A Corpus-Based Study on the Consistency of Business English Terminology Translation

Yang Wang
Zhejiang Shuren University, Hangzhou 310015, Zhejiang, China
Corresponding author: Yang Wang (e-mail: wangyang862026@163.com)

ABSTRACT Accurate semantic representation and terminology consistency are essential for intelligent information processing and cross-domain communication in data-driven engineering systems. To address inconsistencies in business English terminology translation, this study proposes a corpus-based analytical framework that integrates parallel corpus construction, statistical feature extraction, contextual analysis, and quantitative consistency evaluation. A bilingual corpus covering multiple business subdomains is established to investigate the distribution characteristics of terminology translation through lexical overlap analysis, frequency-based screening, expert verification, and entropydriven assessment. By combining translation frequency statistics, dominant translation ratios, Shannon–Wiener diversity indices, and cross-domain distribution entropy, the proposed framework systematically reveals registerdependent translation patterns and semantic constraints underlying terminology variation. Experimental analyses demonstrate that terminology consistency is strongly influenced by document type and contextual characteristics, with legal texts exhibiting significantly higher stability than marketing-oriented documents. The proposed methodology provides an effective data-driven solution for semantic standardization, knowledge organization, and adaptive information management. Beyond translation studies, the framework offers methodological references for intelligent semantic alignment, information fusion, and communication-oriented data processing in engineering systems, supporting potential applications in Electromagnetic Waves, Antennas and Propagation where accurate knowledge representation, adaptive information exchange, and heterogeneous data integration are critical.

INDEX TERMS parallel corpus, semantic consistency analysis, intelligent information processing, knowledge representation

I. INTRODUCTION

AS business English has become inseparably linked with economic globalization and the technical evolution of the industrial heritage sector, it serves as an indispensable language tool for the international communication of fiber manufacturing excellence. This integration ensures the structural consistency of technical lineages and industrial semantics within the broader landscape of global trade and international business discourse. As terminology is the main part of information provided by business text, its correct translation is of the utmost importance. However, the reality of business translation and operation industry has a conspicuous issue: incoherence of terminology translation. It means that one term is always accompanied with several different translations within one project, in different documents, or even in one single document[1-2]. This leads to inconsistent translation quality, confuses readers, and, in legally binding documents such as contracts and letters of credit, can potentially cause serious commercial disputes and economic losses [3-4]. Therefore, standardizing and ensuring the consistency of business English terminology translation is a critical issue that urgently needs to be addressed.

Scholars have explored the issue of consistency in English terminology translation from multiple perspectives [5-6]. Some studies focus on the impact of translation tool use on students' translation behavior. For example, Boonmoh Atipat investigated translation problems caused by the use of tools in online translation classes[7]; Inderawati Rita explored how students can use Google Translate to improve their writing skills[8]; Pratama Aditiya Aziz analyzed the translation quality of cultural vocabulary in tourism texts[9]; and Banat Maysaa examined the linguistic features of GPT-4 (Generative Pre-trained Transformer 4) machine-generated texts from a stylistic perspective[10]. However, although the above studies involve different aspects of translation practice, their focus is either on teaching scenarios or on the surface of texts, and they have not yet touched on the

core issue of consistency in terminology translation. Their conclusions are mostly prescriptive suggestions on "how to translate", which are difficult to systematically reflect the complex situation of "how to translate in practice" in industry practice, and they cannot reveal the prevalence, distribution patterns and underlying constraints of inconsistencies in terminology translation.

In recent years, the rise of corpus-based translation studies has provided a new methodological path for solving the above problems [11-12]. Hu Xianyao, based on the construction topology translation view, examined the distribution and cognitive rationale of specific constructions in translated English [13]; Hu Kaibao explored the image of the Chinese government constructed in the English translation of the Government Work Report [14]; Granger Sylviane systematically defined the research method of learner translation corpus [15]; PIAN YW classified and optimized the English translation strategies of culturally loaded words [16]; Li Qiang analyzed the translation strategies of scientific terms and cultural elements in geological tourism texts [17]. The above studies fully demonstrate the applicability and superiority of corpus methods in translation studies. However, whether it is the distribution of constructions in literary translation, the image construction in political texts, or the translation strategies of cultural words and geological terms, their research objects are all far removed from the business context. More importantly, while existing corpus-based research has focused on terminology translation, there is a lack of systematic empirical studies specifically addressing the consistency of business English terminology, particularly neglecting the dynamic constraints of different registers within business texts (such as legal contracts and marketing letters) on terminology translation choices. The research gap is the basis of this work.

Is the translation of business English terminology really as "chaotic" as the industry thinks it is? Are the differing translations of the same term simply the result of random mistakes, or are they confined to deeper linguistic rules? Are there systematic differences in the translation choice between the same term in a legal contract and in a marketing letter? What characteristics of translation behaviour do these differences represent? This paper will look at these types of questions with the help of an existing parallel corpus of Business English and Mandarin Chinese. The main purposes of this study are to clarify the consistency of Business English terminology by systematically examining the distribution of translations across registers, specifically integrating the technical semantics and historical manufacturing lineages of industrial heritage. By mapping these specialized linguistic patterns, the research provides data and theoretical references to improve terminology management and the digital dissemination of industrial manufacturing excellence.

II. A CORPUS-BASED APPROACH TO EVALUATING THE CONSISTENCY OF BUSINESS ENGLISH TERMINOLOGY TRANSLATION

CORPUS DESIGN AND CONSTRUCTION

Corpus-based studies on consistency of translation of business English terms rely on the corpus to analyse the trends of terms within the corpus[18-19]. Quality of the corpus will affect the results of terminology extraction, identifying a term's translation, and analysing consistency of the translated terminology. Therefore, when constructing a corpus for business terminology research, three guidelines must be followed: the authenticity of the corpus source (texts must be from actual business operations), the representativeness of the register structure (the linguistic characteristics of various business scenarios must be reflected in the texts), and the computability of the data (the data must be in such a structure that enables quantitative analysis). In this case, the corpus will consist of English-Chinese texts concerning business, and will be a medium-sized parallel corpus of business English. The primary use of this corpus is as a systematic record of terminology translation behaviour across different situations involving business, thereby providing a verifiable foundation for analysis[20-21].

A reasonably designed corpus size is a prerequisite for ensuring the stability of statistical results. After collecting the corpus, the overall size must be calculated first to ensure a reasonable distribution of different text proportions.

$$N = \sum_{i=1}^{m} n_i \tag{1}$$

N represents the total number of words in the original English texts in the corpus, n_i represents the number of words in the i-th text, and m represents the total number of texts included in the statistics in the corpus. According to the research design, this study selected five types of business texts: international sales contracts, letters of credit, listed company annual reports, business correspondence, and cross-border e-commerce product descriptions. Ten to fifteen representative texts were selected from each subdomain to ensure the corpus covers the linguistic expression characteristics of different business scenarios, thereby improving the representativeness of the terminology statistics results. Although different sub-domains of business English belong to the same business category in terms of macrofunction, their vocabulary usage characteristics should have identifiable differences—legal contracts tend to use modal verbs and archaic adverbs, while letters of credit are filled with fixed financial terminology. If the vocabulary overlap between two sub-domains is too high, it indicates that they are converging in linguistic features, and the original classification boundaries need to be re-examined, as shown in Figure 1.

