The Construction and Application
of Online Tourism Translation Corpus for MTI Students

LI Yanqin

Hunan Applied Technology University, China

Abstract: In recent years, the application of online technologies such as Big Data, Cloud Computing and the Internet of Things has correspondingly given rise to smart tourism. The rapid growth of international tourism highlights the importance of tourism translation and relevant translators. Meanwhile, the online corpus is one of powerful platforms than can provide amount of information. Therefore, the author constructs three self-built tourism corpora for the application and exploration of translation, and then makes blueprints for MTI curriculum design, supplies data for researches on tourism translation and designs for the training of the tourism industry to supplement new research directions and methods so as to cultivate MTI students and practice teaching activities in the new era from the perspective of the corpus.

Keywords: MTI; online tourism translation corpus; construction and application

Notes on the contributor: LI Yanqin is a college teacher of Hunan Applied Technology University with academic interest in translation theory and practice, and corpus-based translation studies. Her email address is 435173059@qq.com.

1. Introduction

The global tourism has been gradually developed in recent times for the impact of the Internet. Tourism enterprises have expanded their commercial point of view and have found a faster way to attract consumers through increasing tourism advertising on the Web (David Giménez Folqués 2015). However, the problem is how to integrate technology with cultural industry. Then, a new research method has to be mentioned, that is, corpus. The corpus is based on Big Data, which can provide a relatively accurate and persuasive data foundation for practical research. Besides, the tourism industry has a large amount of audiences that can provide people a great number of datum. Therefore, the establishment of a tourism corpus has more advantages over other types of corpora.

The rise of China's tourism industry, the enhancement of national foreign policy and the frequency of international exchanges and cooperation have made China's great demand for translators. As Huang Youyi said, "Since the 1980s, postgraduate education in China has been cultivating academic or research-oriented masters or doctors...... The research personnel are gradually saturated, while the demand for practical personnel is rising." (Cao Xinyu 2018) However, few articles on both of domestic tourism corpora and MTI students has been written yet. Therefore, the author constructs three self-built corpora, i.e. Chinese-English/English-Chinese (C-E/

E-C) bilingual parallel corpus, Chinese monolingual corpus and English monolingual tourism corpus, which can provide rich storage for comparative studies between Chinese and English for MTI students. Meanwhile, students and teachers can both make empirical studies of tourism translation and teaching practice to promote high-level and multifaceted domestic corpus research.

According to thelanguage types of collected datum, it can be divided into the monolingual corpus and bilingual/multilingual corpus. The former one just contains texts of a single language and consists of many language examples by native speakers; the latter composes of two or more language texts and has correspondence/parallel form, analogy form and translation form. (Zhu Xiaomin 2011：32)

2. The Design of Online Tourism Corpora

2.1 *The monolingual corpus project of English and Chinese tourism texts*

Dating back to the late 1950s, the researchon monolingual corpus has a long history. M. Baker pointed out that only by taking the relevant source language or target language into consideration the analogy corpus can we analyze the established language patterns in the translated text and explore the essence of translation. (Baker 1995) This research model obviously differs from traditional research method that focuses on exploring the relationship between the source text and the target text. What's more, it marks the formal establishment of the corpus translation research method. (Huang Libo & Zhu Zhiyu 2012) But so far, few academic papers on monolingual corpora of English tourism texts at home and abroad have been written. (Zhu Xiaomin 2011：32)

Meanwhile, the domestic monolingual corpora are represented by the CCL Chinese corpus of Peking University and the BCC modern Chinese corpus of Beijing Language and Culture University. Both corpora have a capacity of more than fifty billion, including newspapers, literature, microblog, technology, synthesis, ancient Chinese and other fields. However, there is a lack of specific Chinese monolingual corpora and corresponding data support for certain aspects of specific research.

For analogy corpus, it is easy to overlook the value brought by monolingual corpus when people are conducting a research on parallel texts. As far as the research object of this article is concerned, in terms of the cultivation of MTI students, constructing a monolingual corpus of Chinese-English tourism texts can provide them with simulated texts and data, as well as authoritative source language support and target language reference. In terms of teaching practice, a high-quality monolingual corpus can help researchers to illustrate the MTI classroom teaching as a text supplement.

