

ChatGPT 翻译的词汇多样性和句法 复杂度研究*

北京语言大学 于 蕾

提要: ChatGPT 作为一个多用途语言模型,在翻译任务上的应用和表现引起了研究者的兴趣。本研究利用 TAALED 工具和 Python 编程提取并对比 ChatGPT 翻译、人工翻译和 DeepL 翻译的词汇多样性和句法复杂度的指标。结果显示:1) ChatGPT 翻译在词汇密度和内容词的使用方面具有优势,显示出其在词汇选择上的广度和创新性;2)在句法复杂度方面,ChatGPT 翻译的表现与人工翻译和 DeepL 翻译相当,但在并列短语比例和动词短语复杂度上超越了人工翻译,在平均 T 单位长度上超过了 DeepL 翻译,凸显了其在构建复杂句型方面的能力。这些发现揭示了 ChatGPT 翻译在词汇多样性和句法复杂度方面的潜力,对于如何有效利用 ChatGPT 来提高翻译质量具有参考意义。

关键词: ChatGPT 翻译、机器翻译、词汇多样性、句法复杂度

[中图分类号] H085 [文献标识码] A [文章编号] 1000-0429 (2024) 02-0297-11

DOI:10.19923/j.cnki.fltr.2024.02.005

1. 引言

随着人工智能技术的飞速发展,机器翻译已经成为自然语言处理领域的一个重要分支 (Jiang & Lu 2020),极大地促进了跨语言交流的便捷性。然而,在机器翻译技术取得显著进步的同时,其在词汇多样性和句法复杂度方面的表现仍是评价翻译质量的重要挑战。传统的机器翻译评估方法主要侧重于准确性和流畅性 (Ayala & Chen 2017),缺乏对词汇多样性和句法复杂度的评估。在此背景下,ChatGPT 在翻译任务上的应用和表现受到了学界的广泛关注。该模型基于大量数据的预训练,并针对特定翻译任务进行了精细调整 (Ray 2023),可以更有效地理解和处理不同语言间的细微差异。虽然 ChatGPT 并非专门为翻译而

* 衷心感谢《外语教学与研究》编辑部和匿名审稿专家的宝贵修改意见。

设计,但通过精确的指令(prompt)引导,其翻译潜能可以得到充分发挥。最新一代的 ChatGPT-4 在理解逻辑结构、处理语言复杂性、提升词汇丰富度和增强文本连贯性方面已取得显著进步(Herbold *et al.* 2023),这可能意味着它不仅能够提供更准确和流畅的翻译,还能更好地适应各种翻译场景的需求。鉴于此,本研究采用定量分析方法,不仅深入探究 ChatGPT 翻译、人工翻译和 DeepL 翻译在词汇多样性和句法复杂度方面的优势和局限,而且着重考察这几种翻译模式如何在不同应用场景中相互补充,以实现更高质量的翻译输出。深入了解和比较这些翻译模式能够帮助我们更好地运用它们以提高翻译质量,对于提升实际翻译工作的效率具有实践意义。

2. 文献综述

ChatGPT 已在多个领域展现出卓越能力,它能够应对多种对话场景和各类内容创作,如撰写电影评论、创作诗歌,甚至批改作文等(Helberger & Diakopoulos 2023)。其强大性能得益于背后的大规模自然语言模型(朱光辉、王喜文 2023)。在翻译领域,ChatGPT 同样表现出巨大的发展潜力,尤其是在译后编辑教学(王律、王湘玲 2023),以及翻译指令设计与优化(Lü *et al.* 2023)等方面,引起了学界的关注。尽管 ChatGPT 翻译取得了一定进步,但在处理需要深层理解、专业知识和文化敏感性的翻译任务时仍存在局限性。Sanz-Valdivieso & López-Arroyo (2023)的研究指出,尽管 ChatGPT 在减少翻译错误方面表现优于谷歌翻译,但在专业翻译任务中,这些自动化工具仍无法完全替代人类专家的角色。

