Studies in Linguistics and Literature
ISSN 2573-6434 (Print) ISSN 2573-6426 (Online)
Vol. 9, No. 4, 2025
www.scholink.org/ojs/index.php/sll

Original Paper

A Comparative Dependency Analysis of Human Translation and Machine Translation: A Case Study of English translation of *To Live*

Nuo Ding[1] & Jingxiang Cao[1*]

[1] Dalian University of Technology, Dalian, China

[*] Corresponding author

Received: October 30, 2025 Accepted: December 1, 2025 Online Published: December 26, 2025

doi:10.22158/sll.v9n4p130 URL: http://dx.doi.org/10.22158/sll.v9n4p130

Abstract

Recent advances in neural machine translation have significantly improved translation quality, yet its ability to handle syntactic complexity in literary texts remains underexplored. This study examines syntactic differences between human and machine translations of a Chinese literary text from the perspective of mean dependency distance. Drawing on one human translation and four machine-generated translations, the analysis compares dependency distance patterns and investigates how sentence length relates to differences across translations. The findings indicate that although both human and machine translations show a general tendency toward syntactic simplification, notable divergences persist between human and machine output. These divergences are unevenly distributed and are closely associated with sentence length, especially in longer sentences. The study suggests that sentence-level restructuring constitutes a key distinction between human and machine literary translation and remains a challenge for current machine translation systems.

Keywords

mean dependency distance, sentence length, machine translation, literary translation

1. Introduction

1.1 Research Background

The field of Machine Translation (MT) has evolved rapidly from statistical paradigms to Neural Machine Translation (NMT) and, most recently, Large Language Models (LLMs). While these technologies have achieved remarkable semantic accuracy in technical domains, systematic structural differences—often termed "translationese"—persist between human and machine-generated texts

(Volansky et al., 2015). The differences are particularly critical in literary translation, where the reproduction of stylistic nuance requires sophisticated syntactic manipulation beyond surface-level correctness.

To assess the structural features objectively, scholars have turned to Dependency Grammar (DG). DG is a framework for syntactic parsing that examines sentence architecture through binary, asymmetric relations between a governing "head" and a subordinate "dependent" (Tesnière, 1959). Unlike phrase-structure grammars, which focus on constituent hierarchies, dependency analysis prioritizes the functional connections between lexical units, offering quantifiable metrics for cross-linguistic comparison (Hudson, 2007).

A central metric in this framework is Mean Dependency Distance (MDD), defined as the average linear distance between heads and dependents in a sentence. Research in quantitative linguistics has established that MDD is closely correlated with cognitive load; specifically, human languages tend to evolve syntactic structures that minimize dependency distance to respect the constraints of working memory (Liu, 2009). Consequently, applying dependency analysis to translation studies may represent a methodological innovation. It may bridge the gap between syntactic theory and translation practice, allowing researchers to determine if MT outputs adhere to the cognitive constraints typical of human language processing.

1.2 Research Purpose and Questions

Despite the utility of dependency grammar, current research on human-machine comparison primarily focuses on lexical errors or fluency, often neglecting deep syntactic structures. Few studies have applied it to distinguishing human literary translations from NMT translations or LLMs-based translations.

To fill this gap, this study investigates the linguistic differences between human and machine translations of the novel *To Live* by Yu Hua. By employing a quantitative dependency analysis framework, this research addresses two specific questions:

1) Are there differences in the Sentence Length (SL) between machine translations and the human translation?

2) Are there differences in the Mean Dependency Distance (MDD) between machine translations and the human translation?

1.3 Research Significance

This study intends to make three contributions. First, it establishes a replicable framework for quantifying "syntactic naturalness" in literary translation. Second, it uses MDD as a diagnostic feature to reveal how different MT paradigms handle syntactic complexity, particularly in long sentences where machine systems often struggle to maintain coherence. Finally, the identification of systematic dependency divergences provides actionable data for optimizing MT models, supporting the development of translation tools that equals syntactic authenticity alongside semantic accuracy.

