ISSN: 2688-8653

Construction and Analysis of an Intelligent Bilingual Corpus for the International Marketing of Sichuan-Chongqing Specialties

Xuemei Feng[1,*], Tianqiang Ye[1], Ying Yan[1], and Ruifeng Zhou[2]

[1] School of Foreign Languages, Pass College of Chongqing Technology and Business University, Chongqing, 401520, China

[2] School of Computer and Internet of Things, Chongqing Institute of Engineering, Chongqing 400056, China

[*]Corresponding Author: 2606764578@qq.com

Abstract

In the context of cross-border e-commerce, the core bottleneck restricting the international dissemination of Bashu characteristic products lies in the quality of marketing text translation. Existing research lacks a systematic framework integrating "vertical corpus construction – human-AI collaboration – effectiveness validation." To address this gap, the present study adopts a "problem-oriented → resource construction → application validation" pathway. The research is structured into three phases. First, corpus construction involves collecting data from authoritative sources in Sichuan and Chongqing covering eight product categories. After cleaning and standardization, translations are generated through a hybrid approach: matching existing authoritative translations or using AI-generated drafts reviewed by TEM-8-certified translators. Bilingual alignment is achieved via a "tool-assisted + manual verification" process, with sampling accuracy exceeding 99%. Second, system development leverages KIMI AI to generate code for the "Bashu Treasure Repository" corpus system, which supports real-time querying and pagination. Third, the study ultimately constructs a corpus comprising 2,956 bilingual texts across 739 products from eight categories and proposes an "AI draft translation → corpus-based refinement" workflow. Theoretically, it lays a foundation for translation studies in the internationalization of regional products. Practically, it offers an integrated solution combining linguistic resources, technical tools, and workflow optimization.

Keywords

Bashu Characteristic Products; International Marketing; Bilingual Corpus; Translation Quality.

1. Introduction

Against the backdrop of deepening global digital economic integration, cross-border e-commerce has emerged as a pivotal channel for Bashu specialty products-such as geographical indication products like "涪陵榨菜 (Fuling pickled mustard)" and "巫溪板角山羊 (Plate-horned Goat of Wuxi)" to transcend geographical constraints and expand into international markets. As the direct carrier for conveying product value and constructing brand perception, the translation quality of marketing texts is no longer merely a matter of linguistic conversion. It has evolved into a strategic factor determining the effectiveness of international communication and influencing consumer purchasing decisions.

From the perspective of research evolution, academic focus on cross-border e-commerce translation has shifted from "basic linguistic norms" to "deep cultural adaptation". [12] Early

ISSN: 2688-8653

studies predominantly addressed surface-level issues such as spelling errors in titles and syntactic disorganization (Zhao Hongdan, 2019). [13] Current research, however, places greater emphasis on the cultural adaptability of translated texts-for instance, how linguistic adjustments can align with target market consumption habits and stimulate marketing resonance (Zhao Hongdan, 2025). [2] This shift clearly reveals the limitations of generic translation tools and methods: they struggle to accommodate the unique attributes of "geographical indication + cultural symbols" prevalent in Bashu product marketing texts (e.g., the historical connotations of "贡米 (Tribute Rice)" or the regional botanical reference of "臭黄荆 (Vitex negundo)" (Feng Zhiwei, 2023). This often leads to the dissipation of cultural information or terminological inaccuracies, consequently undermining the products' international competitiveness. [7] In cross-border e-commerce, translation quality is no longer merely a matter of language conversion, but a key strategic element that directly affects product competitiveness and consumer perception (Tang, W., & Li, G., 2024).