词汇多样性指文本中使用不同单词的数量,是语言研究中的一个重要指标(Johansson 2008)。它作为评价翻译质量的核心因素之一(Kim 2020),不仅体现了译者的词汇知识深度和语言运用能力,而且对跨文化沟通具有重要意义。为准确测量词汇多样性,研究者已经开发了多种量化指标,如类符/形符比(type-token ratio, TTR)、文本词汇多样性测量(measure of textual lexical diversity, MTLT)等(McCarthy 2005; Koizumi 2012)。尽管如此,研究显示机器翻译文本在词汇多样性方面常常不如人工翻译。例如,Fu & Nederhof (2021)在使用如 BERT 等先进机器学习技术区分机器和人工翻译时发现,机器翻译在词汇多样性上相对逊色。同样,Brglez & Vintar (2022)在分析不同翻译系统时指出,尽管神经网络翻译系统在特定场合能表现出较高的词汇多样性,但在处理创新词汇时依然面临挑战。不过,随着技术的发展,这种差异正在逐渐缩小。例如,ChatGPT 在模拟丰富、多样化的词汇使用方面取得了显著进步,预示着机器翻译在词汇多样性方面的潜力将不断提升。

句法复杂度在翻译质量评估中扮演着关键角色,可直接影响文本的可理解性和表达的丰富性。句法复杂度衡量语言产出中句法结构的多样性和复杂性(Ortega 2003),涉及诸如句子长度、句法结构的复杂性等多个维度(Ortega 2003; Liu & Afzaal 2021; 朱周晔、王金铨 2023)。一般而言,句子越长、使用的句法结构越复杂,其句法复杂度就越高(梁君英、刘益光 2023)。在翻译领域,句法复杂度受多种因素影响,其中包括源语和目标语的固有词汇语法特征、文本体裁和作者的词法语法偏好(Al-Jabr 2006; Xu & Li 2021)。例如,汉语和英语在句法模式上存在显著差异,汉语以话题为主,词序较为自由,而英语具有更严格的句法规则和固定词序(Ma 2021)。Liu & Afzaal (2021)从句法复杂度的角度探讨了翻译简化假设。他们发现,翻译文本在句法上通常比非翻译文本更简单,尤其在从属结构和整体句子复杂性方面。句法复杂度在翻译过程中至关重要,它直接影响目标语言的流畅性和可读性。过于简化的句法可能导致信息丢失或意义改变,而过于复杂的句法则可能影响文本的清晰性。人工翻译者能灵活调整句法结构以适应上下文和目标语言习惯,而机器翻译系统可能因依赖算法和数据集而导致句法过度简化或复杂化。尽管 ChatGPT 翻译在处理句法复杂度方面展现出一定潜力,但与人工翻译及其他翻译系统相比,其表现仍需要进一步评估。

综上,本研究拟回答以下两个主要问题:

- 1) ChatGPT 翻译的词汇多样性是否超越人工翻译及 DeepL 翻译?
- 2) ChatGPT 翻译的句法复杂度是否超越人工翻译及 DeepL 翻译?

3. 研究设计

本研究旨在综合评估并比较 ChatGPT 翻译、人工翻译和 DeepL 翻译在词汇多样性和句法复杂度方面的表现。为此,首先构建了三个语料库: ChatGPT 英译语料库、人工英译语料库和 DeepL 英译语料库,以确保评估覆盖不同翻译来源。继而使用 TAALED 工具对词汇多样性进行全面分析,并利用 Python 编程语言及自然语言处理工具深入探讨句法复杂度。最后,通过 JASP 软件执行贝叶斯 Wilcoxon 符号秩检验,以定量方式对比不同翻译的表现。

3.1 语料来源

本研究自建的三个语料库语料选自中国日报网¹的 200 篇汉英双语文章,每篇的汉语原文和对应的英语译文分别构成两个语料库:汉语原文语料库和人

¹ 经官方确认,该网站的双语文章为纯人工翻译,并非译后编辑版本,确保了译文的可靠性和可追溯性。

工英译语料库。然后,将汉语文本依次输入 ChatGPT-4 中,并使用“汉译英”翻译指令,收集翻译结果,构建 ChatGPT 英译语料库。鉴于 DeepL 在翻译领域表现出色,将其作为比较对象。DeepL 和 ChatGPT 代表了不同类型的语言模型。DeepL 是一个专注于机器翻译的深度学习模型,而 ChatGPT 是一种用于自然语言生成和理解的预训练语言模型。通过对比这两个模型,可以凸显它们在语言处理中的表现差异。依次将汉语文本输入 DeepL 翻译程序中获得译文,以构建 DeepL 英译语料库。