2. Literature Review

2.1 Dependency Grammar

Dependency Grammar (DG) is a syntactic theory emphasizing unequal relations between words in a sentence. The core postulate of DG is that sentence structure can be analyzed through pairwise combinations of words. Moreover, DG supposes that the combination of words is asymmetric: one word is the governor and the other is the dependent (Hudson, 2007). Dependency relation, also called dependency type refers to the relation between the governor (head) and the dependent. It focuses on the connections between words in a sentence rather than the traditional hierarchical structure of constituents.

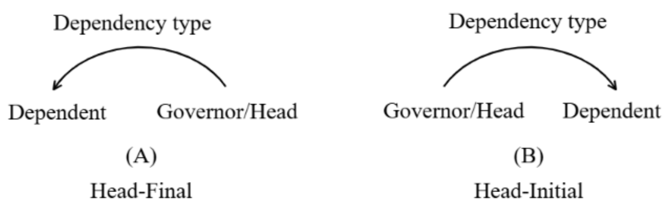

Figure 1. Three Elements of a Dependency

Figure 1 shows the three features of dependency relation. The directed arc from the head to the dependent indicates the asymmetrical relation between the two elements. The label is called dependency type. Dependency type can vary according to the dependency relation. If the head's position is after the dependent, as shown in (A). It is called head-final dependencies (HF). If the head's position is before the dependent, as shown in (B). It is called head-initial dependencies (HI). The position of head in dependency relation denotes dependency direction.

2.1.1 Dependency Distance

Dependency Distance (DD), an indicator describing the linear distance between the governor and the dependent, can be measured by counting the intervening words (Hudson, 1995). All the words in a sentence are numbered in order and starting with 1. The closer the word is to the end of the sentence, the larger its number is.

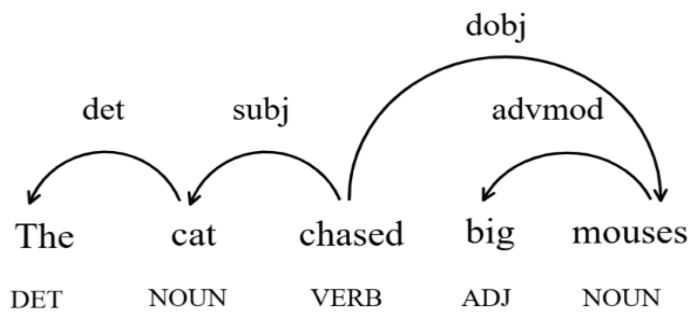

Figure 2. Dependency Analysis of "The Cat Chased Big Mouses"

To be specific, for any dependency relation between the words W_a and W_b, if W_a is a governor and W_b is its dependent, the DD between the two adjacent is a-b (Liu, 2008). With this calculation method, we can calculate DD. if a is greater than b, the DD is a positive number, which means that the governor follows the dependent; if a is smaller than b, the DD is a negative number and the governor precedes the dependent. For example, in Fig. 2.2, in the verbal phrase of "chased big mouses", "mouses" is the governor of "big". The dependent "big" is numbered as 4, and the governor "mouses" is numbered as 5. The DD between the two words is 5-4 equals 1, which is positive number. In the same phrase, "chased" is the governor of "mouses", so the DD between the two words is 3-5 equals -2, which is a negative number.

2.1.2 Mean Dependency Distance

Mean Dependency Distance (MDD) is a widely adopted metric for quantifying syntactic complexity due to its logical consistency and computational efficiency (Liu, 2008). MDD is defined as the average of the absolute values of dependency distances in a sentence or a text sample. For a sentence, MDD is calculated by summing all individual dependency distances and dividing the result by the total number of dependency relations, as shown in Formula 2.1. For a text, MDD is generalized to include the total number of words and sentences, as represented in Formula 2.2 In the two formulas, *n* is the total number of words in the sentence or texts: the total number of sentences in the sample *DDi:* the dependency distance of the i-th syntactic link of the sample (Liu, 2008).