Figure 1 shows the Jaccard overlap coefficients between each domain in a 5×5 matrix on the left heatmap. The overlap coefficient between contracts and letters of credit reaches

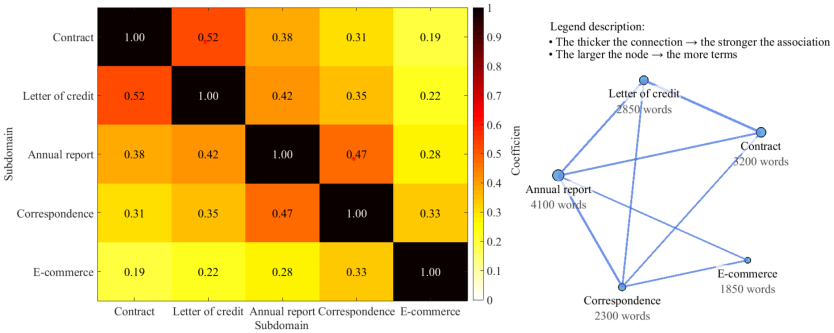

FIGURE 1. Lexical Overlap Analysis of Sub-Registers in the Corpus

0.52, reflecting a high degree of vocabulary correlation in legal and financial texts. The overlap coefficient between annual reports and correspondence is 0.47, reflecting the vocabulary penetration between information disclosure and daily communication texts. The overlap coefficient between the e-commerce domain and other domains is generally low, ranging from 0.19 to 0.33, confirming the unique vocabulary selection in marketing texts. The network diagram on the right documents the relationships between nodes and edges. In the node-edge format, the nodes correspond to the total number of terms that occur in various domains. It can be seen that annual reports contain the most terms and e-commerce contains the least.

This study's scale is limited to an approximately 250,000-word English original text (around 400,000 translations into simplified Chinese), creating a parallel corpus with about 500,000 alignment units. This scale follows the "moderate principle" of corpus construction for specific purposes: a small scale makes it difficult to capture low-frequency terminology translation patterns, resulting in unstable statistical results; a large scale requires significantly higher computational resources and more intensive manual alignment for professional sub-domains, reducing the operability and processing efficiency of the research under specific resource constraints. The corpus contains approximately 250,000 words of original English text, ensuring that the frequency of core business terms meets statistical significance requirements while also accommodating the feasibility of manual verification and annotation [22-23]. To verify whether the frequency of the research subjects in the corpus is sufficient, a terminology coverage index is introduced:

$$\tau = \frac{K_{\text{obs}}}{K_{\text{total}}} \tag{2}$$

K_{obs} represents the number of core terms actually appearing in the corpus, and K_{total} represents the total number of terms in the pre-defined research term list. If the τ value is too low (e.g., below 0.8), it is necessary to retrospectively adjust the term selection criteria or expand the corpus size to ensure that subsequent analysis has a sufficient statistical basis.

To be properly analyzed by a computer system you must take your unstructured document (raw text) and convert it into structured data[24-25]. To begin this process; you must create an electronic (digital) copy of the original document. Usually, you will do this by scanning in your hard copy using OCR (Optical Character Recognition); however, even OCR conversion of your original document and PDF (Portable Document Format) will require a second level of adjustment to get your data to align properly. Also, during the OCR stage there is a high possibility that you will introduce some mistakes into your data. To fix these errors, you will need two researchers to proofread the entire data set to ensure general data cleaning, which includes correcting spelling mistakes in technical terminology, validating numerical indicators, and standardizing document formatting. To prepare a parallel corpus, the alignment stage used a machine-assisted tool (i.e., Trados Aligner) to create the first pass of alignments automatically using the text lengths and bilingual dictionaries; this gave us preliminary alignment projects. Manual adjustments were made for areas of difficulty like multiple clauses in a contract that could not easily be matched or where no two salutation formats could be aligned. Where the sentences to be aligned could not be aligned automatically, the sentences were divided and/or merged together using manual processes to complete the alignment process. Once the entire alignment process was completed the results were stored in a plain text file with UTF (Unicode Transformation Format) – 8 encoding, where the alignment units were on one line each and both the English and Chinese sentences were linked by the same corresponding separation character. Once the structure of the parallel corpus/data was set up, it created a lexical overlap index in order to verify that the two corpus domains have been separated appropriately so that the future comparative analysis in different domains will result in significant differences between each domain. The business English-Chinese parallel corpus created as part of this study was created with the above mentioned planned design and executed adequate control, such that the corpus's size was sufficient for practical use in relation to the domain structure, to provide reliable data for the purposes of extracting terminology and for other activities.

TERM EXTRACTION AND SCREENING

Once the corpus has been created, one then needs to extract terms from the corpus of business that have research value. The definition of business English is very different from the definition of "business" in general. Business English terms are neither synonymous with rare words nor considered to be high frequency words; they are words that represent fixed professional concepts (that are the same between speakers of English) and carry semantic specializations in reference to a specific business context [26-27]. For example, in nonbusiness contexts, "premium" could be defined as "of high quality" however it would be defined as "an insurance premium" when used in a business context by an insurance agent; "futures" is not generally used in non-business contexts but used only in reference to a "futures contract" in a financial context. Because of this, the extraction process cannot only be completed with the use of the frequency rank or the intuition of a human but requires an extraction method that combines both a quantitative screening process and a qualitative identification process in order to create a balance of statistical significance and professional representativeness.

The term extraction process first scans the vocabulary level of the corpus, statistically analyzes the word frequencies of the English portion of the parallel business English-Chinese corpus, and generates a descending word frequency table. This step transforms the text into a lexical probability distribution model, and word frequency can be considered an initial measure of lexical importance [28-29]. However, high-frequency words often include function words (such as "the," "of," and "and") and general content words, which, despite their high frequency, lack subject-specific terminology. Therefore, it is necessary to introduce an external reference system to screen business terminology, comparing the word frequency table with existing business English terminology lists (such as the Business English Keyword List and the Financial Terminology Database), and selecting words that appear simultaneously as candidate terms. To quantify the screening threshold, a word frequency threshold screening function is introduced:

$$I_{\text{freq}}(k) = \begin{cases} 1, & \text{if } f(k) \geq f_{\min}, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

$I_{\text{freq}}(k)$ represents the indicator function for whether candidate term k passes the word frequency threshold, $f(k)$ is the actual frequency of candidate term k in the corpus, and $f_{\min}$ is the preset minimum frequency threshold.