3.2 研究工具

本研究采用 TAALED 工具进行词汇多样性的全面分析。TAALED 是一个专门用于评估文本中词汇丰富性和多样性的分析工具。它提供了一系列指标和计算方法,允许用户深入分析文本的词汇结构和复杂性。这些指标涵盖了从基本的词汇统计(如词汇数和词型数)到更复杂的词汇多样性指标(如 TTR)。为了克服 TTR 在评估不同长度文本时的不稳定性, TAALED 还整合了 MTLD 指标(Kyle, Crossley & Jarvis 2021)。MTLD 可通过计算在文本中维持预定 TTR 值的连续单词串的平均长度来衡量文本的词汇多样性(McCarthy & Jarvis 2010)。总的来说, TAALED 为研究者和教育工作者提供了一个全面而强大的工具,使我们能够深入了解文本的词汇结构,从而更好地评估和理解文本的词汇多样性。

本研究使用多个综合性指标来定量分析句法复杂度。其中,平均句子长度、平均从句长度和平均 T 单位长度用于反映句子结构的基本复杂度;从句比例、并列短语比例和复杂名词短语比例则衡量不同语法结构的使用频率;动词短语数量和复杂度、嵌套从句数量和结构深度进一步揭示了语句的复杂层次。通过这些多元化的指标,我们能够全面评估不同翻译模式对文本句法结构的影响,这对于翻译质量的评估和语言学研究具有重要意义。具体而言,本研究利用 Python 编程语言,使用自然语言处理库,对文本数据进行解析并提取关键的句法特征。这种方法使我们能够以定量的方式深入分析句法特征,并为句法复杂度的评估提供一个全面的度量标准。

本研究运用 JASP 软件来执行贝叶斯 Wilcoxon 符号秩检验,对比 ChatGPT 翻译(CGT)、人工翻译(HT)和 DeepL 翻译(DT)的差异程度。我们提出以下假设:备择假设(alternative hypothesis, 简称为 H1)认为 CGT 在词汇多样性和句法复杂度方面优于 HT 或 DT;相对应的零假设(null hypothesis, 简称为 H0)则认为 CGT 在词汇多样性和句法复杂度方面并不优于 HT 或 DT。假设是否成立主要通过贝叶斯因子(Bayes Factor, BF)的大小来判断。贝叶斯因子定量评估零假设和备择假设的相对支持程度(Masson 2011)。BF 的值超过 1 表明数据

更倾向于支持 H1,小于 1 则表明更支持 H0。BF 的值越大,表示对 H1 的支持越坚定。具体而言,BF 值介于 1—3 之间通常被认为提供了非常微弱的支持;3—20 之间被视为一定程度的支持;20—150 之间被认为是较强的支持;超过 150 则被视为非常强的支持(Kass & Raftery 1995)。Wilcoxon 秩和统计量(W)反映了两组数据间的差异程度(Bergmann, Ludbrook & Spooren 2000);Rhat 值指示模型的稳定性,接近 1 通常表明所有拟合模型都具有良好的模型收敛性(Schad, Betancourt & Vasishth 2021)。

4. 研究结果及讨论

4.1 ChatGPT 翻译与人工翻译的词汇多样性比较

CGT 和 HT 在多个词汇特征指标上的表现见表 1。

表 1. CGT 与 HT 词汇多样性对比

指标	BF	W	Rhat
词汇数	0.04	9,181.00	1.00
词型数	0.07	9,726.00	1.01
内容词词汇数	0.24	10,535.50	1.00
内容词词型数	1.25	10,928.00	1.00
功能词词汇数	0.02	6,823.50	1.01
功能词词型数	0.03	4,529.50	1.02
词型词汇密度	15,544.61	15,074.00	1.01
词项词汇密度	1,134.50	14,384.00	1.01
所有词 TTR	0.08	10,541.00	1.00
内容词 TTR	0.06	10,321.00	1.00
功能词 TTR	0.03	8,317.00	1.00
所有词 MTLD	2.06	11,611.00	1.00
内容词 MTLD	6.11	12,197.00	1.00
功能词 MTLD	0.04	7,772.00	1.00

