ISSN 1799-2591
Theory and Practice in Language Studies, Vol. 15, No. 9, pp. 2867-2875, September 2025
DOI: https://doi.org/10.17507/tpls.1509.08

Lexical Diversity and Syntactic Complexity in AI-Translated Legislative Texts*

Jialei Chen
School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China

Qiufeng Hong
School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China

***Abstract*—This study investigates the performance of artificial intelligence (AI) in the translation of legislative texts, focusing on the quality of translations produced by ChatGPT 4o and DeepL Pro. By using TAALED and NeoSCA, we evaluated and compared a number of indices in lexical diversity and syntactic complexity of AI-generated and human translations of twenty Chinese legislative texts. We used JASP to calculate Bayes Factor and then compare the translations of human and AI. Our findings indicate that while AI models demonstrate notable strengths in function words diversity and coordinate syntactic structures, they still lag behind human translators in overall lexical diversity and syntactic complexity. The study underscores the potential and limitations of AI in legal translation, highlighting the necessity for human-AI collaboration to achieve high-quality translations in this specialized field.**

***Index Terms*—lexical diversity, syntactic complexity, AI, legislative texts**

I. Introduction

Artificial intelligence (AI) has made significant strides in the field of translation, offering innovative solutions and enhancing the efficiency of translating diverse types of texts. However, the application of AI in specialized domains such as legal translation presents unique challenges due to the intricate and system-bound nature of legal language. Legal translation requires not only a high degree of precision but also a deep understanding of both the source and target legal systems, which often include culturally specific references and complex terminologies. This complexity makes legal translation a time-consuming and labor-intensive process.

The unique demands of legal translation mean that traditional AI translation systems often fall short in this domain. For instance, Choudhury and McConnell (2013) pointed out that business translation systems continue to face difficulties in achieving a balance between cost, quality, and translation speed to market. This is particularly pertinent in the legal sector, where inaccuracies can have significant ramifications. As a result, law firms are increasingly interested in integrating AI with legal translation workflows to enhance efficiency and reduce costs without compromising quality. Recent advancements in AI, particularly in natural language processing and machine learning, have opened new possibilities for legal translation. Tools like ChatGPT, Copilot and DeepL are beginning to show promise in handling the nuanced requirements of legal texts. These AI models can assist human translators by automating repetitive tasks, suggesting contextually appropriate translations, and ensuring consistency across documents. However, the effectiveness of these tools in legal translation still requires rigorous evaluation and refinement.

Given this context, our study aims to evaluate the quality of AI-translated legislative texts through a corpus-based analysis. We focus on two key aspects: lexical diversity and syntactic complexity. By comparing translations of twenty Chinese legislative acts produced by human translators, ChatGPT 4o, and DeepL pro, we seek to determine the effectiveness and reliability of AI translators in handling the intricate requirements of legislative language. This evaluation will provide valuable insights into the capabilities and limitations of AI in the realm of legal translation, contributing to the ongoing discourse on the integration of AI in professional translation practices.

II. Literature Review

The emergence of artificial intelligence (AI) has revolutionized translation industry. Studies have demonstrated that properly trained and fine-tuned AI models have considerable promise in translation tasks (Nunes & Alonso, 2020; Yuan, 2023; Herbold et al., 2023). While AI's potential to streamline translation processes is undeniable, the scholar community has been diligent in scrutinizing the myths and realities of AI models, particularly in specialized fields such as legal translation.

Legal translation stands out as a domain replete with challenges due to its system-bound and historically rooted nature, which imbues legal languages with unique cultural references and terminological complexities (Prieto Ramos,

* This article is supported by the 2024 Student Innovative Science Research Project of Southwest University of Political Science and Law (Grant No. 2024XZXS-316).

© 2025 ACADEMY PUBLICATION

2024). Legal translation involves communicating not only between legal languages or discourses but also between legal systems and legal genres (Scott, 2019). Therefore, it has generally been considered unsuitable to employ AI in legal translation. Translators in this field remain skeptical about the quality of AI outputs.