After word frequency screening and comparison with professional thesaurus, an initial list of candidate terms is obtained, but noise still exists. Words like "company" and "business" have high frequency in business texts and are included in professional thesaurus, but they are general terms and not very meaningful for revealing specific business concepts; including them would dilute the focus of the analysis. There are also some words with low frequency, but they are highly specialized in business practice and their translations are easily inconsistent (such as "force majeure" in legal clauses and "deadfreight" in transportation clauses), and may require specialized semantic refinement despite being present in initial lists. A professional manual verification process should be established, whereby two independent expert translators with over five years of experience in translating similar business documents independently examine the terminology used; remove general high-frequency words from the document; create and/or add industry-specific terms to the document; and establish a rating reliability index to measure the degree of consistency among the ratings provided by each of the professional manual verification reviewers.

$$\eta = \frac{P_o - P_e}{1 - P_e} \tag{4}$$

The agreement rate between two experts is represented by P_o, while P_e indicates the expected agreement under random conditions. This has been confirmed through measuring agreement with and without term modifications. For instance, if η< 0.6, then further discussion between experts is required to resolve any discrepancies and ensure that the subjective component of the final term list has a basis in reproducible reliability. Following manual verification and supplementation of terms, the final core term list is to be limited to approximately 200 terms. The selection of 200 terms is not arbitrary but represents a trade-off between size of the term list corpus and depth of research. In order to theoretically show that this is a rational choice, the term index discriminating power is also used:

$$D(t) = \frac{f_{\text{domain}}(t)}{\sum_{i=1}^{N} f_i(t)} \log\left(\frac{N}{df(x)}\right) \tag{5}$$

$f_{\text{domain}}(t)$ denotes the occurrence of a word/phrase within the selected product or service business domain while the denominator is the sum of occurrences of that same word/phrase from all of the different comparative domains (i.e., General English Corpus) and where $df(x)$ is how many of the comparative text corpora contain that word/phrase. Thus, this equation draws together both how frequent a particular word/phrase occurs within the applicable business domain (frequency ratio) as well as how often it appears in comparison with other domains (inverse document frequency). In general - the higher the value of $D(t)$, the more standardised the vocabulary associated with this business terminology can be considered to be. The calculation of the $D(t)$ values for an approved candidate list will indicate whether or not the result of having manually screened them conforms to what can statistically be expected from them as "terminology" - as demonstrated in Figure 2.

As seen in Figure 2, the absolute number of occurrences of each term from the corpus are represented by the blue bars on the left. "Company" and "Business" are far more prevalent than any other terms, with 892 and 756

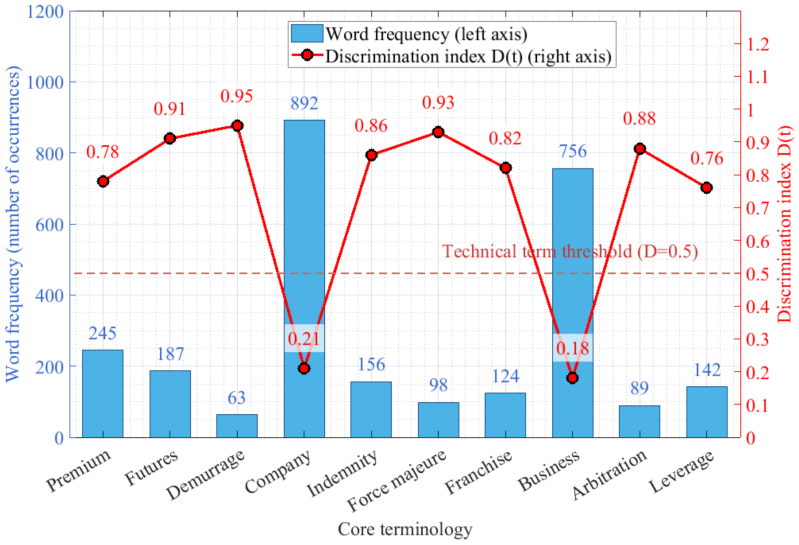

FIGURE 2. Biaxial Plot of Core Term Frequency-discrimination Index

occurrences, respectively, making these two terms high-frequency, generally used terms. On the other hand, the two terms "demurrage" and "force majeure" have far fewer occurrences, with only 63 and 98 occurrences, respectively, indicating that they are low-frequency terms. Additionally, the red line graph and the data point labels (to the right) represent the discriminative power index $D(t)$ for each term. The discriminative power index for "demurrage" is 0.95, for "force majeure" is 0.93, while for "company" is 0.21, and for "business" is 0.18, which means that demurrage and force majeure have very high discriminative power, while company and business do not, suggesting there is an inverse correlation between word frequency and discriminative power.

After identifying 200 core terms, it is also necessary to evaluate their even distribution across the five subdomains to avoid over-concentration of the research object in one type of text, thus losing representativeness. Therefore, the term cross-domain distribution entropy is introduced:

$$H(t) = -\sum_{j=1}^{5} p_j(t) \log p_j(t) \tag{6}$$

$H(t)$ represents the distribution entropy of term t across the five subdomains, and $p_j(t)$ is the proportion of the frequency of term t in the j-th subdomain to its total frequency. The distribution entropy ranges from 0 to $\log 5$; $H(t) = 0$ when the term is completely concentrated in one subdomain, and $H(t) = \log 5$ when the term is evenly distributed across all subdomains. This index can identify two special types of terms: one is cross-domain general terms (high entropy), such as "contract," which appears in almost all business texts; the other is domain-specific terms (low entropy). These two types have different values in consistency analysis. The consistency of cross-domain term translations reflects the degree of standardization in the business domain, while domain-specific terms reveal the terminology management situation in a specific subdomain. Cluster analysis of the distribution entropy of 200 terms can further subdivide the research objects into core terms across the entire domain and domain-specific terms, providing a basis for stratified discussion.

TRANSLATION EXTRACTION AND FREQUENCY STATISTICS

To obtain the Chinese form for the selected business vocabulary, data needs to be compiled from parallel corpora and you must perform a frequency analysis of the data collected. The results of this analysis will provide a foundation for conducting research into consistency in how business terminology will be translated. The goal is to collect data that provides an accurate depiction of the types of translations that ultimately occur in the business sector, as well as to provide a statistical basis for evaluating the translations [30-31]. Given that individual business terms may have multiple translations depending on context, both tool collection and manual reviewing methods have been used in order to verify the accuracy and meaning of each translation. In the translation extraction stage, the English terms in the corpus are first systematically retrieved. This study uses Parallel Concordancer software to locate all occurrences of each core term in the corpus through keyword retrieval. The results are presented in index rows, including the local context of the term and its corresponding translation. By retrieving all parallel sentence pairs containing the target term, a complete set of terminology contexts is formed, providing corpus support for translation identification. In order to measure the degree of use of a term in the corpus, its occurrence frequency can be counted first. Let the occurrence frequency of a term t in the corpus be n_t, then the relative occurrence frequency of the term can be expressed as:

$$RF(t) = \frac{n_t}{N} \tag{7}$$

After completing the terminology retrieval, the Chinese translations in each parallel sentence pair need to be manually identified to determine the corresponding translations of the terms in specific contexts. Because automatic word alignment technology may have errors in handling complex semantic relationships, this study uses manual annotation to confirm translations. During annotation, researchers read each parallel sentence pair containing the term, judging the specific expression of the term translation based on the sentence's semantics, and recording it as a standardized string. For example, for the term "indemnity," if the translation has different expressions such as "peichang" (compensation) or "buchang" (reimbursement), each is recorded as an independent translation. To ensure consistent annotation, the recording must be standardized, unifying synonym forms or merging identical expressions. Simultaneously, the annotation needs to record the subdomain to which the translation belongs, such as contract, letter of credit, business correspondence, etc., to analyze the differences in the use of translations in different business contexts. After the translation annotation is completed, all translations need to be systematically counted. First, the absolute frequency of each translation is counted to clarify the distribution of term translations in the corpus.