表 1 显示,在内容词词型数和所有词 MTLD 这两个指标上,BF 值分别为 1.25、2.06,非常微弱地支持 H1,表明 CGT 相较于 HT 表现出微弱的优越性;在内容词 MTLD 指标上,BF 值为 6.11,一定程度上支持 H1;在词型词汇密度和词项词汇密度这两个指标上,BF 值分别高达 15,544.61、1,134.50,表明有非常强的证据支持 H1。上述各指标的 Rhat 值均接近或略大于 1,说明了模型拟合结果的稳定性。然而,对于其他词汇多样性指标(如词汇数、词型数、内容词词汇数、功能词词汇数、功能词词型数,以及不同类型的 TTR 指标),BF 值均小于

1,表明 CGT 的表现并没有显著超越 HT。总体而言,上述结果表明 CGT 在内容词的多样性和词汇密度方面优于 HT。这些特点说明了 CGT 能够在保持原文意义的同时,还能使用多样化的词汇和结构来表达相同的概念,这可能有助于提高内容的丰富性和词汇的多样性。

4.2 ChatGPT 翻译与人工翻译的句法复杂度比较

表 2 呈现了 CGT 和 HT 在句法复杂度方面的表现。

表 2. CGT 与 HT 句法复杂度对比

指标	BF	W	Rhat
平均句子长度	0.04	2,528.00	1.05
平均从句长度	0.03	5,739.50	1.03
平均 T 单位长度	0.03	2,528.00	1.05
从句比例	0.03	2,486.00	1.05
并列短语比例	45,028.26	16,646.50	1.00
复杂名词短语比例	0.03	2,552.00	1.02
动词短语数量	0.04	4,336.00	1.04
动词短语复杂度	447.46	13,575.50	1.00
嵌套从句数量	0.05	2,765.50	1.04
嵌套结构深度	0.03	3,068.00	1.05

在平均句子长度、从句比例、复杂名词短语比例等指标上,贝叶斯因子均小于 1,说明数据支持 H0,即 CGT 相较于 HT 并没有表现出更高的句法复杂度。然而,在并列短语比例和动词短语复杂度两个指标上,CGT 的句法复杂度显著高于 HT,BF 值分别为 45,028.26 和 447.46,表明有非常强的证据支持 H1,说明 CGT 在这些特定句法结构上表现出更高的复杂度。总的来说,这些结果揭示了 CGT 在句法复杂度方面的两个主要特点:一方面,在某些句法结构(如平均句子长度、从句使用和复杂名词短语)上,它倾向于更简单或直接的风格。这可能是因为 CGT 更注重可读性和准确性,而不是刻意增加句子的复杂性。另一方面,在并列短语和动词短语的使用上,它展现出较高的复杂度和精细度。这可能是由于它在翻译过程中的算法优化,使其能够在保持文本清晰度的同时,也能有效地使用复杂的句法结构。这种平衡可能是影响 ChatGPT 翻译质量的一个关键因素。

4.3 ChatGPT 翻译与 DeepL 翻译的词汇多样性比较

对 CGT 与 DT 在不同词汇多样性指标上性能的比较见表 3。

表 3. CGT 与 DT 词汇多样性对比

指标	BF	W	Rhat
词汇数	0.04	2,669.00	1.06
词型数	4.41	11,064.00	1.00
内容词词汇数	0.02	3,985.00	1.02
内容词词型数	35.88	11,977.00	1.00
功能词词汇数	0.04	2,084.50	1.25
功能词词型数	0.02	6,602.00	1.00
词型词汇密度	2,976.96	13,270.00	1.00
词项词汇密度	72,368.07	16,264.00	1.00
所有词 TTR	97,141.71	18,610.00	1.05
内容词 TTR	5.315×10^{^7}	17,895.00	1.01
功能词 TTR	24,111.18	16,332.00	1.06
所有词 MTLD	9.456×10^{^8}	18,619.00	1.04
内容词 MTLD	1.129×10^{^8}	18,091.00	1.03
功能词 MTLD	22,404.33	13,686.00	1.00

