International Journal of Linguistics, Literature and Translation
ISSN: 2617-0299 (Online); ISSN: 2708-0099 (Print)
DOI: 10.32996/ijllt
Journal Homepage: www.al-kindipublisher.com/index.php/ijllt

IJLLT
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Linguistic Features of Native and Translated Chinese Cancer Information: A Corpus-Based Study

Yi Xiao[1] and Mengdan Zhao[2]
[1]*Chengdu University of Technology, College of Foreign Studies, Chengdu and China*
[2]*Chengdu University of Technology, College of Foreign Studies, Chengdu and China*
Corresponding Author: Mengdan Zhao, **E-mail**: mengdan_zhao@163.com

| ABSTRACT
This study explores the linguistic features of native Chinese and translated Chinese health information related to cancer employing a corpus-based approach. The analysis focuses on three main areas: lexical density, syntactic structures, and discourse features. By comparing texts from authoritative Chinese and international health websites, this research aims to uncover the impact of translation on the complexity and readability of cancer information. The findings reveal significant differences in lexical density, the use of passive constructions, and the frequency of connectives and idioms, providing insights into the translation process and its implications for health communication.

| KEYWORDS
corpus-based study, linguistic features, translation universals, translational Chinese

| ARTICLE INFORMATION
ACCEPTED: 14 October 2026 **PUBLISHED:** 15 November 2026 **DOI:** 10.32996/ijllt.2026.9.11.26

1. Introduction

In the contemporary era of digitalization and globalization, access to accurate and comprehensible health information has become increasingly crucial for public health promotion and disease prevention. Health communication, as an interdisciplinary field bridging medicine, public health, and communication studies, plays a vital role in disseminating knowledge, shaping health behaviors, and supporting clinical decision-making (Kreps & Bonaguro, 2009). With the rising incidence of cancer worldwide, the demand for reliable and accessible cancer-related information has grown significantly, making effective health communication a public health priority.

In this context, translation serves as a key conduit for cross-cultural knowledge transfer, enabling the flow of medical and health information across linguistic boundaries. The quality of translated health materials directly influences their accessibility, acceptability, and efficacy among target audiences (Pöchhacker, 2016). Therefore, understanding the linguistic characteristics of translated health texts is not only a concern of translation studies but also a practical necessity for effective health communication.

As technology develops, a growing number of researchers use computer assistance in linguistics. Applying computerized corpora can help explain the phenomenon of translation, which means it can investigate the features of translated texts. These features are the outcomes of translation as a specific practice. Translation is not just a direct conversion of languages, but a complex process involving multiple cognitive and social factors (Baker, 2020b). McEnery and Xiao (2007) pointed out that the features of translation are different from the features of the native target language. Translated texts tend to be more conservative in vocabulary usage, and the coherence of the text may be affected by the structure of the original text. However, few studies have examined how these features manifest in specialized Chinese health communication, particularly cancer information. This study therefore compares native Chinese and translated Chinese cancer health texts through a self-built corpus.

Copyright: © 2026 the Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) 4.0 license (https://creativecommons.org/licenses/by/4.0/). Published by Al-Kindi Centre for Research and Development, London, United Kingdom.

2. Literature Review

Translation universals are defined as the typical linguistic features that appear in translated texts, which are not present in native texts and are not the result of interference from a specific linguistic system (Baker, 2020b). Baker (2020a) categorizes translation universals into three main dimensions: simplification, explicitation, and normalization.

Most studies on the translation universals hypothesis have been conducted based on English corpora. However, some studies have investigated the linguistic features of native Chinese texts and translated Chinese texts. Xiao (2010) explored the linguistic features of translated Chinese texts through the comparative analysis of two Chinese corpora and found that the lexical density of translated Chinese texts is slightly lower than that of native Chinese texts, and translated Chinese texts tend to use more connectives. Furthermore, Xiao revealed that the English language influences passive constructions in translated Chinese texts; therefore, the hypothesis of normalization does not fully apply to Chinese texts. Xiao et al. (2010) further developed the "third code" perspective by using the ZJU Corpus of Translational Chinese, which provides a useful framework for comparing translated and native Chinese.