After obtaining the absolute frequency, the relative proportion of each translation among all translations also needs to be calculated to more intuitively reflect the degree of use of different translations. The equation for calculating the relative frequency of translations is as follows:

$$P_m = \frac{f_m}{F} \tag{8}$$

P_m represents the relative frequency of the m-th translation among all translations. This indicator can determine whether a particular translation is dominant in translation. If the relative frequency of a translation is much higher than others, it is more common in translation practice and more likely to become the standard translation; conversely, if the relative frequencies of multiple translations are similar, it indicates that there are significant differences in expression of the term during translation. In addition to analyzing the frequency of translations, it is also necessary to count the number of translation types of a term, that is, how many different translations of a term are available in the corpus, which reflects the diversity of term translation. The calculation method is as follows:

$$V = |\{c_1, c_2, \cdots, c_k\}| \tag{9}$$

V represents the number of term translations, c_k represents different translation expressions, and the symbol $|\cdot|$ represents the number of elements in the set. A larger V value indicates that the term has multiple translations during the translation process, suggesting potentially lower translation consistency; conversely, a smaller V value indicates that the term translations are more concentrated, resulting in more stable translations.

The potential for obtaining a complete distribution of business terminology in a text can be achieved through both translation extraction and the use of frequency statistics. Using these two methods, the statistical results reveal the frequency of business terms within the corpus and if there are any patterns in their expression, meaning that the statistical results allow for future analyses of translation consistency and aid in supporting quantitatively those analyses. For example, by comparing the frequency distributions of various translations due to business terminology, it may be able to determine the dominant expressions used in translation practice and show that there are substantial differences in the translation expressions used for business terminology. As such, there is also an opportunity to analyze how various types of business texts affect the selection of terminology in translating business text. Therefore, using both translation extraction and frequency statistics can provide an overall understanding of how frequently a business term is being translated, as well as provide data for the construction of evaluation models and standardization studies.

CONTEXTUAL ANALYSIS AND CONSISTENCY QUANTIFICATION

Following the completion of the frequency statistics and extraction of translation terminology, an assessment of the consistency of translations of business English terminology will require both quantitative indicators and contextual analysis [32-33]. Frequency statistics alone do not adequately explain the reasons for different translations occurring; therefore, this study uses both contextual observation and quantitative geographical indicators for a comprehensive analysis of the usage of terminology translation. In conducting the contextual analysis, the KWIC (Key Word in Context) index rows from corpus retrieval methods are used to determine the context of the terms, as well as to find the correlation between translation selections and the context in which they are used; simultaneously, numerous quantitative indicators will be presented for statistical examination of the translation concentration and diversity, thereby allowing for an assessment of the consistency characteristics of business English terms' translation through data analysis.

For the contextual analysis part of your research, one of the focuses will be on those terms that have multiple translations [34-35]. In general, if a term has three or more translations within a corpus, it's highly probable that it has significant differences in its expressions during translation. Therefore, this study has selected a number of terms with a high number of translations; the objective is to perform the primary analysis using the Parallel Concordancer software to find all of the KWIC index rows that contain those terms, as well as their respective contexts before & after them. The user's interface in the index row will allow re-

searchers to view both the original English sentence and the respective Chinese noun from each KWIC; therefore, able to identify specific contexts in which differing translations appear for each term. Some legal contract materials may require references to certain words which would require those words to be represented by simpler terms in business correspondence or marketing materials, while other sources will link words through their collocation with other words to yield consistent collocated translations over time. The relatively high diversity indices observed in annual reports (0.8) and correspondence (0.93) can be attributed to their communicative functions: annual reports must balance technical precision with narrative disclosure for diverse stakeholders, leading to varied synonymous expressions; meanwhile, business correspondence is highly personalized and context-dependent, often prioritizing stylistic flexibility and interpersonal rapport over rigid terminological standardization. For example, the collocation "indemnifying against losses" tends to be translated as "compensating for losses"; however, it can also appear as "compensation" or "liability to compensate." In turn, a thorough analysis of collocation will allow for an explanation of translation differences, as well as a semantic basis for assessing subsequent levels of consistency.

To conduct a statistical assessment of the translation of a term based on the contextual information, will require the application of quantitative measures to analyze the distribution of the translations. First, calculate the proportion of dominant translations to measure the concentration of a particular translation among all translations. Let $f_{\max}$ be the number of times the most frequent translation of a term appears, and F be the total number of times the term appears in the corpus. Then, the proportion of dominant translations can be expressed as:

$$D = \frac{f_{\max}}{F} \tag{10}$$

When the D value is close to 1, it indicates that most translation instances use the same translation, indicating high consistency in terminology translation; a low D value indicates that translations are scattered and translation differences are significant. This indicator can quickly identify terms with stable or unstable translations. However, relying solely on the proportion of dominant translations is insufficient to fully reflect the distribution of translations, thus further analysis of translation diversity is needed. This study introduces the Shannon-Wiener diversity index, originally used in ecology and later commonly used in linguistics to describe the dispersion of linguistic units, to measure the complexity of translation selection. Compared to the standard Type-Token Ratio (TTR), which primarily measures lexical richness at a surface level, the Shannon-Wiener index accounts for the proportional distribution of specific translation variants. This allows for a more nuanced quantification of translation consistency by penalizing high-entropy distributions where multiple translations are used with similar frequencies, which is critical for identifying register-driven translation patterns in business contexts. Its calculation equation is as follows:

$$H' = -\sum_{m=1}^{k} p_m \ln p_m \tag{11}$$

H' represents the translation diversity index, and p_m represents the relative frequency of the m-th translation among all translations. When the H' value is small, it indicates that the translations are concentrated, with most translation instances using the same expression; when the H' value is large, the translations are evenly distributed, with diverse translation choices, reflecting a low level of consistency. Combining this index with the proportion of dominant translations allows for a more comprehensive assessment of terminology translation consistency. H' integrates the types and proportions of translations; H' increases when the translation distribution is even, and approaches zero when a certain translation is absolutely dominant. A scatter plot of the number of translation types V versus H' is drawn, with a trend line and a $V = 3$ reference line, as shown in Figure 3.

As shown in Figure 3, most terms have between 2 and 5 possible translations. The H' value near V=3 exhibits significant dispersion, accurately illustrating the phenomenon of "different distribution uniformity despite the same number of translation types": even with 3 possible translations, some show low diversity due to a dominant translation, while others show high diversity due to even distribution. This highlights the advantage of the Shannon index. The black dashed line in the figure represents the theoretical maximum curve H'=ln(V), the red dashed line is the V=3 reference line, and the black solid line is the quadratic trend line, indicating that the diversity index grows non-linearly with increasing translation types. The 23 highly diverse terms are marked with red pentagrams and are key management areas.