在词型数上 (BF = 4.41), 数据非常微弱地支持 H1, 在内容词的词型数上 (BF = 35.88), 有较弱的证据支持 H1, 显示出 CGT 在内容词的词型多样性上具有显著优势。在词型词汇密度 (BF = 2,976.96)、词项词汇密度 (BF = 72,368.07)、所有词 TTR (BF = 97,141.71)、内容词 TTR (BF = 5.315 × 10^{^7}) 以及功能词 TTR (BF = 24,111.18) 这些指标上, 有非常强的证据支持 CGT 的表现均优于 DT。此外, 所有词 MTLD (BF = 9.456 × 10^{^8})、内容词 MTLD (BF = 1.129 × 10^{^8}) 和功能词 MTLD (BF = 22,404.33) 的 BF 值大幅超过 150, 进一步强化了这一结论。然而, 词汇数、内容词词汇数、功能词词汇数以及功能词词型数的 BF 值均小于 1, 表明 CGT 在这些维度上并未显著优于 DT。显然, CGT 在维持和增强词汇多样性方面表现出了显著优势, 尤其是在词汇密度、内容词和功能词的使用方面超过了 DT。

CGT 在内容词的词型数上显著优于 DT, 说明 CGT 使用了更丰富多样的内容词汇, 反映出它在理解和生成涵盖广泛主题和概念的文本方面的强大能力。在词型词汇密度和词项词汇密度上, CGT 较 DT 有更好的表现, 表明 CGT 能够更高效地利用语言资源, 通过不同的词型和词项来增强文本的表达力。CGT 在 TTR 和 MTLD 上的优势显示出它在长篇文本中保持词汇多样性的强大能力, 这对于避免重复和俗套表达至关重要。这些结果均表明 CGT 在词型多样性和词汇使用的丰富性方面相较于 DT 有显著优势, 尤其是在处理内容词和保持长文

本中的词汇多样性方面。

4.4 ChatGPT 翻译与 DeepL 翻译的句法复杂度比较

表 4 展示了 CGT 与 DT 在句法复杂度方面的差异。

表 4. CGT 与 DT 句法复杂度对比

指标	BF	W	Rhat
平均句子长度	0.04	870.50	1.02
平均从句长度	0.03	5,237.00	1.02
平均 T 单位长度	6.237×10^{^15}	20,100.00	1.10
从句比例	0.03	726.00	1.00
并列短语比例	0.05	39.00	1.07
复杂名词短语比例	0.04	0.04	1.07
动词短语数量	0.04	5,964.50	1.03
动词短语复杂度	0.06	9,687	1.00
嵌套从句数量	0.02	4,878.50	1.02
嵌套结构深度	0.03	4,509.50	1.00

在平均句子长度、平均从句长度、从句比例、并列短语比例、复杂名词短语比例、动词短语数量、动词短语复杂度、嵌套从句数量以及嵌套结构深度指标上, BF 值均未超过 1,说明 CGT 在这些句法复杂度指标上超越 DT 的证据不充分。然而,在平均 T 单位长度的指标上, BF 值高达 6.237×10^{15} ,有非常强的证据支持 H1。T 单位被定义为“一个主句加上所有从属于它或嵌入其中的从句和非从句结构”(Hunt 1970: 4)。平均 T 单位长度是衡量句子复杂性的一个指标 (Gaies 1980),较长的平均 T 单位长度通常表明文本采用了更多的从句和短语结构,这不仅丰富了信息内容,也提升了表达的精细度。在机器翻译中,不同的算法可能在处理句子结构复杂性方面有不同的表现。CGT 的译文在平均 T 单位长度这一指标上显示出较高的复杂度,这意味着 ChatGPT 在翻译时倾向于生成更长和结构更复杂的句子。这一特点使其适合处理需要详细说明或复杂论述的长文本,同时保持句子的清晰度和连贯性。

本研究发现 ChatGPT 不仅能处理基础的翻译任务,而且在处理复杂和长篇文本方面尤其出色,增加了词汇的多样性和句法的复杂度。在翻译实践中,可以综合应用 CGT、HT 和 DT,以符合不同类型文本的需求,提高翻译质量。人工译者凭借其概念创新、深度共情和语篇重构的能力,展现出不可替代的价值和独特性(杜安 2023)。与此同时,DT 则在准确性和可读性上表现较好(Yulianto & Supriatnaningsih 2021),尤其适用于标准化和通用文本的处理。通过这种分层