Laviosa (1998) pointed out that lexical density, the ratio of content words to total words, is lower in English translations. Chen (2006) investigated the frequency of connectives in translated Chinese scientific texts and argued that connectives are used more frequently in translated Chinese than in native Chinese, treating connectives as one of the explicitation features of English–Chinese scientific and technological translation. These studies suggest that translation universals are not uniform across genres and registers. Pang (2024), for example, showed that the manifestation of translation universals varies across genres, while Wu et al. (2023) found that translated Chinese texts in news and general genres can exhibit greater syntactic complexity on certain measures than native texts. This study therefore examines whether similar patterns appear in cancer health information, a specialized domain characterized by terminological precision and reader accessibility.

Readability is a central concern in health communication. Unlike other categories of materials, health education materials should avoid unnecessary passive voice to improve comprehensibility (Shoemaker et al., 2014). Baur and Prue (2014) similarly emphasized plain-language principles in the CDC Clear Communication Index. Lau (2006) argued that because the semantics of idioms can be more complex than their surface meaning, including idioms in a text may affect readability. If readers cannot understand the meaning behind the idioms, the difficulty of the text will increase. Chapple et al. (2021) further showed that translation in clinical settings requires balancing linguistic nuance with clarity and reader safety. Meppelink et al. (2015) and Kruger and van Rooy (2016) also indicated that simplification can function as a conscious strategy in cross-cultural health communication to reduce cognitive load and mitigate misunderstanding.

Based on previous research (Laviosa, 1998; Xiao, 2010), this paper hypothesizes that translated Chinese texts will exhibit the following linguistic characteristics: lower lexical density, more frequent use of connectives, and a higher occurrence of passive constructions. Accordingly, this study addresses two research questions:

1. Are there significant differences between translated Chinese and native Chinese cancer information in terms of lexical density, syntactic structures, and discourse features?
2. Do these differences reflect the translation universals of simplification, explicitation, and normalization?

3. Reserch Design

3.1 Corpus Building

The study constructed two corpora, a native Chinese (NC) corpus and a translated Chinese (TC) corpus, and collected health materials about cancer from four authoritative websites: Native Chinese - China Anti-Cancer Association (CACA), Chinese Health Education Network; Translated Chinese - National Cancer Prevention and Control Network (chinacpcn), World Health Organization (WHO), MSD manuals. The construction of the two corpora followed a three-phase procedure:

Phase 1: Data Collection. All texts selected for the corpus were public-facing cancer health education materials. Advertisements, duplicate webpages, and specialized research reports were excluded to ensure the relevance and accessibility of the content.

Phase 2: Text Cleaning. The quality of a corpus significantly influences the validity and reliability of research outcomes. To ensure data accuracy, formatting inconsistencies were first corrected using text editing software. This was followed by meticulous manual proofreading at the character level. The cleaned texts were then saved in .txt format for subsequent processing.

Phase 3: Text Processing. This phase involved word segmentation and part-of-speech tagging. To facilitate retrieval and quantitative analysis, cleaned texts underwent tokenization—converting continuous strings of characters into discrete, identifiable tokens(Zhu & Li, 2023). Each processed text was stored individually in the respective corpus. Two corpora (native Chinese and translated Chinese) both contain 100 texts with cancer health information. The distribution of cancer health information data on different websites is shown in Figure 1 below.

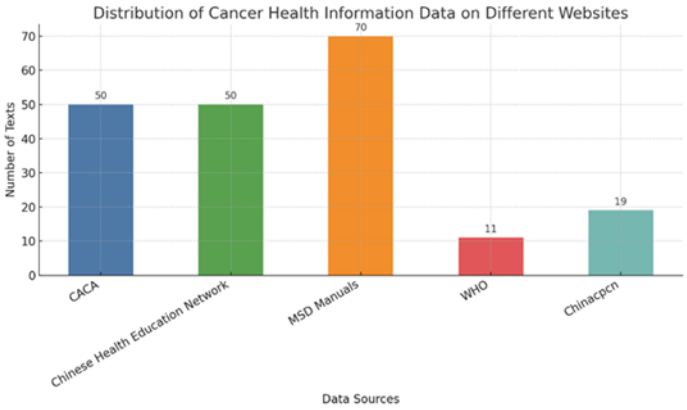


Figure 1: Distribution of Cancer Health Information Data on Different Websites

3.2 Instruments

To conduct a comprehensive analysis of the linguistic features in original Chinese and translated Chinese health information related to cancer, this study employed HanLP, AntConc, Python, and SPSS.