$$MDD\ (the\ sentence) = \frac{1}{n-1}\sum_{i=1}^{n}|DDi|$$

Formula 2.1

$$MDD\ (the\ text) = \frac{1}{n-s}\sum_{i=1}^{n}|DDi|$$

Formula 2.2

MDD provides a more precise measure of syntactic complexity compared to raw dependency distances because it accounts for averages across sentences or samples. It is also closely tied to human working memory limitations, as longer dependency distances require more cognitive effort to process and store information, thereby increasing sentence complexity. Consequently, MDD is an insightful metric for understanding linguistic structures and processing difficulties.

2.2 Literary Translation

Literary translation requires more than semantic transfer; it must recreate tone, imagery and authorial voice while conforming to target-language norms. Translators exercise creative agency in rendering narrative voice and aesthetics, and individual translator style appears in systematic linguistic choices. Corpus studies show that coarse measures such as average sentence length and simple lexical repetition often capture general translationese rather than distinctive translator voice (Fan et al., 2019). By contrast, fine-grained syntactic metrics derived from dependency grammar reveal translator signatures. Tang (2025) found that different translators of Hongloumeng produce distinct profiles of mean dependency distance and dependency direction, which serve as sensitive stylistic fingerprints.

Typological contrasts between Chinese and English intensify the syntactic challenges of Chinese-to-English literary translation. Chinese commonly uses parataxis and ellipsis; English generally prefers hypotaxis, fixed constituent order and explicit subordination. Translators therefore routinely segment clauses, use omitted pronouns, insert subordinators and reorder modifiers to produce idiomatic English. For example, translators of early modern Chinese fiction produce substantially longer average sentences in English than comparable native English texts, reflecting structural unpacking of implicit Chinese clauses (Lu et al., 2023).

Dependency grammar offers operational measures to quantify the adjustments. Dependency distance measures the linear span between a head and its dependent, and the Dependency Distance Minimization hypothesis predicts a tendency toward short dependencies because of human working-memory constraints. Cross-linguistic observation shows that adjacent-word dependencies are the most common pattern (Liu, 2008). In translation practice, this cognitive constraint often leads translators to shorten long source dependencies. Corpus studies report that translated English typically occupies an intermediate position in mean dependency distance: longer than native English but shorter than the Chinese source, a compromise driven by source-language influence and target-language normalization (Ferrer-i-Cancho, 2004). Lu et al. (2023) show that when translators face long, parallel Chinese sentences they commonly break or reorder clauses so that most dependencies remain short, thereby reducing processing load while preserving content.

2.3 Comparison between Human and Machine Translation

Comparative research on human versus machine translation of literary texts has moved from surface adequacy toward structural evaluation. Human translations generally show greater structural diversity and a stronger target-language orientation. Neural machine translation tends to be more literal and uniform, exhibiting source-language interference in word order and clause structure. Sizov and colleagues show that human translators often choose creative syntactic alternatives while NMT systems produce more conservative, word-by-word renderings (Sizov et al., 2024). Large language models narrow the gap: some LLM outputs resemble human translations in fluency and local syntactic arrangement, but measurable differences from native writing remain (He et al., 2024).

Dependency-based analyses provide quantitative insight into these contrasts. Zhang (2025a) compared human, Google Translate and ChatGPT translations of Chinese academic text using mean dependency distance and related metrics and found that ChatGPT and human translations share lower MDD than Google Translate, indicating that ChatGPT, like skilled humans, tends to shorten dependencies to ease processing. Zhang (2025b) further reports that ChatGPT produces fewer head-final dependencies on medium-length sentences than Google, which suggests more English-like ordering and reduced source interference. These findings imply that LLMs can approximate human strategies for reducing syntactic complexity at the sentence level.