III. EXPERIMENT AND RESULTS ANALYSIS

DATA SOURCES

This research utilizes a self-made parallel corpus of both Chinese and English business corpora. The processes of constructing the corpus will follow the principles of authenticity, representativeness of register structure and computability. The parallel corpora contain five types of business sub-registers; including: written documents for international sales contracts; letters of credit; annual reports for listed companies; business correspondence; descriptions of products for sale via cross-border e-commerce. The contracts and letters of credit are primarily derived from the UNCITRAL Open Model Law, the ICC Standard Clauses Library, and anonymised contract templates from very large domestic foreign trade companies. The data for the annual reports are from the previous five years' worth of English language annual reports of publicly traded companies in Shanghai,

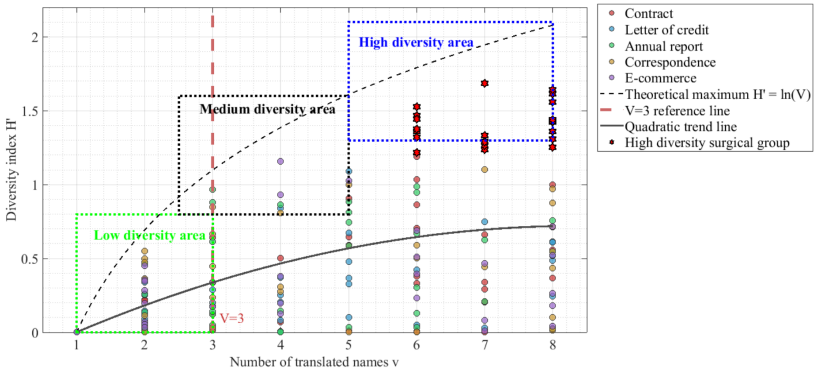

FIGURE 3. Relationship between the Number of Translation Types and the Diversity Index

Shenzhen, and Hong Kong. The information used to generate the business correspondence is based on both foreign trade practice textbooks and a sample of actual corporate email correspondence (all of which were anonymised with consent). Cross-border e-commerce text is collected from a diverse range of product pages on the Amazon platform, including both top-tier international brands and a significant proportion of small-to-medium enterprise (SME) listings. This sampling strategy ensures that the corpus captures a broad spectrum of translation qualities, from professional localized content to the less standardized translations typical of individual sellers, thereby providing a more representative basis for studying terminological inconsistency across the sector. Each piece of text is manually vetted for duplication and confirmed to represent the business practices of that company. The overall amount of original English text has been limited to approximately 250,000 words (which equates to about 400,000 words when translated into Chinese), thereby creating a parallel corpus where the total number of aligned units is over 500,000. The size will allow for sufficient frequencies to statistically signify core business terms, but also to accommodate the need for reliable manual annotation/verification of those terms. To process this corpus, the researchers utilized high-precision optical character recognition (OCR) technologies and converted formats for use throughout the process. To assist with these processes, each researcher was provided copies of all text and their research experience in business translation was used to proofread the text side-by-side; they focused primarily on correcting spelling errors in professional terminology, formatting/number representation errors and errors due to special formatting. The end result of their efforts was a consistent record of the text in plain UTF-8 encoded text files, which will provide reliable resources for subsequent term extraction and quantitative analyses of consistency and frequency.

RESULTS ANALYSIS

ANALYSIS OF THE REGISTER DISTRIBUTION CHARACTERISTICS OF CORE TERMS

To further reveal the structural distribution characteristics of business English terminology across different domains, this study conducted stratified statistical analysis of the frequency distribution of 200 core terms across five sub-domains. Based on the terminology discrimination index $D(t)$, the terms were divided into three levels: "low-frequency, high-specialization terms" ($D(t) \geq 0.8$), "medium-frequency, medium-specialization terms" ($0.4 \leq D(t) < 0.8$), and "high-frequency, general terms" ($D(t) < 0.4$). The frequency of their occurrence in five types of texts—contracts, letters of credit, annual reports, correspondence, and e-commerce documents—was then statistically analyzed. (as shown in Figure 4)

Figure 4 presents the structural distribution characteristics of business English terminology across different registers. The frequency of low-frequency, highly specialized terms in contracts and letters of credit reaches as high as 2450 and 2180 respectively, while e-commerce only has 320, showing a significant difference. However, the frequency of high-frequency, general terms in e-commerce texts reaches 5230, far exceeding the 980 in contract texts. The average terminology discrimination index marked at the top of the bars is 0.82 for contracts and 0.36 for e-commerce, quantitatively demonstrating the "normative gradient" from legal and financial texts to marketing texts. This result strongly confirms the core finding of the paper: the use of business English terminology is significantly constrained by register; legal texts use more highly specialized terms, while marketing texts mainly use general vocabulary, providing direct data for subsequent analysis of the constraints of register on translation selection.

After clarifying the register differences in the specialized attributes of terminology, it is necessary to grasp the overall evenness of the distribution of 200 core terms across the five sub-registers. Cross-domain distribution entropy $H(t)$ is a quantitative indicator measuring this uniformity. When terms are completely concentrated in a certain domain,

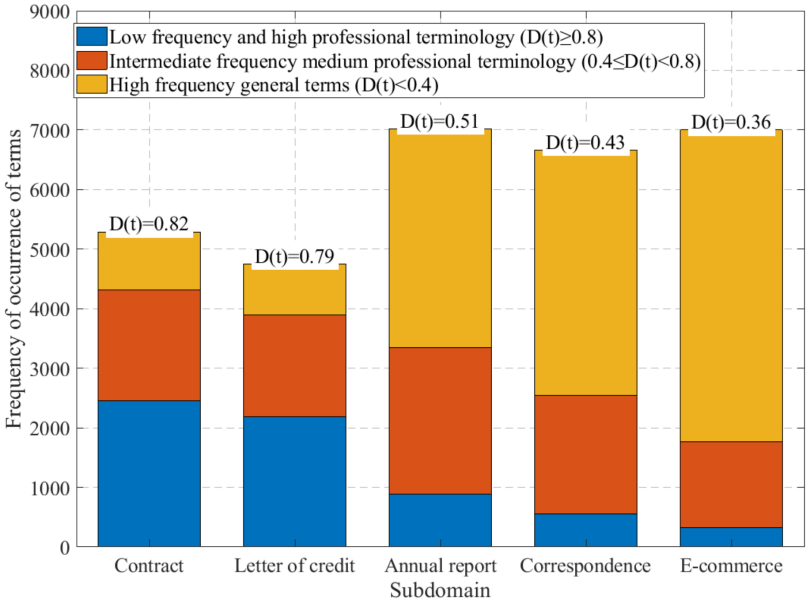

FIGURE 4. Stratification of Register Distribution of Core Terms

$H(t) = 0$; when they are uniformly distributed across all domains, $H(t) = \log 5 \approx 1.61$. By statistically analyzing the distribution entropy of 200 terms, two types of term groups can be identified: "cross-domain general terms" and "domain-specific terms." Figure 5 presents the overall pattern of the cross-domain distribution entropy of the 200 terms.

Figure 5A shows a histogram of frequency distribution of entropy. The two red dashed lines mark the thresholds H=0.8 and H=1.2, dividing terms into low-entropy, medium-entropy, and high-entropy regions. Data distribution shows that terms are mainly concentrated in the 0.7 to 1.1 region, with relatively fewer terms in the high-entropy region. Figure 5B is a pie chart of terminology classification. Low-entropy terms (H<0.8) total 86, accounting for 43%; medium-entropy terms (0.8≤H≤1.2) total 68, accounting for 34%; and high-entropy terms (H>1.2) total 46, accounting for 23.0%. This visually presents the proportion of the three types of terms. Low-entropy terms correspond to domain-specific terms, highly concentrated in specific text types such as contracts and letters of credit, and their translation choices reveal the terminology management characteristics of specific sub-domains. High-entropy terms correspond to cross-domain general terms, commonly appearing in various domains, such as "contract," and their translation consistency reflects the overall standardization of business English. Medium-entropy terms fall between the two.