To address these translation challenges in this vertical domain, corpus-based translation studies provide crucial methodological support. [1] Existing research confirms that domain-specific corpora, by aggregating linguistic data from specialized contexts, can significantly enhance terminology consistency and the normative rendering of culture-specific items (Baker, 1993). [5] For example, in translating promotional materials for Chinese Time-Honored Brands, specialized corpora have proven effective in avoiding mistranslations of culture-loaded terms (Li Dechao, 2020). Nevertheless, a bilingual corpus dedicated specifically to Bashu specialty products remains absent. Most existing corpora cover general e-commerce categories or broad cultural domains, failing to accurately match the textual characteristics of Bashu products—such as complex craft narratives (e.g., intangible cultural heritage tea-making skills) and unique authoritative certification systems (e.g., National Geographical Indication). [8] Consequently, they cannot provide targeted support for translation practices in this field (Takezawa, T., et al. ,2002).

Concurrently, the paradigm shift in translation technology further underscores the urgency of constructing a specialized corpus. While the prevailing translation model has transitioned from computer-aided translation to AI-aided translation, neural machine translation and large language models still exhibit shortcomings when processing texts related to regional specialty products. These include terminology "generalization" (e.g., uniformly translating different varieties of "贡米" as "Gong rice", obscuring varietal distinctions) and syntactic "rigidity" (e.g., literally translating Chinese run-on sentences leading to logical confusion) (Feng Zhiwei, 2023). [9] There is a pressing need to optimize model outputs using high-quality, vertical-domain corpora to establish an efficient human-AI collaborative workflow characterized by "AI draft translation → corpus-based refinement" (O'Brien et al., 2024).

Based on this rationale, this study aims to fill the aforementioned research gap by constructing an intelligent bilingual corpus for the international marketing of Bashu specialty products. It seeks to integrate corpus resource development, application system implementation, and translation effectiveness validation, thereby providing an integrated solution-from "resources" to "tools" to "workflows"-for translation practices in this domain. Ultimately, this research aims to facilitate precise cultural transfer and effective brand communication for Bashu specialty products in the global marketplace.

2. Corpus Collection and Processing Scheme

The quality of a corpus directly determines its application value, particularly for vertical-domain marketing corpora, which must simultaneously meet three requirements: "accurate terminology, complete cultural information, and clear marketing intent" (Li Dechao, 2020). To this end, this study developed a comprehensive workflow of "multi-source collection → layered

ISSN: 2688-8653

processing → dual quality control" to ensure the corpus's authority, accuracy, and scenario adaptability."

2.1. Source Dealing

2.1.1. Corpus Sources and Collection Criteria

[11] Corpus selection strictly adhered to the principles of "relevance, accuracy, and diversity" (Zhang Yujie, 2021), with specific operations as follows:

Source Authority: Core corpus was obtained from official channels in Sichuan and Chongqing, including the official websites of the Sichuan Provincial Government (http://www.sc.gov.cn) and Chongqing Municipal Government (http://www.cq.gov.cn), the "Geographical Indication Product Protection Announcements" issued by their respective Market Supervision and Administration Bureaus, and industry reports from Departments or Bureaus of Agriculture and Rural Affairs, ensuring textual content is administratively reviewed and legally valid. Supplementary corpus came from cross-border e-commerce promotional materials of the China Council for the Promotion of International Trade (CCPIT) Sichuan Sub-council, balancing practical needs of marketing scenarios.

Content Relevance: Focused on text types "directly serving international marketing," covering three categories: core product information (e.g., fermentation process of "郫县豆瓣 (Pixian Bean Paste)," brewing history of "泸州老窖 (Luzhou Laojiao Tequ Baijiu)"), policy support (e.g., export tax rebate policies), and brand stories (e.g., intangible cultural heritage inheritance of "蜀绣 (Shu Embroidery)", excluding purely internal management documents (e.g., production safety regulations).