The quality of AI translation has been a subject of considerable debate and research. Wang and Mao (2023) use the Chinese and English introduction texts from the Zibo Ceramics and Glaze Museum as their corpus. They apply automatic evaluation methods, BLEU and TER to comprehensively assess the translation performance and quality of ChatGPT and three other machine translation tools. They also summarize excellent prompts that can be applied in the field of Chinese-English translation of ceramics-related texts. Wang and Ma (2023) use medical texts as their corpus to compare ChatGPT with commonly used translation tools. They analyze the quality of the translations from the perspective of vocabulary translation. Wen and Tian (2024) employ BLEU and TER to compare the translation generated by ChatGPT of China-specific discourse text. They observe that while ChatGPT offers some advantages over the other three translation tools, it still falls short in areas such as handling ideological content, complex sentence structures, culturally loaded expressions, recognising metaphors and metonymy, and ensuring translation accuracy. BLEU and TER are frequently used to evaluate translation quality. Traditional metrics such as accuracy and fluency are no longer sufficient to capture the nuances of translation quality. Recent studies have begun to incorporate measures of lexical diversity and syntactic complexity as critical indicators of translation quality. For instance, Yu (2024) employs 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. Lin et al. (2023) analyse the linguistic difficulty of translated legal texts in comparison to the syntactic complexity of native English legal writings, aiming to reveal statistically significant differences between the two large datasets.

Lexical diversity (LD) reflects the richness of a writer's vocabulary and is commonly measured by indices such as Type-Token Ratio (TTR) and its variants, which have been widely utilized in writing assessment research. However, the sensitivity of these indices to text length has raised concerns, as they may conflate lexical breadth with fluency. Scholars like McCarthy and Jarvis (2007, 2010) have proposed measures such as MTLD to address this issue, demonstrating greater stability across a range of text lengths. Syntactic complexity, on the other hand, pertains to the intricacy of grammatical structures employed by language learners and has been linked to L2 development. Researchers such as Ortega (2003) and Wolfe-Quintero et al. (1998) have underscored the importance of distinguishing between measures that evaluate first language acquisition and those pertinent to second language learners, given the distinct processes involved. Advancements in computational linguistics have ushered in tools that automate the analysis of these constructs, as illustrated by the work of Lu (2010), which introduces a computational system for the automatic analysis of syntactic complexity in L2 writing. This system, leveraging deep syntactic parsing, offers a significant advancement over manual analysis, thereby facilitating large-scale corpus studies.

In the legal domain, the distinction between human translation and AI translation is not merely a matter of preference but one of necessity. Legal language is rife with technical terms and complex sentence structures that often require a deep understanding of both the language and the legal context. Yang (2023) uses ChatGPT to translate Vietnamese legal texts and compared the results with other machine translation and human translation outputs. By summarizing its strengths and weaknesses, he reflects on the implications of ChatGPT for translators. The integration of AI in translation tasks, particularly with models like ChatGPT, has further revolutionized the way we approach translation quality. Studies, such as the one conducted by Yu (2024), reveal that while AI models show promise in enhancing lexical diversity and syntactic complexity, they also pose challenges in specialized domains like legal translation, which demand a high level of precision and understanding of context-specific language. Her study highlights the potential of AI to improve legal translation; however, it also underscores the challenges AI faces in specialized areas, including the legal field.

Building upon this, Moneus and Sahari (2024) present a comprehensive analysis of human and AI translation within the legal sphere. The study meticulously evaluates the strengths and weaknesses of both translation approaches, providing insights into their effectiveness and applicability in the translation of legal texts. By integrating the discussion on lexical diversity, syntactic complexity, and the role of AI in translation, this document seamlessly transitions into a focused examination of AI's role in legal translation.

While the study shed light on the capabilities of AI in enhancing lexical diversity and syntactic complexity within legal translation, it may have overlooked the multi-dimensional nature of this specialized field by employing a limited set of parameters. The complexity of legal language, with its technical jargon and structured syntax, necessitates a more robust analytical framework that goes beyond traditional metrics. In response to this gap, our research will conduct a comparative analysis of Chinese legal texts translated by human translators and AI, employing an expanded set of parameters. By doing so, the study aims to provide a more comprehensive assessment of AI's efficacy in the domain of Chinese legal translation, offering deeper insights into its potential and limitations in this high-stakes area. This study aims to answer the following two research questions:

(1) What is the difference between AI-translated legislative texts and human translation in terms of lexical diversity and syntactic complexity?

(2) What implications can we get from the difference between AI translations and human translations?

© 2025 ACADEMY PUBLICATION

III. METHODS

A. Corpus

The corpus of this study comprises translations of twenty legislative texts of China. The translations are produced respectively by human translators, ChatGPT 4o, and DeepL pro. We randomly select and download twenty Chinese legislative texts (as shown in Table 1) and their official English translation produced by human translators from the official website of the Supreme People's Court of the People's Republic of China.