CROSS-REGIONAL CONSISTENCY ASSESSMENT

There is a next step to examining the constraints of various types of business text on the consistency of terminology translation is to further explore the limitations imposed on the consistent use of terms due to the differences between the five sub registers. The five sub-registers differ significantly in terms of their textual functions: Contracts and letters of credit are legal regulatory documents and have very strict requirements for the accuracy of terminology; Annual reports are an informative disclosure document and their use of terminology tends to be fairly standardized; Correspondence is a text used on a daily basis for communication; thus, the expression used in this text is generally more flexible; E-commerce product descriptions are predominantly marketing documents and therefore there is a greater emphasis on the readability and attractiveness of the text. These differences in function will likely result in differences in terminology translation strategies. Figure 6 shows the distribution of the diversity index for the 200 core terms across each of the five sub-registers.

The distribution of the translation consistency of business English terminology according to register is shown in Figure 6. The mean H' for contracts is 0.38, with a small concentrated box and with a red median line at approximately 0.3, indicating a relatively high degree of consistency in the translation of legal terminology. The mean H' for letters of credit is 0.44, which is still low but a little higher than contracts. The mean H' for annual reports and correspondence are 0.80 and 0.93 respectively, with higher box heights, indicating differences between these registers. The mean H' for e-commerce texts is as high as 1.19, with the largest and widest box, suggesting that there are many different translations for marketing texts that lack consistency. The dots in the figure represent outliers, which mostly occur in the contracts and letters of credit registers. These findings support the main argument of the research paper: that the consistency of business English terminology in translating is heavily influenced by the constraints imposed by the particular register and that the constraints differ according

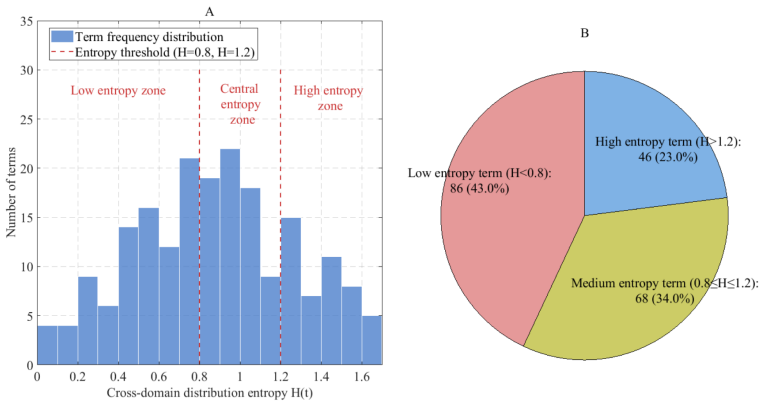

FIGURE 5. Cross-domain Distribution Entropy Analysis of Core Terms: (A) Histogram of Distribution Entropy Frequency Distribution; (B)

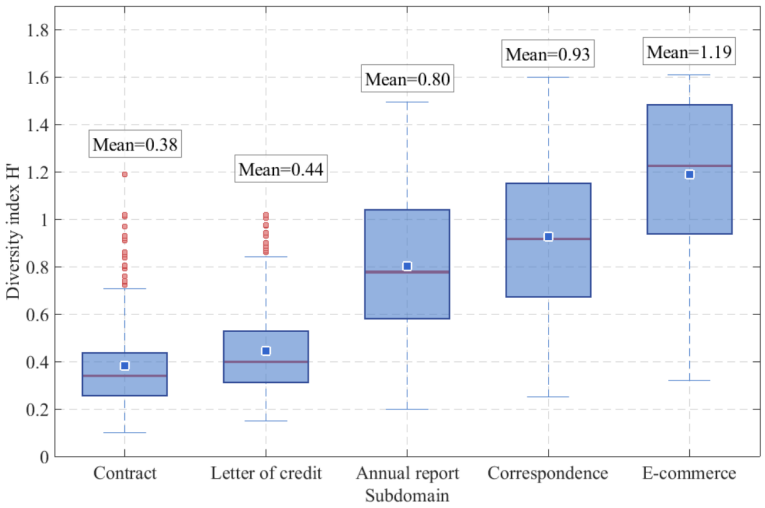

FIGURE 6. Box Plot of the Diversity Index of Translation Names for Five Sub-registers

to the function of the register.

COMPARISON OF INTRA-TEXT AND CROSS-TEXT CONSISTENCY

Once the major discrepancies in the translation diversity are investigated across registers, a further analysis of instances of terminology translation consistency must take place at the textual level through the use of an intratextual consistency coefficient. This coefficient identifies the degree of consistency of how a particular term was translated within one text (i.e., how often was that term translated differently in that same text). A high level of frequency in how different translations of the same term are used in a single text could lead to cognitive confusion and even legal uncertainty for readers. The cross-textual consistency coefficient identifies the variations between the dominant translation of the same term when different texts are considered; therefore, a smaller coefficient of variation indicates a high level of cross-textual stability in terminology translation. These two indicators provide complementarity from micros (i.e., intratextual) and macros (i.e. cross-textual) perspectives, and their respective specific structures are shown in Figure 7.

Figure 7 presents the internal consistency characteristic of the terminology translation at text level from several aspects. As Figure 7A shows, it's the coefficient of intra-text consistency. Red dashed line represents the overall mean 0.923. Most values are right-skewed on the distribution diagram and the center is above 0.92. Yet if observe on layer stack, a significant cluster appears in area of <0.88 on the left and majority terms are contributed by e-commerce text, relatively high in portion. Contracts and letter of credits seem exclusively appear in high-value region. Statistical information box shows contracts mean 0.96, e-commerce mean 0.871, revealing the register differences hidden behind the "high overall mean." As illustrated in Figure 7B, cross text consistency coefficient variation depicts a total average of 0.115 (the red dashed line) and a volatility threshold of 0.15 (the red solid line). Additionally, high-frequency words tend to show a greater coefficient of variation, while low frequency words typically fall below 0.05 coefficient

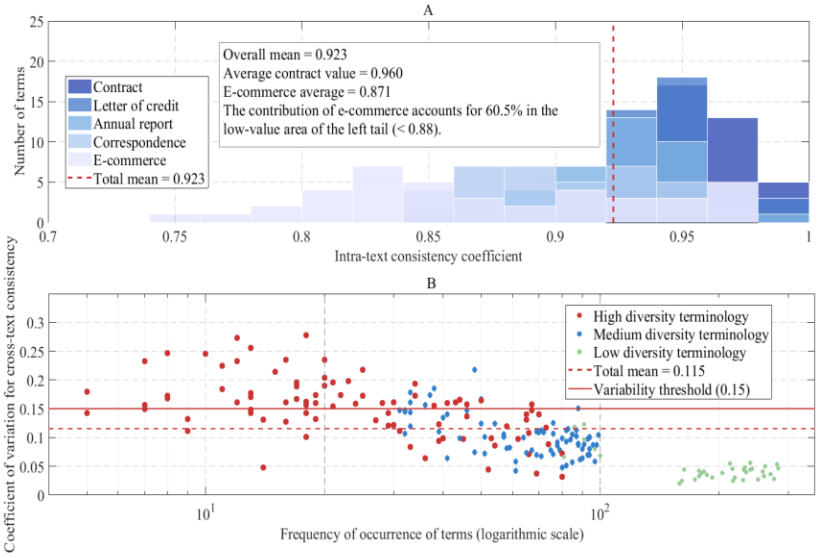

FIGURE 7. Analysis of Intra-Text Consistency Coefficient and Cross-Text Consistency Coefficient of Variation: (A) Intra-Text Consistency

of variation. The intra-text consistency of legal language has a higher level of uniformity due to being governed by register than marketing language which shows a more varied consistency. Cross-text stability is determined by term frequency; however, the specialized nature of many low-frequency terms impacts their final classifications: stable definitive specialized terms remain stable regardless of low frequency; unpredictable cross-domain polysemous terms are frequently charged with variability.