2.2 *The parallel corpus project of Chinese and English tourism texts*

In terms of corpus construction, a number of large-scale and representative bilingual parallel corpora have been built or have been under construction in China, such as the one hundred million Chinese E-C Parallel Corpus developed by Wang Kefei of Beijing Foreign Studies University, the C-E Parallel Corpus of Shakespeare's Dramas developed by Hu Kaibao of Shanghai Jiao Tong University, and the C-E Parallel Corpus of *A Dream in Red Mansions* consrtucted by Liu Zequan of Yanshan University. These parallel corpora are mainly for language (especially contrastive) researches, translation researches, translation practices, and dictionary compilation services (Xiong Bing 2015：3). However, as far as China's bilingual parallel tourism corpora are concerned, few of them are used to cultivate MTI translators and to conduct pedagogical research. Therefore, considering phenome-

non, the last and most important corpus here is the C-E bilingual parallel tourism translation corpus. For the cultivation of MTI students, a large amount of data provided by the bilingual parallel corpus can be used to analyze tourism text, and to compare language research and the implementation of MTI's machine-assisted translation.

Looking at the above three corpora currently built for translation research, most of them are mainly literary texts, or feature encyclopedia texts (including various literary and non-literary works), and they combine regional characteristics or target to a certain genre. However, the bilingual specialized corpus is relatively rare. (Li Dechao and Wang Kefei 2010：47) The establishment of a specialized bilingual corpus is extremely important for current domestic researches and development. Therefore, considering the multidimensionality of student training and the operability of MTI teaching practice, the author believes that the scale of CE monolingual tourism corpus and CE bilingual translation tourism corpus needs to be expanded. After ensuring the standard and authority of the corpus, the creation of a multi-dimensional tourism corpus can integrate tourism translation with MTI practice, provide sufficient data and texts for relevant researches, and actively expand the scope of research.

3. The Construction Process of Online Tourism Translation Corpus

With the renewal and iteration of modern information technology, the increasing data and the high complexity brings out the Big Data (Lu Zhi and Hu Jian 2018). The explosive growth of data has brought not only convenience to the development of corpora, but also challenges. The large amount of input of corpus leads to the corresponding problems in the process of corpus construction. Therefore, the design and editing of corpus can be directly related to the effectiveness and reliability of the research based on a certain corpus. (Kennedy G. 1988)

3.1 Data collecting

Since the corpus is divided into a bilingual parallel corpus and a monolingual analog corpus, it should be collected and screened according to the category of the corpus when collecting data.

The monolingual data of English tourism texts are mainly selected from five regions, that is, Asia, Europe, America, Oceania and Africa. The texts are collected through various channels that based on the originals displayed on the local tourism official website. Secondly, they contain resources from well-known websites such as Wikipedia, Lonely Planet, and the British Encyclopedia. In addition, they also select YouTube videos and travel magazines to extract materials. On the one hand, the diversification of corpus source ensures the diversity of the corpus, and on the other hand it ensures the authority of the corpus, laying a good corpus foundation for the establishment of a monolingual corpus of English tourism texts.

The source of the monolingual corpus of Chinese tourism texts is not that complicated. It is based on the 249 5A-level tourist attractions ranked by the Ministry of Tourism and Culture of the People's Republic of China according to the classification of natural and human landscapes. The collection of data is mainly through national tourism official websites, provincial/regional tourism official websites, tourism books officially published by publishing houses, and well-known tourism websites. Due to the rich tourism resources in China, the collection of data is quite convenient. The storage capacity is larger than the monolingual corpus of English tourist texts, which provides good examples for observing language phenomena in tourism texts.

The data collecting of the Chinese-English tourism translation parallel corpus is more difficult than the above two corpora, for the data collected in this paper will be especially used for research purposes. The quality of the translation version in the parallel corpus must be guaranteed. So, corpus collectors need to be qualified with a

Chinese-English bilingual ability. The data in this corpus comes from online electronic texts, published books, official sites of tourist attractions, and bilingual subtitles and audios of travel documentaries produced by authoritative producers, such as *HJEnglish*, *China Daily*, *BBC*, etc.. Most of these bilingual corpora are given priority to the translation version from Chinese into English.