TABLE 1
LIST OF SELECTED LEGISLATIVE TEXTS

No.	Title
1	Archives Law of the People's Republic of China
2	Customs Law of the People's Republic of China
3	Law of the People's Republic of China on Guarding State Secrets
4	Law of the People's Republic of China on the Prevention and Control of Atmospheric Pollution
5	Compulsory Education Law of the People's Republic of China
6	Frontier Health and Quarantine Law of the People's Republic of China
7	Law of the People's Republic of China on Control of the Entry and Exit of Aliens
8	Law of the People's Republic of China on the Control of the Exit and Entry of Citizens
9	Law of the People's Republic of China on Prevention and Control of Water Pollution
10	Law of the People's Republic of China on Protection of Cultural Relics
11	Marine Environment Protection Law of the People's Republic of China
12	Military Service Law of the People's Republic of China
13	Law of the People's Republic of China on Penalties for Administration of Public Security
14	Law of the People's Republic of China on Public Servants
15	Notarization Law of the People's Republic of China
16	Emergency Response Law of the People's Republic of China
17	Law of the People's Republic of China on Animal Epidemic Prevention
18	Law of the People's Republic of China on Lawyers
19	Law of the People's Republic of China on the Administration of the Urban Real Estate
20	Law of the People's Republic of China on Urban and Rural Planning

First, we use ChatGPT 4o to translate the selected twenty legislative texts. We translate them in one conversation window and use the same prompt. Jiao et al. (2023) evaluate ChatGPT's performance in three areas: translation prompts, multilingual translation, and translation robustness. They find that among three different styles of translation prompts, the prompt "Please provide the [TGT] translation for these sentences" elicits the best machine translation results. Based on their findings, we add the description of the text in the prompt. Thus, the prompt we use is "The following text is [Act Title], please provide the English translation for the text". Next, we use DeepL pro to translate the selected legislative text. DeepL and ChatGPT represent different types of language models. DeepL is a deep learning model focused on machine translation, while ChatGPT is a pre-trained language model used for natural language generation and understanding. By comparing these two models, the differences in their performance in language processing can be highlighted (Yu, 2024).

B. Measures

We evaluate the quality of AI-translated legislative texts from two aspects: lexical diversity (LD) and syntactic complexity (SC). Lexical diversity (LD) refers to the variety of distinct words employed in a text—the broader the range, the higher the diversity (McCarthy & Jarvis, 2010).

For lexical diversity, we use TAALED (Kyle et al., 2021) to examine the lexical structure and complexity of the corpus. TAALED is a tool designed to analyze lexical diversity. It calculates indices like type-token ratio (TTR) and also integrates measure of textual lexical diversity (MTLD). MTLD is a sequentially assessed measure of a text's LD, calculated as the average length of word strings that sustain a predetermined TTR threshold (McCarthy & Jarvis, 2010). The indices selected to demonstrate lexical diversity in this study is shown in Table 2.

© 2025 ACADEMY PUBLICATION

TABLE 2[1]
INDICES FOR LEXICAL DIVERSITY

Index Name	Brief Description
Tokens	Number of tokens.
Types	Number of types.
Tokens of Content Words	Number of content word tokens.
Types of Content Words	Number of content word types.
Tokens of Function Words	Number of function word tokens.
Types of Function Words	Number of function word types.
Lexical Density of Types	Number of content words types divided by number of total number of types.
Lexical Density of Tokens	Number of content words tokens divided by number of total number of tokens.
TTR of All Words	Type-token ratio for all words.
TTR of Content Words	Type-token ratio for content words.
TTR of Function Words	Type-token ratio for function words.
MTLD of All Words	MTLD is based on the average number of tokens it takes to reach a given TTR value.
MTLD of Content Words	MTLD is based on the average number of content word tokens it takes to reach a given TTR value.
MTLD of Function Words	MTLD is based on the average number of function word tokens it takes to reach a given TTR value.

In terms of syntactic complexity, we employ NeoSCA (Tan, 2024) to analyze the corpus. NeoSCA is a fork of Lu's L2 Syntactic Complexity Analyzer (L2SCA) (Lu, 2010) and Lexical Complexity Analyzer, featuring cross-platform compatibility and a graphical interface. Lu (2010) selects fourteen syntactic complexity measures and classifies them into five types (as shown in Table 3).