策略,每种翻译模式都能够在各自擅长的领域内充分发挥自身优势,提升整体翻译质量。例如,可以首先利用 DeepL 或 ChatGPT 快速完成文本的初步翻译。随后,可以借助 ChatGPT 的灵活翻译指令进行优化,特别是在提高词汇多样性和句法复杂度方面。可以通过特定的翻译指令来强调使用专业词汇,如在翻译科技文章时指示 ChatGPT “将以下中文科技文章翻译成英文,并使用丰富的专业术语”。同样,为了保持文本的原始风格和复杂性,可以指示 ChatGPT 生成更复杂的句子结构,如“请翻译以下文本为英文,并保持其原有的句法复杂度和风格”。这样的指令能够使翻译结果更加贴近原文的意图和风格。同时,还可以结合人工翻译的深度理解,对文本进行更深层次的优化和细化。这种综合方法结合了机器翻译的高效率和人工翻译的细腻度,有助于提高翻译质量,体现了在现代翻译实践中 CGT 和 HT 相互辅助、共同发展的趋势。换言之,CGT 能够在与用户的实时交流中,连续产生新的翻译版本。这个过程是动态和持续进化的,意味着翻译不是一成不变的,而是可以根据交流内容不断改进。这种方式不仅增强了人工与机器之间的协作,还促进了它们之间的相互支持和发展,形成了一种新的人机合作翻译模式(杜安 2023)。

5. 结语

通过综合对比 ChatGPT 翻译、人工翻译和 DeepL 翻译在词汇多样性和句法复杂度方面的表现,本研究证实了 ChatGPT 翻译在词汇密度和内容词的使用上相较于传统翻译方法具有明显优势,这可能归因于其在预训练过程中所接触到的丰富语料。此外,在并列短语和动词短语复杂度方面,ChatGPT 翻译的句子结构比人工翻译更丰富、层次感更强。在平均 T 单位长度的句法复杂度方面,ChatGPT 翻译也超越了 DeepL 翻译,展现出其在构建长句和复杂句型方面的优异能力。这些发现为有效利用 ChatGPT 提升翻译质量提供了重要参考。翻译实践可结合 ChatGPT 翻译和人工翻译,以充分发挥各自优势,提高翻译效率和质量。ChatGPT 能够快速生成初步翻译,为人工翻译提供有益参考,同时人工翻译可以优化 ChatGPT 生成的翻译,确保最终结果更符合语境和专业要求。需指出的是,尽管 ChatGPT 在这些方面展现出强大的潜力,但它在某些特定语境和文本体裁中可能存在局限性。因此,未来应深入探究 ChatGPT 在接收特定指令时调整翻译输出的能力,并评估其对跨语言及不同文本体裁的适应性。

参考文献

- Al-Jabr, A. 2006. Effect of syntactic complexity on translating from/into English/Arabic [J]. *Babel* 52(3): 203–221.
- Ayala, B. & Jiangping Chen. 2017. A machine learning approach to evaluating translation quality

