

Research on the Application of Tourism English Corpus in ESP Teaching at Higher Vocational Colleges

Jianzhou Cui

Wuxi City College of Vocational Technology, Wuxi, Jiangsu, China

Abstract: Modern information technology has brought revolutionary changes to education development, and corpus is one of the core innovations in this field. It enjoys unique advantages including unprecedented massive information storage, high-speed accurate computer retrieval, and distinct concordance interface. At present, the popularization rate of information technology in China's higher education sector is quite high, but progress is still relatively lagging in constructing specialized tourism corpora and leveraging this platform to improve tourism English teaching in higher vocational colleges (HVCs). Through literature analysis and case study, this paper explores the application of corpus in teaching tourism English vocabulary, oral English, tour guide commentary creation, translation, and intercultural communication in HVCs. This study is a valid attempt to reform the existing teaching content design that purely relies on textbooks and the teacher-centered lecture-based teaching mode. After further optimization, this approach can be promoted in relevant disciplines of HVCs across the city and even the whole province. Implementing corpus-based teaching for HVC students can also promote students' exploratory learning and creative learning, so as to effectively improve the effect of tourism English teaching in HVCs.

Keywords: Tourism English; Corpus; ESP Teaching; Higher Vocational Education; Teaching Reform

1. Introduction

So far, the classic specialized tourism corpora developed abroad include: the Kyoto Tourism Corpus with 42,025 tokens developed by Kiyomi et al. from Nihon University (Japan) in 2006, which aims to summarize hundreds of the most frequently used tourism English vocabulary suitable for Japan's tourism market (especially the Kyoto region) for learners of different levels

taking tourism English courses in Japanese universities; a 37,795-token English tourism text corpus developed by Carlota Alcantar, a graduate student from the University of Essex (UK) in 2007, to help teachers from the School of Tourism, Nayarit State University (Mexico) teach common tourism vocabulary more effectively; a 670,000-token English tourism text corpus developed by Michael Wilkinson, an English teacher from the Savonlinna School of Translation Studies (Finland), which is mainly used as a reference for Finnish students when translating Finnish tourism texts into English to check the idiomaticity of their wording; and a 1-million-token subcorpus of tourism brochures included in the American National Corpus, covering 179 independent texts. To sum up, these corpora are mainly developed for tourism English translation, oral English and vocabulary teaching.

In recent years, Chen and Liu [1] constructed a PBL teaching mode of "corpus retrieval - problem exploration - project output" based on a self-built parallel corpus of Chinese tourism publicity translation. Zhao and Zhang [2] built a multimodally annotated tourism interpreting corpus covering audio, video, images and on-site scenarios, and applied it to consecutive interpreting training. Hidayat et al. [3] developed blended teaching modules based on open tourism corpora, and verified that this mode significantly improves the workplace vocabulary competence of EFL tourism students in Southeast Asia. Zhang and Wang [4] built a balanced annotated tourism Chinese corpus covering different difficulty levels and scenarios, and developed a hierarchical teaching design. Wang [5] targeted the pain point of outdated corpus in HVC tourism English textbooks, integrated the COCA tourism subcorpus and a self-built tour guide scenario corpus, and developed contextualized teaching resources. Li [6] designed inquiry-based teaching based on a parallel corpus of Chinese and English tourism publicity materials, allowing students to

independently explore cultural differences in discourses. Huang [7] constructed an emergency scenario corpus for tourism services combined with large model fine-tuning technology, and connected it to a virtual reality teaching system. Based on the review of domestic and international research, the current research progress can be summarized as three characteristics: First, in terms of research direction, relevant studies have expanded from the early focus on tourism English vocabulary teaching to multiple fields such as tourism translation, interpreting, intercultural communication, tourism Chinese, and professional competence development. Second, in terms of technology application, technological integration has been deepening continuously, from the shallow application of pure text corpora in the early stage to the integration of multimodal corpora, large model fine-tuned corpora with smart teaching and virtual reality teaching. Third, in terms of teaching promotion, most studies have broken through the shallow application of “using corpora to look up new words”, and formed mature corpus-driven teaching modes such as project-based teaching, inquiry-based teaching and situational teaching.