TABLE 3
LU'S (2010) FOURTEEN SYNTACTIC COMPLEXITY MEASURES AUTOMATED

Measure	Code	Definition
Type 1: Length of production unit		
Mean length of clause	MLC	# of words / # of clauses
Mean length of sentence	MLS	# of words / # of sentences
Mean length of T-unit	MLT	# of words / # of T-units
Type 2: *Sentence complexity*		
Sentence complexity ratio	C/S	# of clauses / # of sentences
Type 3: *Subordination*		
T-unit complexity ratio	C/T	# of clauses / # of T-units
Complex T-unit ratio	CT/T	# of complex T-units / # of T-units
Dependent clause ratio	DC/C	# of dependent clauses / # of clauses
Dependent clauses per T-unit	DC/T	# of dependent clauses / # of T-units
Type 4: *Coordination*		
Coordinate phrases per clause	CP/C	# of coordinate phrases / # of clauses
Coordinate phrases per T-unit	CP/T	# of coordinate phrases / # of T-units
Sentence coordination ratio	T/S	# of T-units / # of sentences
Type 5: Particular structures		
Complex nominals per clause	CN/C	# of complex nominals / # of clauses
Complex nominals per T-unit	CN/T	# of complex nominals / # of T-units
Verb phrases per T-unit	VP/T	# of verb phrases / # of T-units

This study uses JASP 0.18.3 (JASP Team, 2024) to perform a Bayesian Wilcoxon signed-rank test to compare the differences between human translation (HT), ChatGPT 4o translation, and DeepL Pro translation. We propose the following hypotheses: the alternative hypothesis posits that HT is inferior to ChatGPT 4o translation or DeepL Pro translation in terms of lexical diversity and syntactic complexity; the corresponding null hypothesis posits that HT is superior to ChatGPT 4o translation or DeepL Pro translation in the two aspects. The determination of hypothesis validity is primarily judged by the value of the Bayes Factor (BF). The Bayesian approach generates graded evidence regarding both the alternative and the null hypotheses, including the degree of support favoring the null, and provides a simple means of aggregating evidence across replications (Masson, 2011). A BF value greater than 1 indicates that the data tends to support the alternative hypothesis, while a value less than 1 indicates more support for the null hypothesis. The larger the BF value, the stronger the support for the alternative hypothesis. A Bayes Factor (BF) between 1 and 3 typically indicates very weak support; values from 3 to 20 suggest moderate support; 20 to 150 imply strong support; and values exceeding 150 are taken as very strong support (Kass & Raftery, 1995). The Wilcoxon rank-sum statistic (W) measures the extent of difference between two datasets (Bergmann et al., 2000), while the Rhat value assesses model stability—values close to 1 generally signal good convergence across all fitted models (Schad et al., 2021).

IV. RESULTS AND DISCUSSION

A. LD of Human Translation vs. LD of ChatGPT 4o Translation

As shown in Table 4, BF values of TTR of all words and MTLD of all words are 1.682 and 1.792 respectively, which weakly supports the alternative hypothesis; BF value of TTR of function words is 4.935, suggesting moderate support to the alternative hypothesis; BF value of lexical density of tokens is 945.210, which strongly supports the alternative

[1] Table 2 is revised from the list of definitions of Kyle et al. (2021) in TAALED.

© 2025 ACADEMY PUBLICATION

hypothesis. Additionally, all Rhat values of these indices are equal to or higher than 1, indicating good model convergence and stability of the results. BF values of other indices are lower than 1, suggesting that HT is superior to ChatGPT 4o translation in these aspects.

The data suggests that ChatGPT 4o translation is superior to HT in four aspects which are lexical density of tokens, TTR of all words, TTR of function words, and MTLD of all words. Among these indices, the most significant index of ChatGPT 4o translation is lexical density of tokens which refers to number of content words tokens divided by number of total number of tokens. This indicates that ChatGPT 4o translation is capable of maintaining the original meaning while using diverse vocabulary and structures to express the same concepts. This can potentially enhance the richness of the content and the diversity of the vocabulary.

TTR is a measure of vocabulary "flexibility" or variability, designed to indicate certain aspects of language adequacy (Johnson, 1944). A higher value of TTR means more diversified lexical use. In TTR indices, ChatGPT 4o translation is superior in terms of all words and function words, but a higher value in TTR of function words may influence TTR of all words. This left the BF value of TTR of all words with less reference value. Despite of the influence, BF value of TTR of function words provides moderate evidence supporting the alternative hypothesis, indicating that ChatGPT 4o translation uses more diversified function words than HT. Diverse function words can help in constructing more precise and clearer sentences. Variation in function words can make the text more engaging and easier to read. Moreover, function words like conjunctions, prepositions, and articles play a crucial role in linking ideas and maintaining the flow of the text. Using a variety of these words can improve the overall coherence and cohesion of the text. In summary, using a more diversified set of function words can enhance the clarity, readability, cohesion, coherence, and overall sophistication of a text.

ChatGPT 4o also performs better than human translators in MTLD of all words. This demonstrates its powerful ability to maintain vocabulary diversity in long texts, which is crucial for avoiding repetition and clichéd expressions.