HanLP, an open-source NLP library, was used for Chinese word segmentation and part-of-speech tagging, essential for accurate text preprocessing. HanLP provides high-performance tools for various Chinese NLP tasks, leveraging deep learning techniques to ensure precise and efficient processing of large text corpora (HanLP, 2024).

4. The linguistic features of native Chinese and translated Chinese health information related to cancer

4.1 Lexical Features

Laviosa (1998) has pointed out that the ratio of the number of content words to the number of total words (lexical density) is lower in English translations, and Xiao (2010) has also verified that the lexical density of translated Chinese texts is slightly lower. This study will examine the hypothesis proposed by Laviosa (1998) and Xiao (2010); Xiao, He, and Yue (2010). Lexical density refers to the ratio of the number of content words to the total number of words and it can measure the information load of texts (Stubbs 1986).

$$\text{Lexical Density} = \left(\frac{\text{Number of Lexical Words}}{\text{Total Number of Word}}\right) \times 100$$

(Lexical Words include nouns, verbs, adjectives, and adverbs. Total Number of Words refers to the overall word count in the text)

This study uses an independent samples t-test to compare the lexical densities of the NC and TC groups. First, a normality test id conducted to verify the data distribution. Then, descriptive statistics are used to analyze the basic characteristics of lexical density, and finally, an independent samples t-test is used to compare the differences between the two groups.

Normality Test

To check if the data follows a normal distribution, the Shapiro-Wilk normality test was used. The results are shown in Table 1:

Table 1: Normality Test

Group	Shapiro-Wilk Statistic	df	Sig.
NC	0.991	100	0.771
TC	0.981	100	0.148

As shown in Table 1, the Shapiro-Wilk test significance values for the NC and TC groups are both greater than 0.05 (NC: 0.771, TC: 0.148), indicating that the data for both groups follows a normal distribution. Thus, subsequent parametric statistical tests can be performed.

Descriptive Statistics

The descriptive statistics show the lexical density of the NC and TC groups as presented in Table 2.

Table 2: Descriptive Statistics

Group	N	Mean	Std. Deviation	Std. Error Mean
NC	100	0.532109	0.0510661	0.0051066
TC	100	0.568139	0.0579835	0.0057983

Table 2 shows that the mean lexical density of the TC group (0.568139) is higher than that of the NC group (0.532109), and the standard deviation and standard error are also slightly higher in the TC group.

Independent Samples t-Test

An independent samples t-test was conducted to determine whether the difference in lexical density between the NC and TC groups is statistically significant. The results are shown in Table 3.

The results indicate that there is a significant difference in lexical density between the two groups ($t = -4.663$, $p < 0.001$). Since

the significance value of Levene's Test for Equality of Variances is 0.054, which is slightly above the 0.05 threshold, we assume equal variances. Therefore, the lexical density of the NC group is significantly lower than that of the TC group based on the "Equal Variances Assumed" row.

Table 3: Independent Samples t-Test Equal Variances Assumed

Statistic	Value	Statistic	Value
F	3.758	**Two-Sided p**	<0.001
Sig.	0.054	**Mean Difference**	-0.0360300
t	-4.663	**Std. Error Difference**	0.0077265
df	198	**95% CI Lower**	-0.0512667
One-Sided p	<0.001	**95% CI Upper**	-0.0207933

This study found that the lexical density of Translated Chinese (TC) is significantly higher than that of Native Chinese (NC). This finding contrasts with the translation universal of "simplification," which posits that translated texts tend to be lexically less dense than their native counterparts. The observed higher lexical density in corpus may be attributed to the news genre and the professional nature of medical and health content. Prior research suggests that the manifestation of translation universals, including simplification, varies across genres and registers. For instance, while simplification is evident in literary translation, its presence is less marked in technical and academic registers (Pang, 2024). Furthermore, studies have shown that translated Chinese texts in news and general genres can exhibit greater syntactic complexity in certain measures compared to native texts (Wu et al., 2023). This phenomenon suggests that the simplification hypothesis may not be fully applicable in specialized domains.