Broader comparisons yield a nuanced pattern. Jiang (2025) show that both human and LLM translations simplify syntax relative to source Chinese, with human translators often simplifying more aggressively than models; this difference implies that human translators actively optimize structure to a greater degree than current systems. Thus, simplification emerges in both human and machine outputs but varies in extent and in how it interacts with stylistic goals.

Sentence-length behavior also differs. Human translators display individual preferences that affect sentence-length distributions; MT systems lack such idiosyncratic variation and therefore produce more homogeneous length profiles. This uniformity contributes to a narrower cluster of stylistic features in MT output compared with broader dispersion across human translations (Sizov et al., 2024; He et al., 2024). In literary translation this matters because sentence rhythm, clause embedding and long-distance dependencies contribute to narrative voice and rhetorical effect.

Taken together, dependency grammar reveals that human and machine translations diverge in measurable ways. Human translators typically yield shorter MDD, while NMT exhibits source interference. Yet LLMs have reduced some syntactic gaps by generating more target-like dependency profiles (Zhang, 2025a; Jiang, 2025). For literary texts, where long sentences and embedding are frequent, combining dependency metrics with close qualitative analysis remains essential to evaluate whether machine output captures the syntactic and stylistic richness of human translation.

3. Methodology

3.1 Materials

It is generally acknowledged that machine translation systems tend to perform better on informational texts compared to literary ones. The original Chinese novel *To live* was written by Yu Hua, one of the most renowned contemporary Chinese authors, whose works have been translated into over 40 languages. The novel was first published in 1993 and has since gained significant recognition. In 1998, it won the prestigious Grinzane Cavour Prize in Italy and the Prix Courrier International in France. The

English translation of the novel, titled *To Live*, was published in 2003 and was translated by Michael Berry, a prominent translator and UCLA professor specializing in modern Chinese culture. The translation was selected for the National Endowment for the Arts (NEA) Big Read program, highlighting its importance and impact. In China, the novel was named one of the “Ten Most Influential Books of the Decade”, and it is widely adopted in university and high school curriculums, praised as one of modern literary classics.

The Chinese and English translation are collected from the version of *To live* published in 1993. Following the removal of the title page, table of contents, and acknowledgements section, the remaining textual body constitutes the material collected for analysis. We use Google and DeepL to translate the original Chinese text respectively and directly downloaded them from websites. Compared with DeepL and Google translate which are translation tools, DeepSeek and Chat-GPT are both LLMs-generated tools. Prompts determine the translation quality. A prompt with specific roles, aims and cautions to AI tools may help them generate better or more specific translations. To avoid the influence of different prompts, we use Yicat online platform to obtain the translation. Yicat is an online translating platform, specialised in Computer Assisted Translation (CAT) and Machine Translation (MT). The Yicat platform is integrated with DeepSeek API, which allows direct translation. Then, the four machine translations are downloaded for further research.

The statistics of token, type and sentences of each translations are listed in Tab. 1. In the following text, we use CGT (Chat-GPT Translation), DLT (DeepL Translation), DST (DeepSeek Translation), GT (Google Translation) and HT (Human translation) for ease of expression.

Table 1. Research Materials

Versions	Token	Type
Original Chinese	49013	6354
Chat-GPT	61328	4136
DeepL	64219	3780
DeepSeek	60573	3988
Google	62386	3400
Human	71401	4156
Total	368920	25814

3.2 Instruments

To cover the current MT technology paradigms, this study selected two traditional NMT software, namely DeepL and Google Translate, as representatives of NMT, which are the industry benchmarks of the technical routes in Europe and North America respectively. DeepL is a machine translation service that has gained significant recognition for its high-quality translations, especially in European languages. DeepL stands out with its ability to maintain the context and nuances of the original text, making it particularly useful for professional and formal documents. Google Translate is one of the most widely used online translation services, available in over 100 languages. Google Translate is known for its versatility, supporting not only text but also speech, images, and even real-time conversation mode.