Type Diversity: Encompassed eight major categories of Bashu characteristic products, including Geographical Indication foods: "涪陵榨菜 (Fuling pickled mustard)" and "宜宾芽菜 (Yibin Sprout Vegetable)"; intangible cultural heritage crafts: "蜀绣(Shu Embroidery)" and "竹编（Bamboo Weaving）"; characteristic agricultural products: "苍溪猕猴桃(Cangxi Kiwifruit)" and "奉节脐橙(Fengjie Navel Orange) "; and Chinese medicinal materials: "川贝母(Sichuan Fritillaria)" and " 黄 连 (Chinese Goldthread)", ensuring the corpus reflects textual characteristics of different products (e.g., food emphasizing "taste description, " crafts emphasizing "technique narrative").

Collection and Verification Mechanism: Raw corpus required "cross-verification across two platforms" - information for the same product had to match consistently across at least two authoritative channels (e.g., "峨眉山茶 (Emei Mountain Tea)" must appear on both the Sichuan Department of Agriculture and Rural Affairs website and the Emeishan Municipal Government website) before being entered into the raw corpus dataset, eliminating single-source information to avoid bias.

2.1.2. Corpus Cleaning and Bilingual Alignment Methodology

Raw corpus underwent a three-stage processing pipeline of "data cleaning → translation generation → bilingual alignment," combining technical tools and manual intervention to achieve standardization and usability.

(1) Data Cleaning and Standardization

Addressing common format inconsistencies in web texts using a "technical automation + manual sampling verification" model:

Technical Layer: Used Python's Pandas library for data deduplication (based on product name and publication time), and the BeautifulSoup library for batch removal of HTML tags (e.g., <p>,), redundant advertisement text (e.g., "click to buy"), and special characters (e.g., emojis, garbled symbols). All text was uniformly converted to UTF-8 encoding to ensure Chinese-English character compatibility. Long sentences (exceeding 30 characters) were reasonably

ISSN: 2688-8653

segmented (based on punctuation and semantic logic) to lay the foundation for subsequent sentence-level alignment.

Quality Control: Randomly sampled 5% of cleaned texts to check format standardization (e.g., no residual tags) and semantic integrity (e.g., no loss of key information due to segmentation). A pass rate of ≥99% was required to proceed to the next stage.

(2) Translation Acquisition and Generation

A three-tier translation generation mechanism, "authoritative priority, AI-assisted, human final review," was adopted to balance translation authority and coverage:

Priority Matching of Authoritative Translations: Official translations for corresponding products (e.g., "自贡灯会" for "Zigong Lantern Festival") were directly extracted from bilingual white papers published by the Chongqing and Sichuan Foreign Affairs Offices, and English manuals from the CCPIT Sichuan Sub-council for cross-border e-commerce. Such translations accounted for approximately 35% of the corpus.

[6] AI Drafting + Human Refinement: For texts without authoritative translations, initial drafts were generated using DeepL (emphasizing business text fluency) and ChatGPT (emphasizing cultural term explanation). These were then comparatively reviewed by two translators with TEM-8 certification and over five years of cross-border e-commerce translation experience. (Malkawi, M. A. K. M., et al., 2025). [10] This review strictly followed Vermeer's Skopos Theory (Vermeer, 1983): centering on "achieving the persuasive function of international marketing" (Skopos rule). [3] In doing so, ensuring logical coherence in the target language context (Coherence rule), while remaining faithful to the original's cultural characteristics (Fidelity rule) (Du, X., 2012).

Specific Application of Foreignization Strategy: For Bashu culture-specific items, a combination of "literal translation + transliteration + amplification" was used - e.g., "贡米" translated as "tribute rice (a type of rice historically offered to emperors)" (literal + amplifications), "臭黄荆" rendered as "Vitex negundo (a unique herbal tea in Sichuan)" (transliteration + amplification), preserving cultural information while ensuring target reader comprehension.

Dispute Resolution Mechanism: In case of disagreement between translators (e.g., translation of "川剧变脸"), a proofreader with a Ph.D. in Translation Studies made the final decision based on the principle of "maximizing marketing effectiveness." The final translation underwent secondary review, ensuring an error rate ≤ 2%, a threshold considered acceptable in professional translation workflows (O'Brien et al., 2024).