- [R]. Paper presented at the 17th ACM/IEEE Joint Conference on Digital Libraries, Toronto, Canada, June 2017.
- Bergmann, R., J. Ludbrook & W. Spooren. 2000. Different outcomes of the Wilcoxon-Mann-Whitney test from different statistics packages [J]. *The American Statistician* 54(1): 72–77.
- Brglez, M. & Š. Vintar. 2022. Lexical diversity in statistical and neural machine translation [J]. *Information* 13(2): 93–106.
- Du, An [杜安]. 2023. Exploring the value spaces of human translation in the ChatGPT era and the transitions needed for translation education [J]. *Foreign Languages and Cultures* (4): 90–103. [ChatGPT 时代人工翻译的价值空间与翻译教育转型, 《外国语言与文化》4]
- Fu, Yingxue & M. Nederhof. 2021. Automatic classification of human translation and machine translation: A study from the perspective of lexical diversity [R]. Paper presented at the First Workshop on Modelling Translation: Translatology in the Digital Age, Online, May 2021.
- Gaies, S. 1980. T-unit analysis in second language research: Applications, problems and limitations [J]. *TESOL Quarterly* 14(1): 53–60.
- Helberger, N. & N. Diakopoulos. 2023. The European AI act and how it matters for research into AI in media and journalism [J]. *Digital Journalism* 11(9): 1751–1760.
- Herbold, S., *et al.* 2023. A large-scale comparison of human-written versus ChatGPT-generated essays [J]. *Scientific Reports* 13(1): Article No. 18617.
- Hunt, K. 1970. Syntactic maturity in schoolchildren and adults [J]. *Monographs of the Society for Research in Child Development* 35(1): iii–67.
- Jiang, Kai & Xi Lu. 2020. Natural language processing and its applications in machine translation: A diachronic review [R]. Paper presented at the Third International Conference of Safe Production and Informatization, Chongqing, China, November 2020.
- Johansson, V. 2008. Lexical diversity and lexical density in speech and writing: A developmental perspective [J]. *Lund Working Papers in Linguistics* 53: 61–79.
- Kass, R. & A. Raftery. 1995. Bayes factors [J]. *Journal of the American Statistical Association* 90(430): 773–795.
- Kim, H. 2020. The effects of lexical diversity and lexical sophistication of English on Korean-English translation [J]. *The Journal of Translation Studies* 21(2): 43–65.
- Koizumi, R. 2012. Relationships between text length and lexical diversity measures: Can we use short texts of less than 100 tokens [J]. *Vocabulary Learning and Instruction* 1(1): 60–69.
- Kyle, K., S. Crossley & S. Jarvis. 2021. Assessing the validity of lexical diversity indices using direct judgements [J]. *Language Assessment Quarterly* 18(2): 154–170.
- Liang, Junying & Yiguang Liu [梁君英、刘益光]. 2023. Human intelligence advantages in translation competence: A corpus-based comparative study of human translation and machine translation [J]. *Foreign Languages and Their Teaching* (3): 74–84. [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, 《外语与外语教学》3]
- Liu, K. & M. Afzaal. 2021. Syntactic complexity in translated and non-translated texts: A corpus-based study of simplification [J]. *PLOS ONE* 16(6): Article No. e0253454.
- Lü, Qing, *et al.* 2023. Translating radiology reports into plain language using ChatGPT and GPT-4 with prompt learning: Results, limitations, and potential [J]. *Visual Computing for Industry, Biomedicine, and Art* 6: Article No. 9.

- Ma, Xingcheng. 2021. Coping with syntactic complexity in English-Chinese sight translation by translation and interpreting students: An eye-tracking investigation [J]. *Across Languages and Cultures* 22(2): 192–213.
- Masson, M. 2011. A tutorial on a practical Bayesian alternative to null-hypothesis significance testing [J]. *Behavior Research Methods* 43(3): 679–690.
- McCarthy, P. 2005. An Assessment of the Range and Usefulness of Lexical Diversity Measures and the Potential of the Measure of Textual, Lexical Diversity (MTLD) [D]. Ph.D. Dissertation. The University of Memphis.
- McCarthy, P. & S. Jarvis. 2010. MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment [J]. *Behavior Research Methods* 42(2): 381–392.
- Ortega, L. 2003. Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing [J]. *Applied Linguistics* 24(4): 492–518.
- Ray, P. 2023. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope [J]. *Internet of Things and Cyber-Physical Systems* 3: 121–154.
- Sanz-Valdivieso, L. & B. López-Arroyo. 2023. Google Translate vs. ChatGPT: Can non-language professionals trust them for specialized translation? [R]. Paper presented at the International Conference on Human-informed Translation and Interpreting Technology, Naples, Italy, July 2023.
- Schad, D., M. Betancourt & S. Vasishth. 2021. Toward a principled Bayesian workflow in cognitive science [J]. *Psychological Methods* 26(1): 103–126.
- Wang, Lü & Xiangling Wang [王律、王湘玲]. 2023. The study on the training model of machine translation post-editing competence in the era of ChatGPT [J]. *Technology Enhanced Foreign Languages* (4): 16–23. [ChatGPT 时代机器翻译后编辑能力培养模式研究,《外语电化教学》4]
- Xu, Jiajin & Jialei Li. 2021. A syntactic complexity analysis of translational English across genres [J]. *Across Languages and Cultures* 22(2): 214–232.
- Yulianto, A. & R. Supriatnaningsih. 2021. Google Translate vs. DeepL: A quantitative evaluation of close-language pair translation (French to English) [J]. *AJELP: Asian Journal of English Language and Pedagogy* 9(2): 109–127.
- Zhu, Guanghui & Xiwen Wang [朱光辉、王喜文]. 2023. ChatGPT: Operation mode, key technology and future prospects [J]. *Journal of Xinjiang Normal University (Philosophy and Social Sciences)* (4): 113–122. [ChatGPT 的运行模式、关键技术及未来图景,《新疆师范大学学报(哲学社会科学版)》4]
- Zhu, Zhouye & Jinquan Wang [朱周晔、王金铨]. 2023. Effects of dependency distance of source texts on complexity and quality in Chinese EFL learners' Chinese-English translation [J]. *Foreign Language Teaching and Research* (6): 890–900. [源语依存距离对中国英语学习者汉英译文复杂度和质量的影响,《外语教学与研究》6]