So far, about 100,000 words have been collected in English, about 300,000 words in Chinese, and about 200,000 words in bilingual texts. Since research is a process of continuous development, the storage capacity will not remain unchanged, and will continue to be expanded over time.

3.2　Corpus cleaning

In the process of developing a corpus, the input of a great amount of data will inevitably lead to phenomena such as misplacement, garbled characters, and symbol errors. At this moment, texts need to be cleaned. The corpus cleaning is mainly divided into two parts：text digitization and text organization.

The text digitization is very important to the construction of the corpus, because the text format in the corpus is related to whether the program can run successfully. For this reason, the subject needs to build three corpora with diverse text sources and TXT format. Therefore, the corpus texts in PDF format and HTML format need to be converted into Word formant by the Kingsoft format converter, and then be saved in TXT format. However, since none of the videos' subtitles can be collected currently and the video parameters cannot be extracted, the subtitle extraction in the AVI format and WAV format is manually finished.

Although the current software applications are quite mature, it still cannot be 100% accurate. Therefore, the text converted by the software will inevitably have garbled sentences, misplaced words and blank lines. Due to the large amount of texts, the Text Processing Master 8.0 will be used to process the text in batches. The software has many functions, such as deleting all blank lines, adding characters in batches to the first line, processing special characters, etc., which has a very good auxiliary effect on the cleaning of texts. After this, the text will be manually checked, such as text errors, vocabulary spelling and other errors that the software cannot recognize, to ensure that the cleaning of texts can reach a high accuracy.
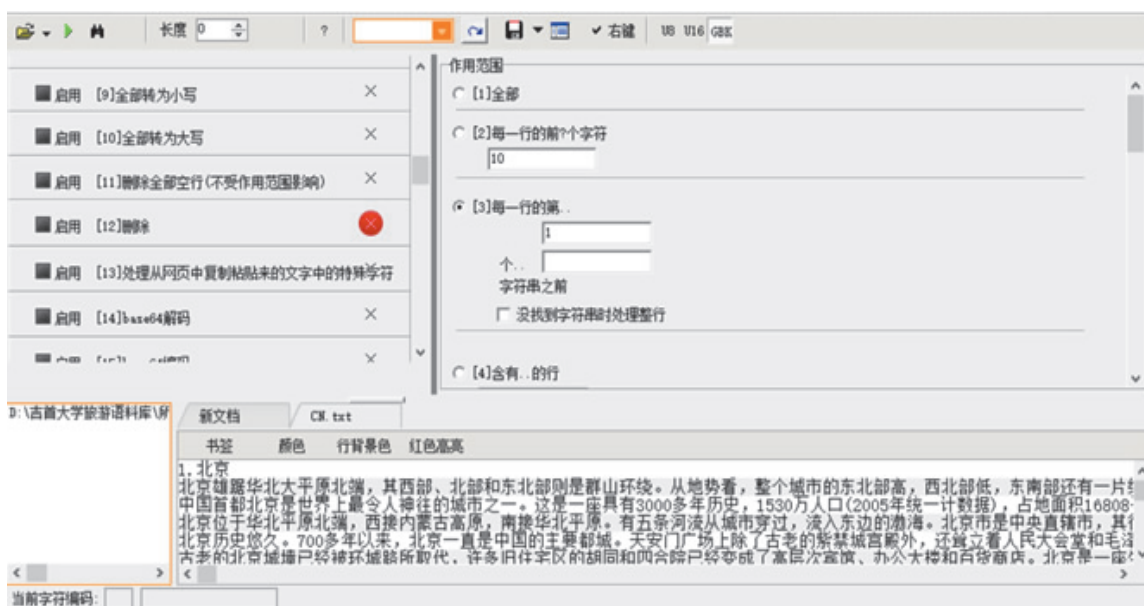