TABLE 4
LD OF HT VS. LD OF CHATGPT 4O TRANSLATION

LD Indices	BF	W	Rhat
Tokens	0.131	140.000	1.000
Types	0.164	165.000	1.000
Tokens of Content Words	0.179	167.000	1.000
Types of Content Words	0.182	167.000	1.001
Tokens of Function Words	0.123	112.000	1.004
Types of Function Words	0.111	120.500	1.002
Lexical Density of Types	0.727	232.000	1.000
Lexical Density of Tokens	**945.210**	399.000	1.016
TTR of All Words	**1.682**	262.000	1.000
TTR of Content Words	0.377	215.000	1.000
TTR of Function Words	**4.935**	280.000	1.002
MTLD of All Words	**1.792**	256.000	1.000
MTLD of Content Words	0.142	144.000	1.002
MTLD of Function Words	0.118	124.000	1.003

B. LD of Human Translation vs. LD of DeepL Pro Translation

Showing in Table 5, lexical density of tokens has a BF value of 1.932, suggesting weak support for the alternative hypothesis that HT is inferior to DeepL Pro translation in this index. In other indices such as basic tokens, TTR of all words, and MTLD of content words, etc., the BF is lower than 1, indicating weak evidence against the null hypothesis that HT is superior. The data indicates that DeepL Pro translation is superior to HT in terms of lexical density of tokens, possibly due to DeepL Pro translation producing more densely packed translations.

The data highlights that DeepL Pro translation's superiority over HT is limited to the lexical density of tokens. A higher lexical density indicates more informative and concise language use. Consequently, DeepL Pro translation translations are often denser and more straightforward, potentially at the expense of some richness and depth. This efficiency and directness in DeepL Pro translation translations might lead to less varied and diverse language use.

In contrast, HT generally produces translations that are richer in vocabulary and exhibit greater linguistic variety, although they might be less dense in terms of information packing. Human translators often prioritize clarity and readability, resulting in translations that are more engaging and natural to read. This approach enhances the overall quality of the translation, making it more accessible and enjoyable for the reader.

In summary, while DeepL Pro translation shows an advantage in producing translations with higher lexical density, it may do so by sacrificing some of the richness and variability that human translations offer. HT remains superior in providing a more varied and nuanced vocabulary, which contributes to clearer, more readable, and more engaging translations. This analysis underscores the strengths and trade-offs between human and AI translations, highlighting the unique value each brings to the translation process.

© 2025 ACADEMY PUBLICATION

TABLE 5
LD OF HT VS. LD OF DEEPL PRO TRANSLATION

LD Indices	BF	W	Rhat
Tokens	0.222	177.000	1.000
Types	0.261	188.000	1.000
Tokens of Content Words	0.251	185.000	1.000
Types of Content Words	0.256	190.000	1.000
Tokens of Function Words	0.200	170.000	1.000
Types of Function Words	0.246	180.500	1.000
Lexical Density of Types	0.300	205.000	1.000
Lexical Density of Tokens	**1.932**	266.000	1.001
TTR of All Words	0.383	214.000	1.001
TTR of Content Words	0.322	202.000	1.000
TTR of Function Words	0.720	233.000	1.001
MTLD of All Words	0.096	99.000	1.002
MTLD of Content Words	0.138	141.000	1.001
MTLD of Function Words	0.103	85.000	1.005

C. SC of Human Translation vs. SC of ChatGPT 4o Translation

Most BF values for syntactic complexity metrics are less than 1, indicating that HT tends to be superior to ChatGPT 4o translation in these aspects. The exception is CP/C (Coordinate Phrases per Clause), with a BF of 1.242, indicating that ChatGPT 4o translation might have a slight advantage in using coordinate phrases within clauses. Only adjective, adverb, noun, and verb phrases are counted in coordinate phrases (Cooper, 1976). It reflects the ability to combine multiple ideas and details within a single clause, contributing to the richness and depth of the text. This suggests that GPT translation can create more complex, coordinated sentence structures.

The analysis shows that HT generally surpasses ChatGPT 4o translation in terms of overall syntactic complexity. the analysis illustrates that human translators excel in overall syntactic complexity. Human translators' context-sensitive syntactic decisions often result in translations that are syntactically superior. They can deftly handle intricate syntactic elements and make sophisticated choices that enhance the readability and coherence of the text. Human translators' expertise in navigating syntactic subtleties ensures that their translations maintain a high level of linguistic finesse and structural sophistication. However, ChatGPT 4o translation demonstrates notable strengths in using coordinate phrases, highlighting its ability to create coordinated and potentially more varied sentence structures. This strength implies that while GPT translation might not yet achieve the same level of nuanced syntactic decisions as human translators, it excels in constructing sentences that are rich in coordinated structures. The use of such structures can enhance the complexity and depth of the translated text, making it more engaging and intricate. This mixed performance underscores the potential and limitations of ChatGPT 4o in handling syntactic complexities. While ChatGPT 4o can adeptly manage coordinate phrases, it still faces challenges in matching the overall syntactic complexity achieved by human translators. The context-sensitive decisions made by human translators contribute to a level of syntactic sophistication that AI has yet to fully replicate. Therefore, while ChatGPT 4o translation demonstrates promising capabilities in specific areas, it continues to rely on human translators for achieving the highest standards of syntactic complexity and overall translation quality.