(3) Bilingual Alignment

A high-precision alignment scheme of "tool-assisted initial alignment + manual bidirectional verification" was employed:

Tool Assistance: ABBYY Aligner software was used for automatic alignment based on lexical overlap and sentence structure, achieving paragraph-level alignment (accuracy ~85%) and sentence-level alignment (accuracy ~70%).

Manual Verification: Two translators performed "forward (Chinese to English)" and "reverse (English to Chinese)" checks respectively, focusing on correcting issues like "one sentence with multiple translations" and "missing translation alignment" (e.g., properly aligning "蜀绣的针法有122种" with its English counterpart).

Quality Acceptance: 5% of the aligned corpus (covering all eight categories) was randomly sampled to calculate alignment accuracy (Formula: Correctly aligned sentence pairs / Total aligned sentence pairs × 100%). A rate of ≥95% was required for acceptance. The final corpus was stored in Excel format with fields: "Product ID - Chinese Source Text - English Translation - Text Type - Source," facilitating system retrieval.

ISSN: 2688-8653

This workflow, through a synergistic model of "technical efficiency enhancement + human quality assurance," addresses the inefficiency of traditional manual processing while avoiding cultural information distortion inherent in purely technical approaches, providing a high-quality data foundation for subsequent corpus system development and translation feature analysis.

2.2. Construction Principles and Technical Implementation of the Corpus System

The development of the corpus system's user end was demand-driven, utilizing AI-assisted programming tools to achieve efficient development. The specific implementation workflow was as follows:

First, based on the corpus structure designed in section 2.1.2, we organized the corpus data comprising four core fields-Chinese Name, English Name, Chinese Introduction, and English Introduction (corresponding to columns B to E in the spreadsheet)-and stored it in Excel format to ensure structured data organization. For details, see Figure 1.

A	B	C	D	
1	石柱长毛兔	Shizhu Long-haired Rabbit	石柱长毛兔是一种具有	Shizhu long-haired rabbit is a
2	巴南银针	Banan Silver Needle tea	巴南银针茶是以重庆市	Banan Silver Needle tea is a h
3	巫溪板角山羊	Plate-horned Goat of Wuxi	巫溪板角山羊主产于巫	Plate-horned Goat of Wuxi is
4	金佛山中华蜂蜜	Jinfoshan Chinese Honey	金佛山中华蜜蜂是我国	Jinfoshan Chinese honey bee i
5	云阳泥溪黑木耳	Yunyang Nixi Black Fungus	云阳泥溪黑木耳，重庆	Yanyang Nixi Black Fungus is
6	秀山毛尖	Xiushan Maojian tea	秀山毛尖采用传统炒青	Xiushan Maojian tea is produc
7	鸡鸣贡米	Jiming Tribute Rice	鸡鸣贡米是云阳县大阳	Jiming tribute rice is a famous
8	静观腊梅	Jingguan wintersweet,	静观腊梅，重庆市北碚	Jingguan wintersweet, special
9	马喇贡米	Mala Tribute Rice	马喇湖贡米是重庆市黔	Malahu Tribute Rice is a spec
10	花田贡米	Huatian Tribute Rice	花田贡米脆酥油糯，冲	Huatian Tribute Rice is crispy
11	巫溪大宁河鸡	Wuxi Daning River Chicken	巫溪大宁河鸡是原产于	Wuxi Daning River Chicken i
12	武隆板角山羊	Wulong Plate-horned Goats	武隆板角山羊主产于区	Wulong Plate-horned Goats a
13	酉阳麻旺鸭	Youyang Mawang Duck	酉阳麻旺鸭，因原产区	Youyang Mawang Duck is nat
14	万州山胡椒	Wanzhou Mountain Pepper	万州山胡椒是我国优良	Wanzhou mountain pepper is
15	万州武陵桂圆	Wanzhou Wuling cinnamon	万州武陵桂圆是重庆市	Wanzhou Wuling cinnamon is