Figure 3.1：*The Screenshot of Text Processing Master* 8.0

3.3 Corpus tagging

Tagging refers to the division of corpus components（usually parts of speech）to facilitate further retrieval and research. This tagging method is called POS tagging（part-of-speech tagging）.（Li Dechao and Wang Kefei 2010：51）. Since this corpus contains bilingual language, it needs to take different methods of tagging for languages. When tagging the English corpus, the text is mainly tagged by CLAWS POS tagger based on the BNC's method of dividing and abbreviating the corpus（such as NN1 for singular nouns and NN2 for plural nouns）（Li Dechao and Wang Kefei 2010：51）. However, since the machine marking cannot achieve 100% accuracy, manual verification is also required.

```
Huangshi_NP0 Village_NP0 Huangshi_NP0 Village_NP0 ,_PUN whose_DTQ original_AJ0
name_NN1 is_VBZ Huangsi_AJ0 village_NN1 ,_PUN is_VBZ also_AV0 called_VVN
Yellow_AJ0 Lion_NN1 Village_NN1 ._SENT -----_PUN
Located_VVN in_PRP the_AT0 middle_NN1 of_PRF the_AT0 Zhangjiajie_NP0
National_AJ0 Forest_NN1 Park_NN1 ,_PUN it_PNP is_VBZ surrounded_VVN by_PRP
cliffs_NN2 and_CJC green_AJ0 jungles_NN2 ._SENT -----_PUN
With_PRP an_AT0 altitude_NN1 of_PRF 1,080_CRD meters_NN2 above_PRP sea_NN1
level_NN1 and_CJC an_AT0 area_NN1 of_PRF about_AV0 16.4_CRD hm2_UNC ,_PUN
Huangshi_NP0 Village_NP0 has_VHZ a_AT0 large_AJ0 number_NN1 of_PRF scenic_AJ0
spots_NN2 such_PRP as_PRP Remaining_AJ0 Piers_NN2 of_PRF the_AT0 Heavenly_AJ0
Bridge_NN1 ,_PUN Natural_AJ0 Fresco_NN1 ,_PUN Front_AJ0 Garden_NN1 ,_PUN
Golden_AJ0 Turtle_NN1 in_PRP the_AT0 Clouds_NP0 Sea_NP0 ,_PUN Conch_NP0 's_POS
Peak_NN1 ,_PUN Flying_AJ0 Cloud_NN1 Cave_NN1 ,_PUN Sealed_AJ0 Book_NN1
Precious_AJ0 Case_NN1 ,_PUN Southern_AJ0 Hearen_NN1 Pillar_NN1 ,_PUN
Souther_NP0 Heaven_NN1 Gate_NN1 ,_PUN Six_CRD Wonders_VVZ Pavilion_NN1 etc_AV0
._SENT -----_PUN
Walking_VVG along_PRP the_AT0 Huangshi_AJ0 Village_NN1 ,_PUN you_PNP will_VM0
feel_VVI the_AT0 mountain_NN1 peaks_NN2 are_VBB below_PRP your_DPS feet_NN2
,_PUN and_CJC all_DT0 the_AT0 peculiar_AJ0 landscapes_NN2 are_VBB within_PRP
your_DPS eyes_NN2 ._SENT -----_PUN
Therefore_AV0 ,_PUN there_EX0 is_VBZ a_AT0 saying_NN1 :_PUN "_PUQ If_CJS
you_PNP went_VVD to_PRP Zhangjiajie_NP0 but_CJC did_VDD n't_XX0 climb_VVI
the_AT0 Huangshi_AJ0 Village_NN1 ,_PUN then_AV0 all_DT0 the_AT0 scenic_AJ0
spots_NN2 you_PNP visited_VVN will_VM0 be_VBI meaningless_AJ0 "_SENT ._PUN
```