In line with these genre-specific tendencies, translators working on medical discourse may prioritize precision and professionalism, often opting to retain or even introduce specialized terminology. This strategy, while enhancing informational accuracy, consequently increases the lexical density and overall complexity of the translated text. This result has important implications for disseminating cancer-related information: while higher lexical density may contribute to greater precision and professional credibility, it could also elevate the cognitive demand on readers, potentially affecting the comprehensibility and accessibility of vital health information.

4.2 Syntactic Features

Unlike other categories of materials, it is essential to avoid using passive voice in health education materials to improve comprehensibility (Shoemaker et al., 2014). The study aims to discuss the frequency of using passive structures in native Chinese and translated Chinese information in online cancer health materials.

$$\text{Frequency of passive constructions} = \left(\frac{\text{Number of occurrences of "被"}}{\text{Total number of words}}\right) \times 10000$$

Table 4: Frequency of Passive Constructions

Corpus Type	Number of Occurrences of "被"	Total Number of Words	Frequency of Passive Constructions (per 10,000 words)
Translated Chinese Corpus	49	28725	17.06
Native Chinese Corpus	52	29381	17.70

The results show that the frequency of passive constructions in translated Chinese texts is approximately 17.06 per 10000 words, while in native Chinese texts it is slightly higher at 17.70 per 10000 words. Although the hypothesis suggests that translated texts might use more passive constructions due to the influence of the source language, the data indicates only a marginal difference between the two corpora.

These findings suggest that in the context of cancer information texts, the influence of the source language on the use of passive constructions in translated Chinese is not as pronounced as hypothesized. Both native and translated texts exhibit similar frequencies of passive constructions, indicating that translators may be effectively adapting the syntactic structures to align with natural Chinese usage.

4.3 Discourse Features

Chen (2006) has investigated the frequency of the use of connectives in translated Chinese texts of scientific texts and proved that the frequency of the use of connectives in translated Chinese texts is much higher than the native Chinese texts. He considered that connectives are one of the explicitation features of English-Chinese translations of scientific and technological texts. Xiao (2010) found that translated Chinese texts contained more connectives with explicit features.

The following analyses are conducted on two specialized corpora. The raw frequency data for connectives in each corpus are first presented, followed by a normalized frequency analysis to ensure a fair comparison. The native Chinese corpus, with a total

of 29,381 words, was analyzed to establish a baseline for connective usage. The following table lists the raw occurrence counts (Freq) of each connective, ranked from highest to lowest. This data provides the foundational evidence for the initial claim that connective frequency differs between native and translated texts. The translated Chinese corpus, comprising 28,725 words, was analyzed using the same methodology. The raw frequency data below allows for a direct, preliminary comparison with the native corpus (Table 5). The higher raw count for the top connective "和 (and)" is immediately apparent.

Table 5: Frequency of connectives

Raw Frequency in Native Chinese Corpus			**Raw Frequency in Translated Chinese Corpus**		
Rank	connectives	Freq	Rank	connectives	Freq
1	和	228	1	和	366
2	与	106	2	与	59
3	而	80	3	并	59
4	级	62	4	及	434
5	并	50	5	而	36
6	以及	20	6	以及	35
7	及其	3	7	及其	3
8	加之	2	8	于是	1

To better compare the usage frequency of Connectives in the two corpora we standardize the frequency to the occurrence per 10000 words. The calculation formula is as follows:

$$\text{Frequency per } 10000 \text{ words } = \left(\frac{\text{Frequency of connectives}}{\text{Total word count}}\right) \times 10000$$

To establish a precise baseline for comparison, the raw frequency data from the Native Chinese Corpus was converted into a standardized frequency per 10,000 words. The results, detailed in the table below, provide a clear profile of connective usage density in native texts. The data shows that "和" is the most densely used connective by a significant margin, followed by "与" and "而".