In addition, we have selected two LLMs tools, namely ChatGPT3.5 and DeepSeek-R1, as representatives of the translation paradigms of LLMs. Chat-GPT, the former, has the strongest zero-shot cross-language transfer ability, while DeepSeek, the latter, specializes in the deep representation of Chinese semantics and is suitable for the English translation of Chinese novels. ChatGPT-3.5, developed by OpenAI, is a transformer-based LLM fine-tuned for conversational tasks. Leveraging 175 billion parameters and trained on diverse multilingual corpora, it demonstrates robust natural language processing capabilities, including cross-lingual translation. DeepSeek-R1, developed by the Chinese AI company DeepSeek, is a transformer-based LLMs tool optimized for reasoning and multilingual tasks. It delivers competitive natural language processing performance, particularly in MT. Notably, frequent version updates in LLMs lead to significant performance discrepancies—particularly notable between ChatGPT-3.5 and ChatGPT-4. Our selection of GPT-3.5 is justified by its role as OpenAI's first massively deployed conversational model, which established the technical paradigm for LLMs-driven translation. Compared with multimodal elements in ChatGPT-4, it allows focused analysis of syntactic generation mechanisms. Meanwhile, DeepSeek-R1 was chosen for its demonstrated superiority in Chinese syntactic parsing, ensuring robust native dependency modeling of dialectal expressions and culture-bound terms. This combination can not only reflect the influence of the evolution from NMT to LLMs based translation on the dependency structure, but also be applicable to literary works.

PyCharm is used to conduct comparative dependency analysis using spaCy for English translations and jieba for original Chinese text. PyCharm is a Python IDE developed by JetBrains, offering advanced code editing, debugging, version control integration. SpaCy is a Python library for natural language processing that provides efficient tools for tasks like tokenization, dependency parsing, and text classification. Jieba is a Python library for Chinese text processing, offering precise and flexible tokenization with multiple modes, support for custom dictionaries and keyword extraction.

Tmxmall is used for alignmnet of translations. It is a cloud-based translation data platform specializing in AI-powered corpus processing, with its Parallel Corpus Alignment Engine as a flagship feature.

3.3 Research Procedures

This study employs both qualitative and quantitative methods. For the quantitative methods. We establish each translation as one corpus, use notepad to clean the corpora, Tmxmall to alignment each corpus and Pycharm based on python scripts to calculate the MDD, DD and Dependency type of each corpus.

Firstly, we clean the materials collected. To address heterogeneous formatting errors. Notepad is used to clean the corpora. We apply regular expressions to clean these errors first and then manually check each translation to ensure better dependency analysis. SpaCy is used to obtain the basic information of English translations, including type, token and number of sentences. For the original Chinese text, jieba is used. Secondly, we use Tmxmall to realize the translation alignment.We use the automatic alignment function of Tmxmall to realize the basic alignment, while there are still some sentence alignment errors. Thus, we manually aligned each sentence, and finally constructed 2088 sentences of five translations and the Chinese original text alignment. Thirdly, we run the python script through Pycharm to get the dependency feature table, which was used to calculate the MDD of each dependency pair for each translation. Since MDD is affected by the sentence length, we again calculated the sentence length and MDD of each sentence.

4. Results and Discussion

4.1 Mean Dependency Distance

We first calculate and compare the MDD of six versions, and the results are shown in Tab. 2.

Table 2. MDD of the Original Chinese Text and Five Translations

Versions	MDD
Original Chinese	3.54
ChatGPT	2.50
DeepL	2.89
DeepSeek	2.48
Google	2.44
Human	2.41

As mentioned before, a longer MDD means that the sentence structure is more complex. The longer dependency distance caused by complex sentence structure puts more pressure on human beings' working memory, so sentences are more difficult to understand as well. Thus, MDD is closely related to the capacity of human cognition, particularly to that of working memory. Cowan (2005) assumes that the MDD within human languages should remain below a threshold, which we expect to be smaller than 4. Liu (2008) found that the MDD of natural languages is shorter than that of random languages, thus proposing that "minimization of dependency distance" is a universal phenomenon in human languages. Notably, all six versions exhibit MDD values below Cowan's cognitive threshold of 4, validating the MDD principle in literary works. This convergence suggests that in literary works, machine translations inherently optimize syntactic structures to align with human working memory.