TABLE 6
SC OF HT VS. SC OF CHATGPT 4O TRANSLATION

SC Indices	BF	W	Rhat
MLS	0.111	284.000	1.003
MLT	0.170	251.000	1.003
MLC	0.243	209.000	1.000
C/S	0.116	303.000	1.003
VP/T	0.174	242.000	1.001
C/T	0.141	263.000	1.005
DC/C	0.114	382.000	1.006
DC/T	0.129	354.000	1.010
T/S	0.186	229.000	1.001
CT/T	0.124	381.000	1.007
CP/T	0.341	206.000	1.000
CP/C	**1.242**	140.000	1.001
CN/T	0.147	269.000	1.008
CN/C	0.145	253.000	1.000

D. SC of Human Translation vs. SC of DeepL Pro Translation

© 2025 ACADEMY PUBLICATION

The BF value for CP/C is 1.018, which is slightly above 1. This suggests weak evidence supporting the alternative hypothesis that HT is inferior in terms of the number of coordinate phrases per clause. This implies that DeepL Pro translation may have a slightly higher number of coordinate phrases per clause compared to HT. Most BF values are less than 1, suggesting more support for the null hypothesis that HT is superior to DeepL Pro translation in terms of syntactic complexity. The Rhat values for all indices are very close to 1, indicating good model convergence, which means the Bayesian Wilcoxon signed-rank test results are reliable and stable.

The quantitative analysis overwhelmingly supports the null hypothesis, demonstrating that HT generally outperforms DeepL Pro translation in terms of syntactic complexity. The strong or very strong support for the null hypothesis across most indices suggests that HT maintains a comparable or superior quality in these aspects when compared to DeepL Pro translation. This indicates that human translators are more adept at managing the syntactic intricacies of language, producing translations that are syntactically complex.

Overall, Human translators, with their ability for conceptual innovation, deep empathy, and discourse reconstruction, demonstrate irreplaceable value and uniqueness (Du, 2023). AI still has weaknesses compared to human translators in translating legislative texts, but it shows certain advantages in aspects like efficiency, the use of function words, and handling of long texts, etc. Additionally, ChatGPT 4o demonstrates superior handling of legal texts compared to DeepL pro. In terms of lexical diversity, ChatGPT 4o is capable of maintaining the original meaning while using diverse vocabulary and structures to express the same concepts. ChatGPT 4o can also generate a more readable, cohesive, and coherent translation by using a variety of function words. DeepL pro only shows little significance in lexical density of tokens. When it comes to syntactic complexity, both ChatGPT 4o and DeepL pro are superior to HT in use of coordinate phrases, but ChatGPT 4o translation's BF value is slightly higher than DeepL Pro translation's. Based on above data analysis, we can have a comprehensive understanding on the quality of AI translation compared with human translation in LD and SC.

TABLE 7
SC OF HT VS. SC OF DEEPL PRO TRANSLATION

SC Indices	BF	W	Rhat
MLS	0.126	269.000	1.000
MLT	0.154	245.000	1.001
MLC	0.913	153.000	1.003
C/S	0.091	317.000	1.007
VP/T	0.59	335.000	1.000
C/T	0.097	310.000	1.005
DC/C	0.114	277.000	1.000
DC/T	0.105	292.000	1.001
T/S	0.157	249.000	1.001
CT/T	0.104	301.000	1.011
CP/T	0.205	225.000	1.001
CP/C	1.018	148.000	1.001
CN/T	0.156	243.000	1.000
CN/C	0.844	167.000	1.000

V. IMPLICATIONS FOR LEGAL TRANSLATORS AND AUTOMATION OF LEGAL TRANSLATION

The analysis above demonstrates that AI has some strengths in translating legislative texts, but compared with human translation, AI-generated translations still have some deficiencies. However, the application of AI in legal translation is inevitable, as AI technology developing inexorably. Translators need to learn how to work with AI models. The insights from these analyses have several implications for the automation of legal translation and the role of legal translators.