CONTEXTUAL DISTRIBUTION AND TYPICAL CASE ANALYSIS OF TRANSLATION VARIATIONS

Once it has performed the quantitative assessment of the consistencies of the translations it has now done a detailed examination at the micro-level and examined how the translated variants of specific terms have been distributed across the various domains. There will be a number of case study examples of the diverse types of terms that were discovered and categorised for the purposes of this paper. However, although the overall statistical evidence shows there to be a high level of diversity in translations, do it also see that the distribution of the various translation variants of the same terms are consistent? Or are they just random mixes of translated terms or are they systematically governed by some contextual or systematic constraints? Table 1 will show the specific frequency distributions of the main translated variants of three of these types of terms: "indemnity," "Premium," and "Force majeure," within five different sub-domains.

In Table 1, it can be seen that the translations of "indemnity" are distributed across many different contexts with great variability. The translation "peichang" for compensation has been found to be used 102 times with 85 of those uses occurring within the context of contracts, and 12 of those usages occurring in letters of credit. There are no usages of "peichang" recorded in the contexts of annual reports, correspondence, or e-commerce. On the other hand, "buchang" was found to be used a total of 30 times, with 12 usages recorded in annual reports and 4 usages recorded in correspondence. This usage distribution suggests that different translations for "indemnity" are not selected through random means but rather display a consistent relationship to specific contexts. For instance, "peichang" is predominantly used in the context of contract scenarios as a legal terminology, while "buchang" has been used in the context of information disclosure texts, such as annual reports and correspondence. The pattern is demonstrated again by the translation for "premium." The 88 occurrences of the term "premium" are almost exclusively within the domains of contracts and letters of credit (90.91%). Thus, it can be presumed there is a total dominance within the insurance industry. The term "high-quality"appears 45 times that 22 of which occurred within e-commerce and 15 in correspondence, thus offering "youjingpin" (high-quality) as the preferred term for marketing. The usage of "force majeure" is also consistently demonstrated by the 92 occurrences of the term "force majeure" of which 90.22% are contained within the domains of contracts and letters of credit. The findings of this research indicate that concepts that are clearly defined and are often used together may lead to greater consistency within translation standards. Standardization efforts for terminology management should avoid a mechanical pursuit of "one word, one translation" but focus on establishing and then building upon each dominant translation's contextual rules for the domain. At the same time, usage guidelines for polysemous terms when used across multiple domains should be developed using context-based factors.

TABLE 1. Register Distribution (Frequency) of Typical Term Translation Variants

Terminology	Translation variants	Contract	Letter of credit	Annual report	Correspondence	E-commerce	Total
Indemnity	Indemnity	85	12	3	2	0	102
	Compensation	8	5	12	4	1	30
	Compensation payment	4	2	1	0	0	7
	Exemption from liability	1	0	2	0	0	3
Premium	Premium	42	38	5	2	1	88
	High quality	0	0	8	15	22	45
	Additional cost	1	2	4	6	8	21
	Mark-up	0	0	2	3	10	15
Force majeure	Force majeure	45	38	6	3	0	92
	Uncontrollable situation	0	1	2	1	0	4
	Exemption clause	2	3	1	0	0	6
	Force majeure event	3	2	0	0	0	5

IV. CONCLUSION

This study investigates the semantic inconsistencies found in the translation of business English terminology, particularly where standardized technical lexicons intersect with the specialized nomenclature of high-precision fiber manufacturing and engineering. By analyzing these linguistic discrepancies, the research aims to establish a more rigorous translation framework that maintains conceptual integrity across global industrial trade and technical documentation. Based on a self-built English-Chinese parallel corpus, it systematically examines the actual situation and constraints of terminology translation consistency by comprehensively utilizing corpus linguistics methods and quantitative indicators. The consistency level of terminology translations is highly negatively correlated with the regulatory intensity of the text. In legally regulated texts, terminology translations exhibit extremely high stability: the average intratextual consistency coefficient of contract texts is as high as 0.96 (corresponding to an average diversity index of only 0.38), indicating that texts with strong legal force have strict inherent constraints on translation standardization. Conversely, in marketing-oriented texts, translation diversity increases significantly: the average diversity index of crossborder e-commerce description texts is as high as 1.19 (average intratextual consistency coefficient 0.871), while annual reports and business correspondence fall between the two (average diversity indices of 0.8 and 0.93, respectively). This "consistency gradient" from legal texts to marketing texts reveals the dynamic changes in the regulatory intensity in business translation practice. By validating the effectiveness of refined corpus design and hybrid screening strategies within the domain of high-performance fiber engineering, this study establishes a reusable analytical framework for the precise translation of technical terminology. This systematic approach ensures linguistic consistency across complex industrial datasets, facilitating the seamless integration of material science specifications into global academic and commercial discourse.

AUTHOR CONTRIBUTIONS

Yang Wang designed, collected and analyzed the data, and drafted the manuscript. Yang Wang conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Yang Wang participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] D. S. Dhivya, A. Hariharasudan, W. Ragmoun, and A. A. Alfalih, "ELSA as an education 4.0 tool for learning business English communication," Sustainability, vol. 15, no. 4, pp. 3809-3826, 2023, doi: 10.3390/su15043809.

[2] L. Huang, "The Ethical Choice of Business English Interpreters under Chesterman's Model of Translation Ethics," Journal of Literature and Art Studies, vol. 14, no. 9, pp. 808-814, 2024, doi: 10.17265/2159-5836/2024.09.010.

[3] X. Liu, "BUSINESS ENGLISH TRANSLATION STRATEGIES FROM THE PERSPECTIVE OF PSYCHOLOGY," Psychiatria Danu bina, vol. 33, no. Suppl 7, pp. 149-150, 2021.

[4] X. Ren, "We want but we can't: measuring EFL translation majors' intention to use ChatGPT in their translation practice," Humanities and Social Sciences Communications, vol. 12, no. 1, pp. 1-11, 2025, doi: 10.1057/s41599-025-04604-6.

[5] M. Alieva, Y. Godis, I. Lazurenko, H. Konopelkina, and O. Lytvynko, "The use of translation transformations in different styles of the English language for teaching written translation," Apuntes Universitarios, vol. 13, no. 3, pp. 92-104, 2023, doi: 10.17162/ au.v13i3.1526.

[6] H. Yin, "Balancing Accuracy and Authority in Legal Terminology Translation Under the Functional Equivalence Theory," Social Sciences and Humanities, vol. 3, no. 1, pp. 51-58, 2026, doi: 10.63313/SSH.9063.

[7] A. Boonmoh and I. Kulavichian, "A study of Thai EFL learners' problems with using online tools and dictionaries in Englishto-Thai translation," Kasetsart Journal of Social Sciences, vol. 44, no. 2, pp. 497-508, 2023, doi: 10.34044/j.kjss.2023.44.2.20.

[8] R. Inderawati, R. Hayati, R. Marlina, N. Novarita, A. Awalludin, and S. Anam, "Argumentative Essay and vocabulary enrich ment of English students by utilizing Google Translate," English community Journal, vol. 6, no. 2, pp. 131-141, 2023.