Figure 1. Screenshot of the original corpus in Excel

Subsequently, code for the user interface and core functionalities was generated using the KIMI AI tool. The development instructions specified six key requirements: 1) A basic interface containing a search box, query button, and results list; 2) A fuzzy search function supporting queries by Chinese name, returning complete entry information; 3) Real-time search effects, dynamically updating results as the user types; 4) Adoption of a mainstream UI framework to ensure an aesthetically pleasing interface; 5) Loading corpus data by reading an Excel file to avoid hardcoding data into the application; and 6) Implementation of pagination, displaying 10 results per page, with page navigation showing the current page and two preceding/succeeding pages, representing out-of-range pages with ellipses.

Based on these requirements, the KIMI AI tool automatically generated the code for a single-page application using HTML and JavaScript. It employed Tailwind CSS as the UI framework for responsive design and integrated the Sheet JS library to handle reading and parsing the Excel file. The generated code implements three core functional modules: the File Parsing Module, responsible for loading and parsing the Excel corpus data to build an in-memory database; the Search Module, which implements a real-time fuzzy matching algorithm supporting dynamic result updates during input; and the Display Module, which manages the logic for paginating search results and rendering page navigation.

The finalized corpus user end possesses the following characteristics: a clean and intuitive interface conforming to modern UI design standards; rapid interactive response and a smooth search experience; a flexible data loading method supporting dynamic corpus updates via local

ISSN: 2688-8653

Excel files; and clear pagination logic facilitating user navigation through large result sets. This implementation successfully translates the design concept into a functional user end, laying a foundation for the subsequent application and potential expansion of the corpus. For details, see Figure 2, which is a screenshot of the corpus during use.

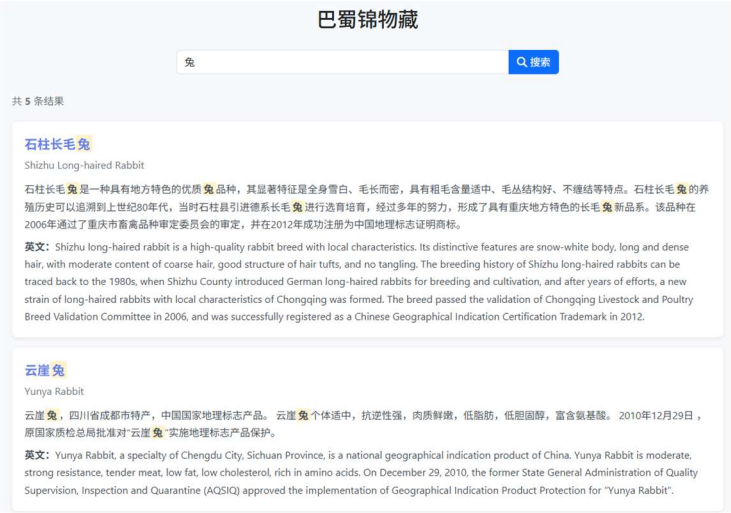

Figure 2. Corpus Screenshot

3. An Analysis of the Linguistic Features and Translation Challenges in Marketing Texts for Bashu Specialty Products

As a distinct genre of regional marketing discourse, the language of Bashu specialty product marketing texts is deeply rooted in a unique local culture and directly impacts the efficacy of cross-cultural communication. This chapter systematically analyzes its linguistic features across four dimensions-nomenclature mechanisms, craft narration, culture-loaded terms, and the construction of authority-while deconstructing the core challenges these features pose to translation practice. [4] Current research places greater emphasis on the "deep cultural adaptation" of translations, which refers to the process of adapting translations to the cultural context of the target market (Kim, Y. Y., 2017).