The same normalization process was applied to the Translated Chinese Corpus. Placing the results side-by-side with the native corpus baseline (Table 6) allows for a direct and fair comparison. The data reveals a dramatic increase in the usage density of several connectives in translated texts, most notably "和", "并", and "以及". This pattern provides concrete, quantitative evidence for the explicitation strategy employed in translation.

Table 6: Normalized Frequency of Connectives

Normalized Frequency of Connectives in Native Chinese Corpus				Normalized Frequency of Connectives in Translated Chinese Corpus			
Rank	**connectives**	**Freq**	**Connectives Frequency per 10,000 Word**	**Rank**	**connectives**	**Freq**	**Connectives Frequency per 10,000 Word**
1	和	228	77.58	1	和	366	127.41
2	与	106	36.08	2	与	59	20.54
3	而	80	27.22	3	并	59	20.54
4	级	62	21.1	4	及	434	14.97
5	并	50	17.02	5	而	36	12.53
6	以及	20	6.81	6	以及	35	12.18
7	及其	3	1.02	7	及其	3	1.04
8	加之	2	0.68	8	于是	1	0.35

A comparative analysis of the use of connectives in native and translated Chinese cancer message texts reveals that connectives are generally used more frequently in translated texts than in native Chinese texts. The normalized data in Table 6 shows that the core connective "和" has a much higher frequency in translated texts (127.41 per 10,000 words) than in native texts (77.58). Similarly, connectives like "并" and "以及" also show markedly elevated frequencies in the translated corpus. This supports the hypothesis that translated texts use more connectives in order to enhance coherence and explicitness. The higher frequency of connectives suggests that translators are more inclined to clarify the logical relationships between sentences by adding linguistic markers in the translation process making it easier for readers to understand.

Lau (2006) argued that due to the semantics of idioms can be more complex than the surface meaning then containing idioms in the text may affect the readability of the text. If the texts include idioms that the readers are unable to understand the meaning behind then the difficulty of the text will increase. The following analysis examines the use of idioms in native and translated Chinese corpora, building upon the premise that idiom usage can impact text readability.

The native Chinese corpus exhibits a rich diversity of idiomatic expressions. The following table lists all identified idioms and their raw frequency counts, demonstrating a wide variety of 24 unique idioms with a total occurrence of 27 times(Table 7).

Table 7: Raw Frequency of Idioms in Native Chinese Corpus

Idioms	Freq	Idioms	Freq	Idioms	Freq
罪魁祸首	3	大腹便便	1	罪责难逃	1
持之以恒	2	当务之急	1	老生常谈	1
治亲带故	1	抑恶扬善	1	行之有效	1
万般无奈	1	掉以轻心	1	防不胜防	1
不知不觉	1	日新月异	1	防患未然	1
势不可挡	1	泾渭分明	1	雪上加霜	1
可乘之机	1	浅尝辄止	1	力所能及	1
名列前茅	1	漏网之鱼	1	缺一不可	1

In contrast, the translated Chinese corpus shows a markedly limited use of idioms. Only three unique idioms were identified in this corpus, each occurring exactly once: "兼而有之", "久治不愈", and "截然不同".

To objectively compare the density of idiom usage between the two corpora, the total frequency was normalized to a standard rate per 10,000 words. The results, summarized in the table 8, reveal a profound disparity. The native corpus uses idioms at a rate over eight times higher than the translated corpus.

Table 8: Normalized Frequency of Idioms (per 10,000 words)

Type of corpus	Frequency of idioms (total)	Idiom frequency per 10,000 words
Native Chinese Corpus	26	8.85
Translated Chinese Corpus	3	1.04

A comparative analysis of the use of idioms in native Chinese and translated Chinese cancer information texts reveals that idioms are used significantly more frequently in native Chinese texts than in translated Chinese texts. This suggests that translated Chinese cancer information may be more inclined to use simple, clear and explicit expressions and less likely to use culturally laden or metaphorical expressions such as idioms. This strategic avoidance in translated health communication can be interpreted as an effort to enhance informational transparency, accessibility, and cross-cultural applicability, aligning with the established plain language principles crucial for effective public health messaging (Baur & Prue, 2014; Shoemaker et al., 2014). Recent studies further indicate that such linguistic simplification is a conscious strategy in cross-cultural health communication to reduce cognitive load and mitigate potential misunderstandings arising from cultural or conceptual disparities (Meppelink et al., 2015; Kruger & van Rooy, 2016). This finding underscores a key consideration for translation practice in specialized domains: achieving an optimal balance between conveying necessary cultural-linguistic nuances and prioritizing clarity, comprehensibility, and reader safety for the target audience (Chapple et al., 2021).

