2024年6月1日, 下同。)
- ② LLM 性能天梯榜网址为 <https://arena.lmsys.org/>。
- ③ CRISPE 指令框架网址为 <https://github.com/mattnigh/ChatGPT3-Free-Prompt-List>。
- ④ LTP 提供一系列中文自然语言处理工具, 可以对中文文本进行分词、词性标注、句法分析等。该工具的 Base2 深度学习模型在分词、词性标注、依存句法的性能上分别达到了 99.18、98.69、90.19, 相较于其他自然语言处理工具, 属于时下 SOTA 水平。详情参考 <https://github.com/HIT-SCIR/ltp/>。
- ⑤ 1) 词汇特征比例=词类频次/总词数, 标点符号和空白字符不算入总词数。2) 句法特征比例=句型频次/总句数, 以句号、问号、叹号、分号、冒号作为判断句子结尾的标准。3) 符号比例=符号频次/总符号数。4) 依存特征比例=关系特征频次/总依存关系频次。5) N 元语法特征中的每个词项均为该项与样本总词数的比例。完整表格及补充资料详见 [https://github.com/DanielKong1996/WoN\\_MLTransStyle/](https://github.com/DanielKong1996/WoN_MLTransStyle/)。
- ⑥ N 元语法 (N-gram, 另译 N 元模式或 N 元组) 事实上关注的仍是词汇层面的词频共现情况, 每个词项都可以作为考察译文风格的一个向量。然而 N 元语法包含较多特征, 若纳入词汇特征层面恐怕会使其略显臃肿, 因此将其列为单独层级。
- ⑦ Weka (Waikato Environment for Knowledge Analysis) 是新西兰怀卡托大学开发的一款免费开源的机器学习数据挖掘软件, 详情参考 <https://www.cs.waikato.ac.nz/ml/weka/>。

## 参考文献:

- [1] 蔡强, 徐恩. 基于语料库的郭大力翻译风格研究 [J]. 江西理工大学学报, 2020, (2): 79-84.
- [2] 冯婷婷, 孙智帅. 经济学翻译家杨敬年学术翻译摭谈 [J]. 中国翻译, 2018, (2): 90-94.
- [3] 冯志伟, 张灯柯. 人工智能中的大语言模型 [J]. 外国语文, 2024, (3): 1-29.
- [4] 韩秋红. “五四时期”国语(文)类教科书实践新式标点符号的讨论 [J]. 辽宁大学学报(哲学社会科学版), 2011, (2): 151-153.
- [5] 贺阳. 现代汉语欧化语法现象研究 [J]. 世界汉语教学, 2008, (4): 16-31.
- [6] 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, (5): 10-14.
- [7] 胡开宝, 田绪军. 基于多语语料库的翻译研究: 议题与意义 [J]. 北京第二外国语学院学报, 2023, (2): 3-17.
- [8] 胡显耀. 基于语料库的汉语翻译语体特征多维分析 [J]. 外语教学与研究, 2010, (6): 451-458.
- [9] 黄立波, 石欣玉. 《到灯塔去》两个汉译本基于语料库的翻译风格比较 [J]. 解放军外国语学院学报, 2018, (2): 11-19.
- [10] 黄伟, 刘海涛. 汉语语体的计量特征在文本聚类中的应用 [J]. 计算机工程与应用, 2009, (29): 25-27.
- [11] 蒋跃, 张英贤, 韩红建. 人机翻译英语被动句相似度的向量距离测算 [J]. 西安交通大学学报(社会科学版), 2016, (3): 110-116.
- [12] 孔德璐. 基于机器学习方法的《德伯家的苔丝》中文译本翻译风格考察 [J]. 数字人文研究, 2024, (1): 58-73.
- [13] 乐明, 余潇慧. 语言计量指标在翻译风格研究中的应用: 以意识流小说《到灯塔去》为例 [J]. 解放军外国语学院学报, 2022, (3): 136-143.
- [14] 雷蕾, 韦瑶瑜, 刘康龙. Alpha Readability Chinese: 汉语文本可读性工具开发与应用 [J]. 外语与