Then we compare the MDD of five translations and original Chinese text. As evidenced in Table 2, the original Chinese text exhibits the MDD value of 3.54, while all five English translations demonstrate smaller MDD values ranging from 2.41 to 2.89. This result appears consistent with Liu's assertion that Chinese typically exhibits greater MDD than other languages. Liu (2008) also calculates the naturally and randomly structured samples of English and Chinese, where their MDD are 2.543 and 3.662, respectively. MDD of the six versions are all smaller than his assertion, it may proves the trend of minimization of dependency distance in human languages, more specifically, in literary works.

We also compare differences between five translations. Fan and Jiang (2024) proposed a hypothesis that the MDDs of translations are basically between MDD of target language and source language. In this study, we take the human translation as the target languages and original Chinese as the source language. The result is divergent from their research because the MDD values of four MT are greater than the HT. Thus, we remain neutral attitudes towards the hypothesis. As shown in table.4.1, the differences of MDD among MT are obvious. The MDDs of English MT are ranging from 2.40 to 2.50, except DLT, which is greater than others. This divergence may suggest NMT systems prioritize source-language syntax over target-language optimization in literary contexts.

To sum up, six versions show differences in MDDs. First, the six versions including original Chinese exhibit smaller MDDs than the threshold of 4, validating dependency distance minimization as a cognitive universal in literary works. However, the MDDs of English translations are smaller than that of the original Chinese, suggesting the difference between English and Chinese. Secondly, MDDs of MTs demonstrate differences from 2.40 to 2.50, with DLT as an outlier. Crucially, four MTs display higher MDDs than HT, contradicting Fan & Jiang's hypothesis that translations exhibit intermediate MDD between source and target languages.

4.2 Mean Dependency Distance with Sentence Length

Dependency distance is also influenced by other non-cognitive factors. Longer sentences are presumed to have longer dependency distance, and also greater MDD, since it is the Sentence Length (SL) that determines MDD. A long sentence is the precondition of a long dependency distance. Thus, we investigate the relation between SL and MDD. First, we calculate the distribution of sentence length, which is listed in Figure 3.

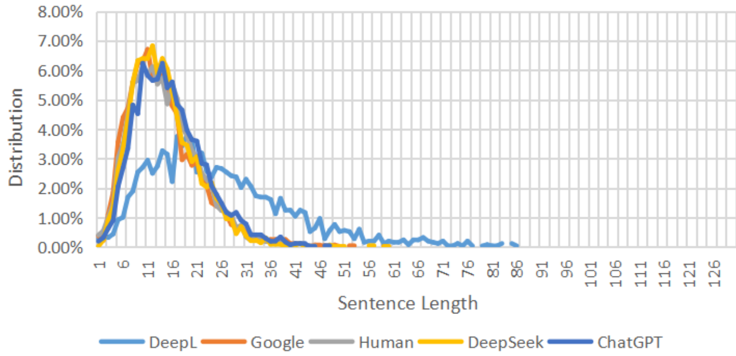

Figure 3. Distribution of Sentence Length of Five Translations

It illustrates the distribution of SL across five translations. All distributions exhibit a right-skewed pattern, wherein shorter sentences dominate, and frequency gradually declines as SL increases. The peaks of the distributions—representing the most frequent SLs—are concentrated between 10 and 20 words. CGT and GT demonstrate the sharpest peaks, approaching 7%. In contrast, DLT shows a flatter and more dispersed distribution. It extends further to the right, with small proportions of SLs exceeding 30, 50, or even 80 words. It can be seen that there is an extremely similar distribution trend between GT, DST, CGT and HT, while DLT shows a ratherly different trend.