3.1. Nomenclature Mechanisms: Semantic Condensation of Geographical Indications and Varietal Information, and the Problem of Decoding

Product naming acts as the "meta-discourse" of marketing texts. The nomenclature of Bashu specialty products exhibits a high degree of semantic condensation and dependence on geographical indications. Its typical structure is "Precise Geographical Name + Product Category/Varietal Characteristic" (e.g., "涪陵榨菜 (Fuling pickled mustard)" and "巫溪板角山羊 (Plate-horned Goat of Wuxi)"). This naming strategy not only asserts the product's inseparability from its place of origin (i.e., its terroir or "sense of place") but also implicitly guarantees quality through the "halo effect" of the geographical name. To further enhance cultural added value, names often incorporate historical symbols such as "贡" (e.g., "Mala Tribute Rice", "Jiming Tribute Rice") and "老" (e.g., "Jiangjin Baijiu"), transforming the name into a vessel of local historical memory.

However, this highly condensed and culturally embedded naming mechanism presents structural challenges for translation. The first challenge is the lack of standardized equivalents for proper nouns. Terms like "大宁河" (Daning River) and "九叶青" (Jiuyeqing) lack widely accepted translations, easily leading to informational emptiness after transliteration or deviation in meaning-based translation. The second challenge is the dissipation of cultural

ISSN: 2688-8653

symbols. For instance, the historical evaluative framework of "皇家贡品" (imperial tribute) embedded in the character "贡" is difficult to convey equivalently to target-language readers when simply translated as "Tribute," resulting in a loss of the underlying historical and cultural context.

3.2. Craft Narration: Information Density Arising from Technicality, ICH Status, and Run-On Sentence Structures

Descriptions of production techniques blend technical specifications with narratives of Intangible Cultural Heritage (ICH) status, creating information-dense discourse. For example, the text for " 巴南银针(Banan Silver Needle Tea)" specifies raw material standards like "one bud and one leaf picked before Qingming Festival" and craftsmanship heritage such as "non-legacy techniques passed down through four generations"; "秀山毛尖 (Xiushan Maojian Tea)" highlights scarcity with quantified descriptions like "50,000 pickings yield just 1 jin of dry tea." Syntactically, such texts frequently employ run-on sentences, presenting multiple points about selection, processing, and characteristics in a linear sequence, aligning with Chinese's paratactic nature but resulting in high information density.

The difficulties in translation conversion lie in: 1) Equivalency of technical terminology: Terms like "炒青 (pan-firing to halt oxidation)" and "破壁 (cell wall disruption)" require finding the closest professional equivalent in the target language without distorting the original meaning. 2) Explanatory translation of ICH concepts: "非物质文化遗产" cannot be merely translated as "Intangible Cultural Heritage"; it often requires brief explanation of its "transgenerational transmission" characteristic to avoid vagueness. 3) Explicitation of logic in run-on sentences: The implicit logical relationships (e.g., condition, purpose) need to be made explicit and reorganized within the hypotactic framework of English grammar to prevent informational confusion.

3.3. Culture-Loaded Terms: The Dilemma of Context Recontextualization in Historical Intertextuality and Literary Allusions

Marketing texts extensively use culture-loaded terms to construct historical depth and cultural sophistication. Primary methods include intertextual references to historical documents and allusions to literature/folklore. "云阳龙缸云雾茶 (Yunyang Longgang Misty Tea)" cites the "Classic of Tea" and "Supplements to the National History of the Tang Dynasty", while "涪陵百花露酒(Baijiu Fermented with Myriad Flowers)" traces its origin to its use as an imperial tribute during the Jiaqing era; this constitutes an authoritative intertextuality based on historical texts. Conversely, the naming legend of "巫山翡翠凉粉(Wushan Jade Jelly)" and the association of "忠县花粉酒 (Zhongxian Pollen Wine)" with the poet Bai Juyi are literary allusions aimed at adding narrative interest and emotional connection.

These culture-loaded terms represent the "hard nuts to crack" in translation. The core challenge is the absence of cultural context. For example, literally translating "臭黄荆" as "Stinky Vitex" conveys only the literal meaning, completely losing its referential meaning as a specific plant and potential local cultural connotations. The solution often requires striking a balance between foreignization and domestication strategies, or employing methods like literal translation with annotation, to maximize the transfer of cultural imagery.