- 外语教学, 2024, (1): 83-93.
- [15] 李梅, 孔德璐. ChatGPT 人机对话式翻译研究 [J/OL]. (2024-05-21) [2024-06-01]. <https://doi.org/10.13256/j.cnki.jusst.sse.230918456>.
- [16] 梁君英, 刘益光. 人类智能的翻译能力优势——基于语料库的人机翻译对比研究 [J]. 外语与外语教学, 2023, (3): 74-84.
- [17] 刘瑾玉. 严复手批《国富论》英文底本研究 [J]. 中国翻译, 2015, (5): 33-39.
- [18] 刘瑾玉, 王克非. 岂一个“富”字了得? ——《国富论》百年汉译史述论 [J]. 上海翻译, 2020, (2): 62-67.
- [19] 刘雪梅. 结构助词“的”“地”“得”用法分析 [J]. 现代语文, 2006, (12): 98-99.
- [20] 刘颖. 统计语言学 [M]. 北京: 清华大学出版社, 2014.
- [21] 沈梦菲, 黄伟. 寻找机器翻译痕迹——神经机器翻译文本的句法特征研究 [J]. 外语教学与研究, 2024, (3): 429-441.
- [22] 王梦瑶. 基于历时语料库的英诗汉译词汇特征研究 [J]. 外语导刊, 2024, (2): 49-57.
- [23] 谢耀基. 汉语语法欧化综述 [J]. 语文研究, 2001, (1): 17-22.
- [24] 于蕾. ChatGPT 翻译的词汇多样性和句法复杂度研究 [J]. 外语教学与研究, 2024, (2): 297-307.
- [25] 詹菊红, 蒋跃. 机器学习算法在翻译风格研究中的应用 [J]. 外语教学, 2017, (5): 80-85.
- [26] Baker, M. Towards a methodology for investigating the style of a literary translator [J]. *Target: International Journal of Translation Studies*, 2000, 12 (2): 241-266.
- [27] Daems, J., O. De Clercq & L. Macken. Translationese and post-edited: How comparable is comparable quality? [J]. *Linguistica Antverpiensia, New Series-Themes in Translation Studies*, 2017, 16 (1): 89-103.
- [28] Gazzola, M., T. Templin & L. J. McEntee-Atalianis. Measuring diversity in multilingual communication [J]. *Social Indicators Research*, 2020, 147 (2): 545-566.
- [29] Ilisei, I. & D. Inkpen. Translationese traits in Romanian newspapers: A machine learning approach [J]. *International Journal of Computational Linguistics and Applications*, 2011, 2 (2): 319-332.
- [30] Lynch, G. & C. Vogel. The translator's visibility: Detecting translatorial fingerprints in contemporaneous parallel translations [J]. *Computer Speech & Language*, 2018, 52 (6): 79-104.
- [31] Maesse, J. Economic experts: A discursive political economy of economics [J]. *Journal of Multicultural Discourses*, 2015, 10 (3): 279-305.
- [32] Mikhailov, M. & M. Villikka. Is there such a thing as a translator's style? [C] // P. Rayson & A. Wilson. *Proceedings of the Corpus Linguistics 2001 Conference*. Lancaster: UCREL, 2001: 378-385.
- [33] Volansky, V., N. Ordan & S. Wintner. On the features of translationese [J]. *Digital Scholarship in the Humanities*, 2015, 30 (1): 98-118.
- [34] Zhang, W., T. Yoshida & X. Tang. A comparative study of TF-IDF, LSI and multi-words for text classification [J]. *Expert Systems with Applications*, 2011, 38 (3): 2758-2765.
- [35] Žižka, J., F. Dařena & A. Svoboda. *Text Mining with Machine Learning: Principles and Techniques* [M]. Boca Raton: CRC Press, 2019.

(责任编辑 刘 心)

metacognitive awareness, i.e., person knowledge, problem solving, and L2 listening achievement, with their mediating effects accounting for 50% and 40% of the total effect respectively. This study further reveals the influencing mechanism of metacognitive awareness and self-efficacy on L2 listening comprehension and provides useful implications for L2 listening teaching and research practices.

**The Acquisition of Word Order as Information Structure Mechanism in Russian by Chinese Learners**, by FEI Jun-hui, p. 091

Using a felicity judgement task, the present study investigates the sensitivity of 92 Russian-learning undergraduates and graduates to discourse-related word order variation in Russian. The results show that L1 Chinese learners are generally insensitive to the information structure in L2 Russian. They always prefer the SVO order, no matter whether the subject encodes new information or not. Their knowledge of information structure is not related to language proficiency. The results indicate that information structure is difficult to be fully acquired because it involves knowledge from multiple modules like syntax, phonology and discourse.

**A Study on the Construction and Translation of International News Discourse from the Perspective of Narrative Competition**, by SI Xian-zhu, p. 101

News reports for international audience on events such as national territorial and border disputes are the narrative discourse construction, which involves narrative events, narrative perspective, and narrative structure, namely, what to tell, from which perspective, and how to tell. These three elements reflect the narrative competition of the previous news discourse from different dimensions. Starting from the above elements, this article examines the narrative characteristics and differences of international English news discourses between China and India media. Based on this, it explores how to improve the narrative construction and translation of China's news discourse for international audience, "tell the story of China well", and enable the world to better understand and accept China.