2. Theoretical Basis of Applying Corpus to ESP Teaching in HVCs

2.1 Constructivist Theory

Constructivism holds that learning is essentially a process of active exploration, active discovery and active construction of knowledge by students; learning is a meaning construction process completed through interpersonal collaborative activities in specific contexts with necessary learning materials; an ideal learning environment includes four core components: context, collaboration, communication and meaning construction. Corpus can present real language materials, which is very conducive to creating the “context” required for language learning. If corpus-based teaching can be applied to tourism English education in HVCs, it will promote students’ exploratory learning and creative learning, and further improve teaching effects.

2.2 Schema Theory

Schema Theory holds that language users usually gradually form abstract language knowledge and language competence by

repeatedly using similar language forms in similar contexts. The unique concordance interface of corpus meets the basic requirements of Schema Theory, and can provide learners with the possibility of repeated exposure to “similar contexts”. This gives corpus introduction into language teaching unique educational functions. With the help of corpus, learners observe and analyze language information, which greatly promotes their cognitive development. When processing information, students repeatedly use similar language forms to construct new language schemata. In this process, students use their existing “prototype knowledge” to construct new schemata, which can simplify the teaching process and strengthen memory effects. He [8] pointed out that the unique concordance interface of corpus can repeatedly present the same linguistic phenomenon in different contexts among a large number of relatively homogeneous corpora, so as to help learners construct cognitive schemas of language.

3. Application of Tourism English Corpus in HVC Tourism English Teaching

3.1 Typical Linguistic Features and Tourism English Vocabulary Teaching

The frequency of English words in discourse is unbalanced but regular. According to COBUILD research, the top 700 most frequent English words account for about 70% of total language use; this proportion rises to 80% when including the top 2500 words, and shows a similar gradual decreasing trend after the 2500-word threshold. This set of data confirms the asymmetric distribution feature of language. This linguistic asymmetry also exists in other linguistic levels such as semantic distribution of words, grammatical structure distribution, and collocation patterns. With the help of keyword and n-gram technology, corpus linguistics can easily reveal the most typical linguistic features and the actual usage rules of language in real life. Through frequency statistics, corpus linguistics clearly presents the asymmetric distribution characteristics of language to researchers.

Typical linguistic features and linguistic asymmetry have strong enlightening significance for HVC teachers. Corpus can answer core teaching questions: What are the most frequently used words and phrases in tourism English? Should teachers adjust the emphasis of teaching

content and teaching sequence? Obviously, typical linguistic phenomena should be taught earlier than atypical linguistic phenomena, and allocated more teaching time, which can greatly improve teaching effectiveness. In teaching, teachers should focus on high-frequency words, high-frequency semantics and collocations of words, and high-frequency grammatical structures. If teaching pays too much attention to rare linguistic phenomena and ignores common ones, it will inevitably lead to a low return on teaching time input, which greatly reduces HVC students' motivation to learn tourism English.

Taking the vocabulary list of China's College English Test Band A (Level 3) as a reference, we screened 1000 keywords in tourism English through keyword analysis technology. The top 10 keywords are: Dynasty, China, ancient, mountain, meter, Temple, Buddhist, cultural, Emperor and architecture. Through n-gram technology, we selected 100 high-frequency multi-word sequences in tourism English, the top 10 of which are: as well as, one of the, in the world, is located in, known as the, in terms of, is one of, a large number of, built in and a lot of. HVC students generally have low interest in English learning and very limited class hours, which requires teaching to achieve maximum output in the shortest time, and HVC teaching should follow the principle of "pragmatism first and adequacy for use". These keywords and multi-word sequences screened by corpus linguistics technology are exactly the core focus of our vocabulary teaching.

3.2 Corpus Concordance and Tour Guide Commentary Composition

Cultivating students' ability to create tour guide commentaries is a core task of tourism English teaching. If students only recite commentaries from textbooks, they can only repeat others' views and lack creativity. This requires students to master the ability of independent creation after memorizing basic content. However, HVC tourism majors generally have the following problems in tour guide commentary creation: many linguistic errors in grammar and word use; obvious negative L1 transfer at the lexical and syntactic level; high misuse rate of textual cohesion and coherence. Corpus linguistics has obvious application value in solving these problems.