First, legal translators need to fully understand the strength and weaknesses of AI translation. AI models like ChatGPT and DeepL are continuously upgrading. They are powerful tools for translators. Understanding the capabilities and limitations of these tools is crucial for effectively integrating them into the translation process. AI can provide insightful suggestions of translation from perspectives where human translators cannot think of.

Human-AI collaboration is another vital implication. In the era of artificial intelligence, translation technologies emerge and evolve continuously, with the degree of technological sophistication in the translation process constantly increasing. This has prompted the post-editing competence to become an important component of the professional quality of translators and an indispensable part of the training of translation talents (Wang & Wang, 2023). The complementary strengths of AI and human translation highlight the importance of human-machine collaboration. While AI can handle repetitive and standardized texts, human translators are essential for ensuring the translation's accuracy, style, and contextual appropriateness, particularly in complex legislative texts. The advancements in AI translation tools push translators to continuously improve their skills and stay updated with the latest technologies.

AI presents unique promise in legal sector, but ethical issues such as infringement of translators' intellectual property rights, improper handling of personal privacy information data, and exaggerated claims about translation technology products are emerging frequently (Zhang & Qu, 2021). These ethical challenges, resulting from the immature

© 2025 ACADEMY PUBLICATION

development of translation technology and the failure of technical risk management, have garnered significant attention from the academic community. Given the potential risks to data security when using AI tools, especially in legal sector where privacy, liability and confidentiality are important, and with low risk tolerance, translators must carefully consider the sensitivity and confidentiality of the information they work with. Ensuring data security and privacy should be a priority when utilizing AI for legal translation.

VI. Conclusion

Employing two computational corpus tools, the study compares AI-translated legislative texts with relative human translations from two aspects: lexical density and syntactic complexity. We find that AI is able to use coordinate structures and tends to generate more informative and concise translation. What's more, ChatGPT as a pre-trained large language model, performs better in the study than DeepL as a neural machine translation tool. Compared with DeepL, ChatGPT can generate more cohesive and coherent translation by using diverse function words.

In terms of lexical density, ChatGPT 4o demonstrates a superior capability to maintain the original meaning while employing a diverse vocabulary and complex structures. This results in translations that are not only accurate but also enriched with varied expressions, enhancing the overall richness and diversity of the content. The study indicates that AI-generated translations, particularly from ChatGPT 4o, manage to strike a balance between efficiency and depth, producing texts that are both concise and rich in information. The analysis of syntactic complexity reveals that ChatGPT 4o excels in using coordinate phrases within clauses, reflecting its ability to combine multiple ideas and details seamlessly. This syntactic flexibility contributes to the richness and depth of the text, making AI translations comparable to, and sometimes exceeding, human translations in certain syntactic aspects. The study shows that while human translations generally surpass AI in overall syntactic complexity, AI tools like ChatGPT 4o and DeepL Pro show strengths in specific areas, such as the use of coordinate phrases. Furthermore, ChatGPT 4o' s use of diverse function words significantly enhances the readability, cohesion, and coherence of the translations. The study also highlights the differences between ChatGPT 4o and DeepL Pro. While DeepL Pro shows some advantages in lexical density, it tends to produce more straightforward and dense translations, potentially sacrificing some of the richness found in human translations. In contrast, ChatGPT 4o manages to maintain a high level of lexical diversity and syntactic complexity, making it a more versatile tool for legal translation.

In conclusion, the automation of legal translation through AI tools like ChatGPT 4o and DeepL Pro presents both opportunities and challenges. By leveraging the strengths of AI and combining them with human expertise, translators can enhance their productivity and maintain high-quality translations. This study underscores the potential of AI tools to complement human translation efforts, suggesting a future where AI and human translators work in tandem to adapt to the evolving landscape of the translation industry.