**A Corpus-Based Translation Criticism of the Nonce Words in Two Chinese Versions of *Ulysses***, by DING Li, LIU Ze-quan, p. 111

Nonce words (NWs) are writers' creative use of language and manifestation of the artistry of literary works. They not only provide readers with aesthetic enjoyment, but also bring great challenges to translation. Following the research paradigm of corpus-based literary translation criticism, the present study aims to compare and evaluate the translation of NWs in the two Chinese versions of *Ulysses*. Six types of NWs—created through blending, compounding, imitation, clipping, intentional misspelling, and derivation—are identified in the source text, each exerting a distinct poetic function depending on the context. The two Chinese versions mainly adopt four translation methods: overall reconstruction, partial reconstruction, normalization and omitting. A considerable number of NWs are normalized or omitted in both of the versions, while Jin Di's translation manages to reconstruct and keep more NWs than Xiao Qian & Wen Jie-ruo's. The result can be accounted for by the translators' contrasting views on poetics and stylistics.

**A ML-Based Investigation into the Translation Style of HT and MT on Political Economic Discourse: A Case Study on Chinese Translations of *The Wealth of Nations***, by KONG De-lu, p. 120

This research applies machine learning (ML) algorithms, namely classification, clustering, and attribute



selection experiments, to explore stylistic differences in human translations (HT) and large language model machine translations (MT) based on a parallel corpus of Chinese translations of *The Wealth of Nations*. The results show that the HT-MT inter-group comparison performs the best, where N-gram features reveal that MT translate the original in a literal and direct manner, while HT exhibit more adaptability. Comparison within HT group performs well, with features like auxiliary ratio, comma ratio, and average word length highlighting unique characteristics of Guo Dali's translation—clear, concise, and less explicit and Europeanized than the others in linguistic terms. Comparison within MT group performs relatively the worst, manifesting the least stylistic differences.

**Modern Auditory Subject in an Auditory Community: Re-listening to *Mrs. Dalloway***, by LONG Dan, ZENG Qi, p. 131

With prospering sound studies in China and abroad, literary critics have been making efforts to build discourses of auditory narratology and carry out textual study from this perspective. Proposing the two concepts “auditory community” and “auditory subject”, this study re-listens to Virginia Woolf's *Mrs. Dalloway* by focusing on the soundscape and auscultation in the novel, exploring the cultural connotation of auditory narrative and its critical power in this novel. Sounds made by the Big Ben, the car and the plane stimulate shared auditory experiences, historical memories and affect responses among the Londoners, which constitutes a multidimensional and heterogeneous auditory community, with which the auditory subjects in the novel have critical identification; they draw guidance from the community, obey its order, and at the same time, they keep vigilance against its discipline, and protect their free spirit and independent critical thinking.

**On the Animal Ethic Thoughts in Thomas Hardy's Poetry**, by LONG Yu, ZHANG Yan, p. 140

Thomas Hardy's poetry displays prominent ethical concerns over the relationship between humans and animals. Hardy's poems reveal his view of nature that all things on earth are interconnected, and that he is committed to breaking the boundaries between humans and animals. The animals in his poems, with their independent subjectivity and intrinsic value, have the equal status with human beings. Hardy's advocacy of loving-kindness to animals combines the spirit of altruism and the thought of respecting animal rights. His animal ethic thoughts bear the traces of the scientific and the philosophical development of the 19<sup>th</sup> century, and also echo the relevant postmodernist ideas of Derrida and Haraway. Hardy's animal ethic thoughts might be found to be ahead of his time. And his poetry might be perceived as enlightening for the contemporary animal studies.

**Spiritual Home or Cultural Trap: The Cultural Reflection on Black Music in Colson Whitehead's *The Underground Railroad***, by DENG Qin, p. 149

Colson Whitehead's *The Underground Railroad*, the winning novel of the National Book Award for Fiction, presents a cultural reflection on black music while retelling the history of slavery in the United States. By means of historical recontextualization, this paper reveals, on the one hand, the cultural significance of black music as African American people's spiritual home and, on the other hand, its hidden cultural traps as demonstrated by its escapism from reality and manipulation of racism. It is believed that the cultural reflection on black music in this novel suggests a reflection on African American cultural tradition, i. e., how to inherit and renew the national cultural tradition, in the new century.