5. Conclusion

This study has explored the linguistic features of native Chinese and translated Chinese health information related to cancer employing a corpus-based approach. The analysis focused on lexical density, syntactic structures, and discourse features, revealing significant differences between the two types of texts. Translated Chinese texts were found to have higher lexical density, slightly fewer passive constructions, more frequent use of connectives, and significantly fewer idioms compared to native Chinese texts.

These findings suggest that the translation process affects the complexity and readability of health information, with implications for both translation studies and health communication. The higher lexical density in translated texts contrasts with the simplification hypothesis proposed in previous literature based on English corpora (Laviosa, 1998; Xiao, 2010), suggesting that translators of specialized health information may prioritize terminological accuracy over lexical economy. This increased lexical load, while enhancing scientific precision, may potentially affect comprehension accessibility for general readers.

The minimal difference in passive construction frequency between the two corpora indicates that translators effectively adapt syntactic structures to align with natural Chinese usage patterns, demonstrating successful normalization in this aspect. Both native and translated texts maintain relatively low frequencies of passive constructions, consistent with established principles for developing comprehensible health materials (Shoemaker et al., 2014).

The significantly higher frequency of connectives in translated texts provides strong support for the explicitation hypothesis (Chen, 2006; Xiao, 2010), as translators tend to enhance textual coherence by making logical relationships more explicit. Conversely, the substantially lower use of idioms in translated texts reflects a tendency toward simplification in handling culture-specific expressions, potentially to reduce comprehension difficulty for target readers.

Theoretically, this study extends the framework of translation universals to the domain of medical and health discourse, providing corpus-based evidence for these phenomena within Chinese health information. It demonstrates that simplification, explicitation, and normalization manifest in distinct, domain-shaped ways across different linguistic dimensions, moving beyond prior research focused mainly on general texts. The findings suggest that translation universals should be re-examined in relation to register and communicative function, thereby enriching the theoretical scope and explanatory power of the concept. Practically, the findings illuminate the complex decision-making process of translators in specialized domains, who must balance competing priorities such as terminological accuracy, textual readability, and cultural adaptation. This enhanced understanding offers an empirical foundation for developing more nuanced translation guidelines tailored to health communication, ultimately supporting practices that better reconcile linguistic norms with effective public outreach.

Despite these contributions, this study has several limitations that should be acknowledged. First, the analysis focused primarily on quantitative differences, while qualitative examination of how specific linguistic choices affect reader comprehension was beyond its scope. Second, the corpora were limited to cancer information from selected websites, which may restrict the generalizability of the findings to other health domains or text types.

Future research could build upon these findings by incorporating reader response studies to examine how the identified linguistic differences actually affect comprehension and perception of health information among diverse reader groups. Comparative studies across different health domains would help determine whether the observed patterns are consistent in other specialized medical contexts. Furthermore, research could investigate the potential interaction effects between different linguistic features and their collective impact on text comprehensibility.Translators should aim to balance the linguistic and cultural features of the source text with the need for clarity and comprehensibility in the target text to ensure effective communication of health information. The development of evidence-based translation guidelines that consider these complex linguistic interactions would significantly benefit health communication practices.

Funding: This research was funded by grants from Humanities and Social Sciences Project of the Chinese Ministry of Education (No. 22YJCZH258)

Conflicts of Interest: The authors declare no conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Baker, M. (2020a). Corpus-based translation studies: The challenges that lie ahead. In K. H. Kim & Y. Zhu (Eds.), *Researching translation in the age of technology and global conflict: Selected works of Mona Baker* (1st ed., pp. 44–54). Routledge.

[2] Baker, M. (2020b). Corpus linguistics and translation studies: Implications and applications. In K.-H. Kim & Y. Zhu (Eds.), *Researching translation in the age of technology and global conflict: Selected works of Mona Baker* (1st ed., pp. 9–24). Routledge.