References

[1] Bergmann, R., Ludbrook, J., & Spooren, W. P. (2000). Different outcomes of the Wilcoxon—Mann—Whitney test from different statistics packages. *The American Statistician*, *54*(1), 72-77.
[2] Choudhury, R., & McConnell, B. (2013). *Translation technology landscape report*. TAUS BV, DeRijp.
[3] Cooper, T. C. (1976). Measuring written syntactic patterns of second language learners of German. *The Journal of Educational Research*, *69*(5), 176-183.
[4] Du, A. (2023). Exploring the value spaces of human translation in the ChatGPT era and the transitions needed for translation education. *Foreign Languages and Cultures*, *7*(4), 90-103. Doi: 10.19967/j.cnki.flc.2023.04.009.
[5] Herbold, S., Hautli-Janisz, A., Heuer, U., Kikteva, Z., & Trautsch, A. (2023). A large-scale comparison of human-written versus ChatGPT-generated essays. *Scientific Reports*, *13*(1), 18617.
[6] JASP Team. (2024). *JASP* (version 0.18.3) [Computer software]. https://jasp-stats.org/download/
[7] Jiao, W., Wang, W., Huang, J. T., Wang, X., Shi, S., & Tu, Z. (2023). *Is ChatGPT a good translator? Yes with GPT-4 as the engine.* arXiv. org. https://arxiv.org/abs/2301.08745
[8] Johnson, W. (1944). Studies in language behavior: A program of research. *Psychological Monographs*, *56*(2), 1-15.
[9] Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773–795. https://doi.org/10.1080/01621459.1995.10476572
[10] Kyle, K., Crossley, S. A., & Jarvis, S. (2021). Assessing the validity of lexical diversity using direct judgements. *Language Assessment Quarterly*, *18*(2), 154-170. https://doi.org/10.1080/15434303.2020.1844205
[11] Lin, X., M. Afzaal, & Aldayel, H. S. (2023). Syntactic complexity in legal translated texts and the use of plain English: a corpus-based study. *Humanities and Social Sciences Communications*, *10*(1), 1-9.
[12] Lu, X. (2010). "Automatic analysis of syntactic complexity in second language writing." *International Journal of Corpus Linguistics*, *15*(4), 474-496. https://doi.org/10.1075/ijcl.15.4.02lu
[13] Masson, M. E. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*, 679-690.
[14] McCarthy, P. M., & Jarvis, S. (2010). MTLD, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, *42*(2), 381-392.
[15] Moneus, A. M., & Sahari, Y. (2024). Artificial intelligence and human translation: A contrastive study based on legal texts. *Heliyon*, *10*(6), 1-14.

© 2025 ACADEMY PUBLICATION

[16] Nunes Vieira, L., & Alonso, E. (2020). Translating perceptions and managing expectations: an analysis of management and production perspectives on machine translation. *Perspectives*, *28*(2), 163–184. https://doi.org/10.1080/0907676X.2019.1646776
[17] Prieto Ramos, F. (2024). Revisiting translator competence in the age of artificial intelligence: the case of legal and institutional translation. *The Interpreter and Translator Trainer*, *18*(2), 148-173.
[18] Scott, J. R. (2019). *Legal Translation Outsourced*. London: Oxford Studies in Language and Law.
[19] Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. *Psychological Methods*, *26*(1), 103-126.
[20] Tan, L. (2024). *NeoSCA* (version 0.1.4) [Computer software]. GitHub. https://github.com/tanloong/neosca
[21] Wang, H., & Ma, K. (2023). The application of artificial intelligence in biomedical text translation: A comparative study. *Chinese Science & Technology Translators Journal*, *36*(3), 23-26. https://doi:10.16024/j.cnki.issn1002-0489.2023.03.001.
[22] Wang, L., & Wang, X. (2023). The study on the training model of machine translation post-editing competence in the era of ChatGPT. *Technology Enhanced Foreign Language Education*, *(4)*, 16-23+115.
[23] Wang, Z., & Mao, C. (2023). Quality evaluation and improvement of ChatGPT translation——A case study on ceramics texts. *Shandong Ceramics*, *46*(4), 20-27.
[24] Wen, X., & Tian, Y. (2024). The effectiveness of ChatGPT in translating China-specific discourse text. *Shanghai Journal of Translation, (2)*, 27-34+94-95.
[25] Yu, L. (2024). Lexical diversity and syntactic complexity in ChatGPT translation. *Foreign Language Teaching and Research*, *56*(2), 297-307+321. Doi: 10.19923/j.cnki.fltr.2024.02.005.
[26] Yuan, Y. (2023). Beyond Chatbots and towards artificial general intelligence (AGI): the success of ChatGPT and its implications for linguistics. *Contemporary Linguistics*, *25*(5), 633-652.
[27] Zhang, F., & Qu, X. (2021). Ethical review of legal translation technology. *Foreign Languages in China*, *18*(6), 17-22.

Jialei Chen is a postgraduate student majoring in Foreign Linguistics and Applied Linguistics from School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China. His research interests focus on Legal Translation and Corpus Linguistics. He has a strong passion for Chinese-English legal translation and corpus-based studies. He has attended three research projects and published two research articles in corpus-based legal translation studies.

Qiufeng Hong is a postgraduate student majoring in Foreign Linguistics and Applied Linguistics from School of Foreign Languages, Southwest University of Political Science and Law, Chongqing, China. His research interests focus on Legal Translation and Machine Translation.

© 2025 ACADEMY PUBLICATION