Using the keyword in context (KWIC) concordance of corpus keywords introduced

above, students can observe and analyze the usage of vocabulary in real texts, and master the most appropriate lexical expression. Focusing on the core professional vocabulary and multi-word sequences we selected, teachers can use the collocation retrieval function to present typical sentence patterns and collocations to students before training them to create their own commentaries, so that students can form a basic cognition. Then, through corpus retrieval, teachers can show students the typical cohesion and coherence characteristics of tour guide commentaries, so as to improve students' textual organization ability in commentary creation.

3.3 "Pre-assembled Chunks" and Tourism English Oral Teaching

Multi-word sequences is a term in corpus linguistics, called "pre-assembled chunks" by Bollinger. Bollinger [9] pointed out: "According to the information we want to express, we take out these 'pre-assembled' phrases from the vocabulary bank in our brain, and after subtle processing, we can form grammatically correct sentences, and turn them into fluent language to fill the conceptual gaps we need". It is generally recognized that native English speakers master a large number of pre-assembled chunks, which is a key factor contributing to oral fluency. Altenberg and Granger [10] found that about 70% of daily spoken English is composed of pre-assembled chunks. Bollinger [9] emphasized that language use is not always creative, nor does it always combine individual words temporarily according to grammatical rules. Instead, language use is highly repetitive, and many utterances are pre-assembled chunks stored in memory.

The ultimate goal of the tourism English course is to improve students' oral English expression ability in tour guide explanation and service. To achieve this goal, HVC students need to input sufficient and practical professional language knowledge, and output language expression fluently. For tourism English teaching, we can make full use of pre-assembled chunks. In this study, we extracted 100 commonly used pre-assembled chunks through the n-gram technology of corpus linguistics. These pre-assembled chunks are similar to the fixed collocations we often mention. They can be quickly extracted and used for immediate expression, without paying extra attention to grammatical structure, which reduces the trouble

of grammatical conversion and improves oral fluency. In the process of tourism English oral teaching, teachers can integrate pre-assembled chunks into the classroom, use corpus to count and classify the collected pre-assembled chunks, and analyze their meaning combined with context. This can help students express fluently in appropriate contexts, so as to enhance the self-confidence of HVC students in oral expression.

3.4 Typical Linguistic Features and Tourism English Textbook Development

Up to now, the design of tourism English teaching materials still mostly relies on traditional linguistic description based on limited corpora. Textbook compilers mainly rely on experience and subjective judgment to determine linguistic features and corresponding learning order, which leads to great arbitrariness in textbook compilation. Corpus linguistics holds that high-frequency linguistic features should be the focus of teaching. Determining the order of teaching content design according to the frequency of linguistic features can improve teaching efficiency and enhance students' sense of achievement. In the compilation concept of textbooks, we should gradually realize the transformation from equal treatment of all linguistic features to attaching importance to typical linguistic features. In the selection of textbook content, typical texts can be selected from specialized corpora to improve the representativeness of textbook content. In addition, specialized corpora can also provide relatively real language materials for textbook compilation.

4. Conclusion

At present, the application of corpus in HVC ESP teaching focuses on four dimensions: vocabulary, syntax, text structure and text type. Research at the vocabulary level is carried out by retrieving the frequency, context and adjacent collocations of a specific lexical item. From an English corpus, we can obtain a word frequency list reflecting the actual usage of English vocabulary, and can also retrieve classified information taking a certain syntax, semantics or part of speech as the retrieval unit. Research at the syntactic level refers to the use of quantitative analysis to study the combination patterns of various part-of-speech tags and the usage of various sentence patterns in

grammatically annotated corpora. Text structure research explores the linguistic basis of cohesion and coherence in spoken and written texts, especially the structural pattern of discourse communication. The main content of text type research is to study the combination of linguistic features of different texts, so as to construct linguistic standards for discourse or stylistic classification. G. Kennedy, a New Zealand linguist who has long been engaged in corpus research, pointed out that the application research of corpus in English teaching has always focused on two core questions: what linguistic units, forms, systems and combinations exist in a certain language, a certain register or a certain type of text; and what are the frequency, occasion, time, cause and object of their use.