[3] Baur, C., & Prue, C. (2014). The CDC clear communication index is a new evidence-based tool to prepare and review health information. *Health Promotion Practice*, 15(5), 629–637. https://doi.org/10.1177/1524839914538969

[4] Chapple, A., Rhodes, P., & Shaw, S. (2021). "Getting the words right": A qualitative study of language translation in clinical trials. *Sociology of Health & Illness*, 43(5), 1187–1203. https://doi.org/10.1111/1467-9566.13285

[5] Chen, J. W. (2006). *Explicitation through the use of connectives in translated Chinese: A corpus-based study* [Doctoral dissertation, University of Manchester].

[6] HanLP. (2024). *HanLP: Han Language Processing*. GitHub. https://github.com/hankcs/HanLP

[7] Kreps, G. L., & Bonaguro, E. W. (2009). Health communication as applied communication inquiry. In L. R. Frey & K. N. Cissna (Eds.), *Routledge handbook of applied communication research* (pp. 410–429). Routledge.

[8] Kruger, H., & van Rooy, B. (2016). Constrained language: A multidimensional analysis of translated English and a non-native indigenised variety of English. *English World-Wide*, 37(1), 26–57. https://doi.org/10.1075/eww.37.1.02kru

[9] Lau, T. (2006). *Chinese readability analysis and its applications on the Internet* [Master's thesis, The Chinese University of Hong Kong].

[10] Laviosa, S. (1998). Core patterns of lexical use in a comparable corpus of English narrative prose. Meta: *Translators' Journal*, 43(4), 557–570.

[11] McEnery, T., & Xiao, R. (2007). Parallel and comparable corpora: What is happening? In G. Anderman & M. Rogers (Eds.), *Incorporating corpora* (pp. 18–31). Multilingual Matters.

[12] Meppelink, C. S., Smit, E. G., Buurman, B. M., & van Weert, J. C. M. (2015). Should we be afraid of simple words? The influence of vocabulary and text simplification on recall and appraisal of health information among people with low health literacy. *Journal of Health Communication*, 20(10), 1162–1171. https://doi.org/10.1080/10810730.2015.1018600

[13] Pang, S. Z. (2024). A diachronic and multi-genre corpus-based investigation of "translation universals." *Foreign Languages and Their Teaching*, (5), 132–144, 150–151. https://doi.org/10.13458/j.cnki.flatt.005063

[14] Pöchhacker, F. (2016). *Introducing interpreting studies* (2nd ed.). Routledge.

[15] Shoemaker, S. J., Wolf, M. S., & Brach, C. (2014). Development of the Patient Education Materials Assessment Tool (PEMAT): A new measure of understandability and actionability for print and audiovisual patient information. *Patient Education and Counseling*, 96(3), 395–403. https://doi.org/10.1016/j.pec.2014.05.027

[16] Stubbs, M. (1986). Lexical density: A computational technique and some findings. In M. Coulthard (Ed.), Talking about text: Studies presented to David Brazil on his retirement (pp. 27–48). *English Language Research*, University of Birmingham.

[17] Wu, J. F., Liu, K. L., Hu, R. F., & Zhou, W. (2023). A contrastive study of syntactic complexity in translated and original Chinese. *Foreign Language Teaching and Research,* 55(2), 264–275, 320–321. https://doi.org/10.19923/j.cnki.fltr.2023.02.011

[18] Xiao, R. (2010). How different is translated Chinese from native Chinese? A corpus-based study of translation universals. *International Journal of Corpus Linguistic*s, 15(1), 5–35. https://doi.org/10.1075/ijcl.15.1.01xia

[19] Xiao, R., He, L., & Yue, M. (2010). In pursuit of the "third code": Using the ZJU Corpus of Translational Chinese in translation studies. *In Using corpora in contrastive and translation studies*. Cambridge Scholars Publishing.

[20] Zhu, C., & Li, R. (2023). A corpus-based analysis of Ezra Pound's translator style in English translations of Chinese classics. *Foreign Language Education*, 44(4), 75–82. https://doi.org/10.16362/j.cnki.cn61-1023/h.2023.04.018