收稿日期: 2023-11-11;修改稿 2024-01-14;本刊修订 2024-01-20

通讯地址: 100083 北京市 北京语言大学外国语学部

effect analysis revealed that PEU and PU mediated the relationship between CSE and BI, with the former being an antecedent to the latter. The pathway with the largest indirect effect size is CSE→PU→ATT→BI. The implications for teachers and platform designers are also discussed.

A study of the transfer of L1 and L2 on L3 writing task in terms of aspectual usage pattern (p. 274)

SUN Yuliang (School of Foreign Languages, Soochow University, Suzhou 215000, China)

This study aimed to investigate the transfer of L1 Chinese and L2 English on the use of grammatical aspect in L3 Spanish using guided writing compositions by 18 Chinese learners of Spanish with varying linguistic proficiency levels. The results suggest that there is no significant difference between the productions of Chinese and Spanish native speakers when the perfective and imperfective aspects are aligned in form and meaning across these three languages. By contrast, when discrepancies are present at the form or meaning level, a negative transfer from learners' L1 and L2 is observed in their production of L3 Spanish. Furthermore, this study revealed that negative transfer caused by discrepancies at the form level was observed only in the group with lower proficiency. However, a negative transfer from the meaning level persisted in the group with higher proficiency.

Human-AI interactive negotiation competence: ChatGPT and foreign language education (p. 286)

WEN Qiufang (National Research Centre for Foreign Language Education/National Research Centre for State Language Capacity, Beijing Foreign Studies University, Beijing 100089, China)

LIANG Maocheng (School of Foreign Languages, Beihang University, Beijing 100191, China)

Generative Artificial Intelligence (AI) systems such as ChatGPT are widely applied worldwide. At the same time, the focus has shifted from “whether AI should be used” to “how to effectively utilize AI.” In practice, we found that human-AI interactive negotiation competence (HAINC) is a key factor affecting the effectiveness of AI applications. Based on our experiential study, the HAINC was divided into five essential components: understanding AI, setting goals, issuing instructions, analyzing feedback, and adjusting strategies. Together, these five elements constitute the dynamic process of human-AI interactive negotiation. To enhance the HAINC, we proposed the establishment of a specialized course in foreign language education. The course aims to achieve two main objectives: to enhance students' HAINC and improve their English learning efficiency. The course content includes three modules: an introduction to AI, case-based teaching, and project operations. In the teaching process, we emphasized the teaching principles of practice-oriented, incremental learning, and timely feedback.

Lexical diversity and syntactic complexity in ChatGPT translation (p. 297)

YU Lei (Faculty of Foreign Studies, Beijing Language and Culture University, Beijing 100083, China)

The versatile language model ChatGPT has captured the attention of researchers owing to its application and performance in translation tasks. This study employed the TAALED tool and Python programming to extract and compare metrics of lexical diversity and syntactic complexity in translations produced by ChatGPT, human translators, and DeepL. The findings indicate that ChatGPT translation has a significant advantage over the other two methods in terms of lexical density and content words, showcasing its breadth and innovation in vocabulary selection. In terms of syntactic complexity, ChatGPT translation is similar to both human translation and DeepL translation. Notably, ChatGPT translation excels in the proportion of coordinate phrases and complexity of verb phrases compared to human translation. Furthermore, it surpasses the DeepL translation in terms of the average length of T units, underscoring its proficiency in formulating intricate sentences. These findings highlight the potential of ChatGPT translation to effectively handle linguistic diversity and complexity, shedding light on how ChatGPT can be strategically employed to enhance translation quality.