[9] A. A. Pratama, T. B. L. Ramadhan, F. N. Elawati, and R. A. Nugroho, "Translation quality analysis of cultural words in translated tourism promotional text of Central Java," Journal of English Language Teaching and Linguistics, vol. 6, no. 1, pp. 179-193, 2021, doi: 10.21462/jeltl.v6i1.515.

[10] M. Banat, "Investigating the linguistic fingerprint of GPT-4o in Arabic-to-English translation using stylometry," Journal of Translation and Language Studies, vol. 5, no. 3, pp. 65-83, 2024, doi: 10.48185/jtls.v5i3.1343.

[11] J. Luo and D. Li, "Universals in machine translation? A corpus-based study of Chinese-English translations by WeChat Translate," International Journal of Corpus Linguistics, vol. 27, no. 1, pp. 31-58, 2022, doi: 10.1075/ijcl.19127.luo.

[12] P. Giampieri, "Volcanic experiences: comparing non-corpus-based translations with corpus-based translations in translation training," Perspectives, vol. 29, no. 1, pp. 46-63, 2021, doi: 10.1080/0907676X.2019.1705361.

[13] X. Hu and M. Zheng, "A Corpus-Based Study on the Alternating Prototype Inhibition Effect in English Construction Trans lation," Foreign Languages, vol. 48, no. 4, pp. 95-104, 2025.

[14] K. Hu and X. Li, "The image of the Chinese government in the English translations of Report on the Work of the Government: a corpus-based study," Asia Pacific Translation and Intercultural Studies, vol. 9, no. 1, pp. 6-25, 2022, doi: 10.1080/23306343.2022.2066814.

[15] S. Granger and M. A. Lefer, "Learner translation corpora: Bridging the gap between learner corpus research and corpusbased translation studies," International Journal of Learner Corpus Research, vol. 9, no. 1, pp. 1-28, 2023, doi: 10.1075/ ijlcr.00032.gra.

[16] Y. W. Pian and W. Chen, "English translation of culture-loaded words-A corpus-based study," Journal of Literature and Art Studies, vol. 12, no. 6, pp. 667-673, 2022, doi: 10.17265/2159-5836/2022.06.012.

[17] Q. Li, R. Wu, and Y. Ng, "Developing culturally effective strategies for Chinese to English geotourism translation by corpusbased interdisciplinary translation analysis," Geoheritage, vol. 14, no. 1, pp. 6-29, 2022, doi: 10.1007/s12371-021-00616-1.

[18] L. Zhang, "A Study on Improving Semantic Consistency of Translation Systems by Combining Dynamic Computing Methods in English Corpus," J. Combin. Math. Combin. Comput, vol. 127, no. 1, pp. 3509-3525, 2025, doi: 10.61091/ jcmcc127b-196.

[19] X. Lin, M. Afzaal, and H. S. Aldayel, "Syntactic complexity in legal translated texts and the use of plain English: a corpusbased study," Humanities and social sciences communications, vol. 10, no. 1, pp. 17, 2023, doi: 10.1057/s41599-022-01485-x.

[20] M. Rikters, R. Ri, T. Li, and T. Nakazawa, "Japanese-English conversation parallel corpus for promoting context-aware ma chine translation research," Journal of Natural Language Processing, vol. 28, no. 2, pp. 380-403, 2021, doi: 10.5715/jnlp.28.380.

[21] F. S. Mauri, P. Sanchez-Gijon, and A. Oliver Gonzalez, "Cadlaws-An English-French parallel corpus of legally equivalent documents," Mutatis Mutandis: Revista Latinoamericana de Traduccion, vol. 14, no. 2, pp. 494-508, 2021, doi: 10.17533/ udea.mut.v14n2a10.

[22] P. Giampieri, "Is machine translation reliable in the legal field? A corpus-based critical comparative analysis for teaching ESP at tertiary level," Esp Today, vol. 11, no. 1, pp. 119-137, 2023, doi: 10.18485/esptoday.2023.11.1.6.

[23] T. Li and F. Pan, "Reshaping China's image: A corpus-based analysis of the English translation of Chinese political discourse," Perspectives, vol. 29, no. 3, pp. 354-370, 2021, doi: 10.1080/0907676X.2020.1727540.

[24] T. Haulai and J. Hussain, "Construction of mizo: English parallel corpus for machine translation," ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 22, no. 8, pp. 1-12, 2023, doi: 10.1145/3610404.

[25] L. Lowphansirikul, C. Polpanumas, A. T. Rutherford, and S. Nutanong, "A large English-Thai parallel corpus from the web and machine-generated text," Language Resources and Evaluation, vol. 56, no. 2, pp. 477-499, 2022, doi: 10.1007/ s10579-021-09536-6.

[26] R. R. Austin, C. L. Martin, R. C. Jones, S. C. Lu, R. Jantraporn, K. S. Martin, and K. A. Monsen, "Translation and validation of the Omaha System into English language simplified Omaha System terms," Kontakt, vol. 24, no. 1, pp. 48-54, 2022, doi: 10.32725/ kont.2022.007.

[27] M. M. Roshid, S. Webb, and R. Chowdhury, "English as a business lingua franca: A discursive analysis of business e-mails," International Journal of Business Communication, vol. 59, no. 1, pp. 83-103, 2022, doi: 10.1177/2329488418808040.

[28] M. Zhao and D. Li, "Translator positioning in characterisation: a corpus-based study of English translations of Luotuo Xiangzi," Perspectives, vol. 30, no. 6, pp. 1074-1096, 2022, doi: 10.1080/0907676X.2021.2000626.

[29] J. Duan, "A Corpus-Based Study on Modal Verbs in the Chinese-English Translation of the Book of Contracts in Chinese Civil Code," Theory and Practice in Language Studies, vol. 15, no. 1, pp. 262-271, 2025, doi: 10.17507/tpls.1501.29.

[30] H. S. Mahdi and Y. Sahari, "A corpus-based study of translating idioms from English into Arabic using audio-visual translation," The International Journal of Information and Learning Technology, vol. 41, no. 3, pp. 244-261, 2024, doi: 10.1108/ IJILT-07-2023-0128.

[31] A. Zhang and X. Zhu, "Analysis of English translation of corpus based on blockchain," International Journal of Web-Based Learning and Teaching Technologies (IJWLTT), vol. 18, no. 2, pp. 1-14, 2023, doi: 10.4018/IJWLTT.332767.

[32] X. Han and Y. Ran, "Intelligent recognition English translation model based on speech recognition," International Journal of Computational Systems Engineering, vol. 10, no. 1-4, pp. 90-103, 2026, doi: 10.1504/IJCSYSE.2026.151338.

[33] J. Yang, N. Husin, and A. M. Yusof, "Exploring the Consistency between Translation Style Attitudes and Practices," Interna tional Journal of Language Education and Applied Linguistics, vol. 15, no. 1, pp. 40-51, 2025, doi: 10.15282/ijleal.v15i1.11711.

[34] C. Mao, X. Gao, Z. Yu, Z. Wang, S. Gao, and Z. Man, "Bilingual parallel sentence pair extraction under structural feature consistency constraints," Journal of Chongqing University, vol. 44, no. 1, pp. 46-56, 2021, doi: 10.1145/3465740.

[35] L. Li and C. Li, "A Corpus-Based Study of Commonly Used Business English Vocabulary," Foreign Language Journal, vol. 4, no. 1, pp. 64-69, 2021.