HTs exhibits a near-normal distribution centered around a mode of 10 words, with tight dispersion and minimal right-skewness. Only 4.5% of sentences exceed 30 words. In contrast, MTs show significant distributional divergence from HT. DLT displays extreme right skew, whose medians are 18 and 19 and mode of 19 with heavy right-tailed distribution. The SL of DLT over 30 accounts for 26%, indicating systematic long-sentence generation. GT exhibits left skew, whose medians are 8 and 9 and mode of 10, over-indexing on short sentences, indicating short-sentence generation. DeepSeek and ChatGPT approximate human central tendency, yet show higher variance in mid-length sentences and moderately heavy tails. The SLs over accounts for 10-15%, which is twice or three times that of the HTs.

In a short conclusion, there is an extremely similar distribution trend between GT, DST, CGT and HT, while DLT shows a relatively different trend. While all systems are right-skewed, DLT's extreme long sentences and GT's short sentences generation reveal differences with HT. DST and CGT are basically consistent with HT in terms of SL distribution but fail to achieve the consistency of mid-length sentences.

As we mentioned earlier, MDD is a measure influenced by SL. To reveal the MDD difference in each sentence, we calculate MDD of each sentence and compare them with the corresponding SL. MDD and SL can are listed in Figure 4.

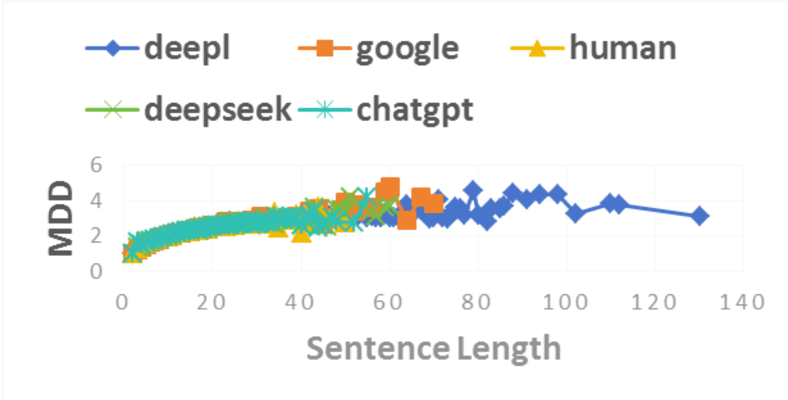

Figure 4. MDD of Different SL of Five Translations

The relationship between sentence length and dependency distance is assumed to be linear, i.e., y = a + bx, where x stands for SL, y stands for the corresponding MDD. After calculation, we find that when the SL is shorter than 45, the MDD and SL of each version show a linear relationship, which conforms to the power function model. (y=a· xb, R2>0.98), and after the medium and long sentences, that is, 45, we find that all four versions gradually deviate from the power function model. We can suggest that a long sentence may have many short dependency pairs with short MDD, but a long MDD calls for a long SL. In this study, when the SL ranges from 2 to 30, the MDD and SL is in a linear relationship. When the SL reaches 45 or even longer, the MDD goes up in a slow trend, which is deviation of linear relationship.

DLT presents the longer MDD, with MDD value exceeding even 4.5 for sentences over 100 words. This suggests that DeepL may attempt to preserve complex syntactic constructions in long sentences. GT's MDD value are generally smaller, rarely exceeding 3.5 even for longer sentences. Even if Google control the SL, there are still some long sentences with MDDs higher than 4. DST exhibits a mid-range MDD, increasing steadily with SL but without the dramatic peaks seen in DLT or the flattening observed in GT. HT demonstrates a smooth trend, with a consistent MDD rise that reflects nuanced control over sentence structure. Even in long sentences, MT's MDD remains moderate and controlled.

When the SL is less than 30, the MDD of the five versions exhibit a relatively high level of convergence. This indicates a notable degree of correspondence in handling short sentence. As the sentence length falls within the interval of 31-70, the MDD values of these five models display significant fluctuations and divergences. This phenomenon suggests the presence of substantial disparities in how each machine translation processes sentences of this length, with their translation quality and consistency varying. For sentences with a length exceeding 70, only DLT generates translations or has measurable MDD for sentence lengths beyond 70. This reflects that DLT may follows the SL of the original Chinese version. SL significantly impacts the consistency of machine translation model performance stability for short sentences, high variability for medium-length sentences, and DLT demonstrates a unique advantage in handling long sentences.