3.4. Authority Construction: Cross-Cultural Conversion of Credibility in Quantitative Data and Certification Systems

To establish objective credibility, Bashu marketing texts construct a multifaceted discourse of authoritative certification. One aspect is scientific quantitative endorsement, using specific data like "amino acid content of 3.10%" or "gastrodin content 8.2 times the national standard" to

ISSN: 2688-8653

establish the product's superiority in composition. The other is the citation of institutional certifications, such as "National Geographical Indication Product," "China Well-Known Trademark," and "Organic Product Certification," which carry strong signaling power within the Chinese market context.

The translation challenge lies in the potential diminution of this authority in an international context. Firstly, the asymmetry between Chinese and foreign certification systems means concepts like "China Well-Known Trademark" may lack equivalent perceived authority among foreign consumers. Secondly, differences in measurement units and standards (e.g., converting "亩" [mu] and "斤" [jin]) can create comprehension barriers. Therefore, translation often necessitates providing brief contextual background for the certification systems or seeking functionally equivalent international certification terms for analogy, to achieve an effective conversion of credibility.

The features of Bashu specialty product marketing texts are a direct reflection of their unique regional culture. This very nature dictates that their translation is far from a simple linguistic transfer; it constitutes a complex cross-cultural practice involving cultural decoding, context recontextualization, and authority reconstruction. Subsequent research on translation strategies must be grounded in a precise understanding of these deep-seated features to formulate approaches that remain faithful to the original's cultural essence while achieving effective communication in the target context.

4. Conclusion and Future Research Directions

This study, situated within the context of the rapidly evolving global digital economy and cross-border e-commerce, addresses a critical bottleneck in the international dissemination of Bashu characteristic products: the quality of marketing text translation. It presents a systematic investigation, spanning theoretical exploration to practical application, following a progressive "problem-oriented → resource construction → application validation" pathway. The research successfully designed and constructed a vertical-domain intelligent bilingual corpus named "Bashu Treasure Repository." This corpus encompasses 2,956 high-quality bilingual texts covering 739 products across eight categories. An application system featuring real-time retrieval and a user-friendly interface was developed using modern front-end technologies, thereby transforming static linguistic resources into a dynamic, usable translation aid tool. Methodologically, the study established a rigorous and transparent pipeline for corpus collection, cleaning, alignment, and annotation, ensuring the authority and accuracy of the linguistic assets. It innovatively proposed a human-AI collaborative workflow termed "AI draft translation → corpus-based refinement." Through controlled experiments, the research empirically validated the significant efficacy of the corpus and its associated workflow in enhancing translation quality across four dimensions: translation efficiency, terminology consistency, textual fluency, and AI optimization capability. The results demonstrated not only a reduction in time consumption and error rates for human translation but also an effective mitigation of the inherent limitations of free AI tools, such as terminology generalization and inadequate handling of culture-loaded terms.

On a theoretical level, this study conducted an in-depth analysis of the profound features of Bashu product marketing texts. It systematically elucidated, from multiple dimensions including nomenclature mechanisms, craft narration, culture-loaded terms, and the construction of authority, how their regional cultural uniqueness translates into specific challenges in cross-lingual communication. This analysis not only provided a theoretical foundation for the corpus construction in this study but also laid a solid scholarly groundwork for research on translation strategies for the internationalization of regional specialty products. Overall, this study provides an integrated solution for the international marketing of Bashu and

ISSN: 2688-8653

other regional specialty products, combining linguistic resources, technological tools, and practical workflows, possessing both significant theoretical value and broad practical implications.