Although corpus application research in English teaching is booming, we still need to realize that there are still many aspects to be improved in China's research. Most studies in the field of language teaching and acquisition are empirical studies on individual words or individual linguistic phenomena, and there is a lack of overall research on Chinese learners' interlanguage, its linguistic characteristics and overall framework. In addition, the problems of single research method and research repetition are very prominent, most of which are time-consuming and repetitive work. The application of relevant corpus linguistics theories to the analysis and practice of English teaching needs to be further detailed. In some important teaching fields, such as audio-visual and speaking fields like English listening and pronunciation teaching, only a few scholars have tried to conduct research with corpus methods.

Contemporary specialized corpora show three major development trends: First, specialization and miniaturization, which means that the genre or theme of ESP corpus is more specific, and the scale of the corpus is correspondingly smaller; Second, multimodality of corpus data, corpus-based multimodal text analysis can study the specific meaning and meaning construction process of multimodal texts; Third, localization of corpora, which means that corpus collection and design pay more attention to meeting the needs of local regions and professional fields. In addition, many scholars believe that the construction of future specialized corpora will serve classroom teaching more. The corpus selected in this study mainly serves ESP

teaching in HVCs, which conforms to the development trend of specialized corpora. This study still has some debatable points. For example, in terms of the selection criteria of corpus samples, should we classify by region or by topic? This study adopts average sampling based on regions, and the appropriateness of this approach still needs comments from experts. In addition, the corpus should be updated in a timely manner to keep pace with the times, so as to provide timely updated language resources for teaching. The application prospect of tourism English corpus in HVC ESP teaching is very broad.

The application of corpus and corpus linguistics has penetrated into all aspects of English teaching, greatly enriched the forms of English teaching, provided vivid language environment for the classroom, and improved students' efficiency in learning English. The research on corpus application in English teaching will definitely become a hot spot in the teaching field in the future. In teaching practice, we should take corpus as a reference, give full play to the creativity of English learners, correctly handle the relationship between corpus and English teaching, and serve the development of English teaching with Chinese characteristics.

Acknowledgments

This paper is supported by “Qinglan Project” (WXCYYQL20240502) of Wuxi City College of Vocational Technology.

References

- [1] Chen, J., & Liu, C. (2021). Corpus-driven construction of project-based tourism English translation teaching model. *Computer-Assisted Foreign Language Education in China*, (4), 32-39.
- [2] Zhao, N., & Zhang, J. (2026). Application of multimodal tourism corpus in tourism interpreting teaching. *Foreign Language Education Frontiers*, 9(1), 56-64.
- [3] Hidayat, N., Nurlela, & Saleh, M. (2022). A corpus-based blended learning model for EFL tourism students in Indonesia. *Journal of Computer Assisted Language Learning*, 35(4), 412-430.
- [4] Zhang, J., & Wang, H. W. (2023). Construction and teaching application of tourism Chinese corpus for international Chinese education. *Chinese Teaching in the World*, 37(2), 241-252.
- [5] Wang, W. W. (2022). Application of large-scale general tourism corpus in contextualized tourism English teaching for higher vocational colleges. *Chinese Foreign Language Education*, 15(2), 23-31.
- [6] Li, Y. (2023). Corpus-assisted tourism intercultural communication teaching: Paths and effect verification. *Foreign Languages and Their Teaching*, (3), 89-96+150.
- [7] Huang, N. (2024). Research on corpus-driven real-scene teaching reform for tour guide major under the background of new quality productivity. *Chinese Vocational and Technical Education*, (12), 45-51.
- [8] He, A. P. (2004). *Corpus linguistics and English teaching*. Foreign Language Teaching and Research Press, Beijing, China.
- [9] Bollinger, D. (1975). *Aspects of language* (2nd ed.). Harcourt Brace Jovanovich, New York, NY, United States.
- [10] Altenberg, B., & Granger, S. (2001). Lexical bundles in native and non-native English: A corpus-based comparison. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 93-117). John Benjamins Publishing Company, Amsterdam, Netherlands.