The results demonstrate consistent and meaningful differences between human and machine translations in terms of dependency distance and sentence-length behavior. All English translations show a reduction in mean dependency distance compared with the Chinese source text, indicating a general tendency toward syntactic simplification in translation. However, the human translation exhibits the lowest dependency distance, while machine translations remain syntactically heavier, suggesting that current MT systems do not fully replicate human strategies of syntactic reorganization in literary translation. This finding directly addresses the first research question and provides limited support for the “intermediate” hypothesis, as machine output does not consistently occupy a middle position between source and human translation.

Regarding the second research question, sentence length emerges as a key factor shaping these differences. Machine translations diverge from the human version in their handling of medium and long sentences, either by excessive segmentation or by preserving long, complex structures from the source text. The strong association between increasing sentence length and rising dependency distance further indicates that syntactic complexity in MT is driven less by local grammatical choices than by global sentence planning. Overall, these findings suggest that while machine translation performs comparably to human translation at the clause level, it remains limited in managing long-distance dependencies typical of literary prose, pointing to sentence-level restructuring as a critical area for future improvement.

References

Cowan, N. (2005). *Working memory capacity*. Hove: Psychology Press.

Fan, J., Liu, H., & Sun, Y. (2019). Distinguishing translational language from native language using dependency distance and direction. *Journal of Quantitative Linguistics*, *26*(3), 256-275.

Fan, L., & Jiang, Y. (2024). “In-between Hypothesis”—A novel proposition on the feature of translational language: A quantitative study based on dependency grammar. *Foreign Language Teaching and Research*, *56*(1), 136-147, 161. (in Chinese)

Ferrer-i-Cancho, R. (2004). Euclidean distance between syntactically linked words. *Physical Review E*, *70*(5), 056135.

Futrell, R., Mahowald, K., & Gibson, E. (2015). Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, *112*(33), 10336-10341.

He, Z. W., Liang, T., Jiao, W. X., et al. (2024). Exploring human-like translation strategy with large language models. *Transactions of the Association for Computational Linguistics*, *12*, 229-246.

Hudson, R. (1995). *Measuring syntactic difficulty* (Research report). London: Department of Phonetics and Linguistics, University College London.

Hudson, R. (2007). *Language networks: The new word grammar.* Oxford: Oxford University Press.

Jiang, Z. (2025). Does LLM translation align with translation universals? A cross-genre simplification study on English-Chinese translation based on dependency grammar. *PLOS One*, *20*(6), e0324830.

Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159-191.

Liu, H. (2009). *Dependency grammar: From theory to practice*. Beijing: Science Press.

Lu, F., et al. (2023). Syntactic dependency strategies in translation of long Chinese sentences. *Digital Scholarship in the Humanities*, *38*(2), 123-139.

Sizov, F., España-Bonet, C., van Genabith, J., Xie, R., & Dutta Chowdhury, K. (2024). Analysing translation artifacts: A comparative study of LLMs, NMTs, and human translations. *Computational Linguistics*, *50*(1), 45-72.

Tang, H. (2025). Can syntactic complexity distinguish translator style? A dependency-based stylometric study. *SAGE Open*, *15*(3), 1-18.

Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris: Klincksieck.

Volansky, V., Ordan, N., & Wintner, S. (2015). On the features of translationese. *Digital Scholarship in the Humanities*, *30*(1), 98-118.

Zhang, Y. (2025a). A dependency-based comparison of human and machine translation. *Modern Linguistics*, *13*(7), 491-498.

Zhang, Y. (2025b). ChatGPT and human translation: Dependency distance comparison. *Language Technologies*, *9*(1), 77-90.