Despite the achievements outlined above, this study has ample room for further deepening and expansion in several aspects, pointing to clear directions for future research. First, regarding the evolution of the corpus itself, future efforts should aim to transform it from a static repository into a dynamic, self-evolving intelligent entity. This entails exploring dynamic updating and expansion pathways for the corpus based on incremental learning and crowdsourcing mechanisms, enabling it to automatically capture and integrate emerging marketing texts and user reviews generated in real-time on e-commerce platforms, thereby ensuring the timeliness and vitality of the corpus. Simultaneously, introducing large language models for automated corpus quality assessment and enhancement will be a key technological direction for improving the scale and quality of the corpus.

Second, concern information alignment and explore automated marketing content (e.g., short video scripts) generation technology based on a multimodal corpus. On the other hand, to address the personalized translation needs of different product categories, target markets, or specific merchants, research can move towards corpus-based personalized model fine-tuning or adaptive retrieval technologies, achieving an upgrade from a "generic tool" to a "personalized intelligent translation advisor." At the theoretical level, further integrating interdisciplinary theories from computational communication studies and cognitive linguistics, and empirically testing the cognitive and affective impacts of different translation strategies on target consumers using methods such as eye-tracking and EEG, will propel the evaluation of translation effectiveness from "linguistic equivalence" towards "communicative efficacy," thereby constructing a more explanatory and guiding theoretical framework.

Acknowledgments

This work was supported by the Pass College of Chongqing Technology and Business University [Grant Number KY202420]. The project title is "The Construction and Application of an Intelligent Corpus: International Marketing of Bashu Characteristic Products Based on Data Mining."

References

[1] Baker, M. (1993). Corpus linguistics and translation studies: Implications and applications. In M. Baker, G. Francis, & E. Tognini-Bonelli (Eds.), *Text and technology: In Honour of John Sinclair* (pp. 233–250). John Benjamins.

[2] Feng, Z. W. (2023). Current status and challenges of neural machine translation. *Chinese Translators Journal*, 44(5), 5–15.

[3] Du, X. (2012). A brief introduction of Skopos theory. *Theory and Practice in Language Studies*, *2*(10), 2189-2193.

[4] Kim, Y. Y. (2017). Cross-cultural adaptation. *Oxford research encyclopedia of communication*.

[5] Li, D. C. (2020). A corpus-based study on the external publicity translation of Chinese time-honored brands. *Shanghai Journal of Translators*, (4), 68–73.

[6] Malkawi, M. A. K. M., Altalahin, A., Abushunar, M. S. H., & Aldelaa, A. S. (2025). AI-Aided Translation: An Eye on ChatGPT as a Tool of Translating Literary Texts. *Technium Soc. Sci. J.*, *75*, 311.

[7] Tang, W., & Li, G. (2024). Enhancing competitiveness in cross-border e-commerce through knowledge-based consumer perception theory: an exploration of translation ability. *Journal of the Knowledge Economy*, *15*(3), 14935-14968.

ISSN: 2688-8653

[8] Takezawa, T., Sumita, E., Sugaya, F., Yamamoto, H., & Yamamoto, S. (2002, May). Toward a Broad-coverage Bilingual Corpus for Speech Translation of Travel Conversations in the Real World. In *LREC* (pp. 147-152).

[9] O'Brien, S., Moorkens, J., & Castilho, S. (2024). The impact of large language models on the translation profession: A meta-analysis. *The Journal of Internationalization and Localization*, 11(1), 1–25.

[10] Vermeer, H. J. (1983). *Skopos und kommission in der translatorialen handlung* [Skopos and commission in translational action]. Linguistische Arbeiten, 147, 173–187.

[11] Zhang, Y. J. (2021). A corpus-based study of the translation of Chinese cultural-specific items in food advertising. Springer.

[12] Zhao, H. D. (2019). Human-computer interactive translation of cross-border e-commerce product copy. *Chinese Science & Technology Translators Journal*, 32(1), 26–28.

[13] Zhao, H. D. (2025, June 30). Cross-cultural real-time translation strategies in cross-border e-commerce live broadcasts. *Tonghua Daily*, p. 4.