Figure 3.2：*The Screenshot of English Tagging Software*

Compared with general-purpose texts, tourism texts are a special style for specific groups, and their syntax, vocabulary and rhetoric are also unique（Tang Fang and Li Dechao 2016）. Therefore, when tagging Chinese texts, the corpus first marks the part-of-speech tagging（including nouns, verbs, prepositions, strings, etc.）on the ICTCLAS platform according to the traditional POS tagging method, and then, it uses corresponding rhetorical means to mark the corpus based on a prominent rhetorical feature in the tourism text. Since there is no ready-made software that can complete this kind of labeling, it is necessary to rely on manual work. In the process of manual review, by studying the phonetic, vocabulary and grammatical characteristics in the text, the rhetorical characteristics contained in the sentence are obtained. Setting the sentence as a unit, specific tags would be made for different rhetorical techniques. This corpus mainly uses autonomous coding：for example, use <SIM> for similes, <PUN> for puns, and <PARA> for parallelism. Although manual coding is a huge project for large-volume texts, once it is completed, these annotations will reveal the differences in rhetorical usage of English and Chinese tourism texts and provide strong evidence for the differences in the aesthetic concepts of the two ethnic groups behind the texts. It is of great benefit to comparative rhetoric and comparative aesthetics research.

Figure 3.3：*The Screenshot of Chinese tagging software*

Figure 3.4：*The Screenshot of Tagged Text of Chinese Rhetorical Devices*

Figure3.5：*The Screenshot of Tagged Text of English Rhetorical Devices*

3.4 Corpus alignment

The main steps involve sentence division and sentence coding. The Chinese in the bilingual tourism translation corpus needs to be divided into five wildcards in word，which are respectively .^p，? ＾p，！^p，"^p and ^p. A better effect of sentence division can be achieved by replacing the five wildcards with the corresponding five punctuation marks. The sentence division of English corpus needs only four wildcards，namely .^p，? ＾p，！＾p，"^p，again，just repeat the steps above. Finally，import the text into WordSmith Tool 6.0，click "right click" next to the corpus，and then click"view（Viewer）" to see the text that has been divided into high sentences. Click "Save" to export the text.
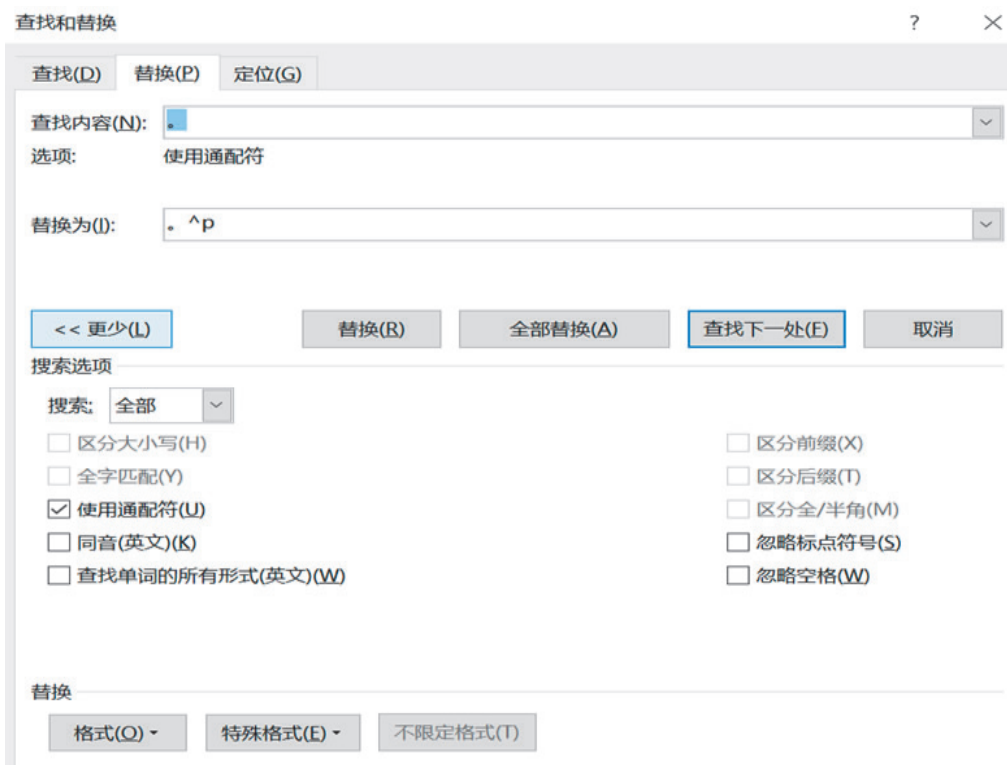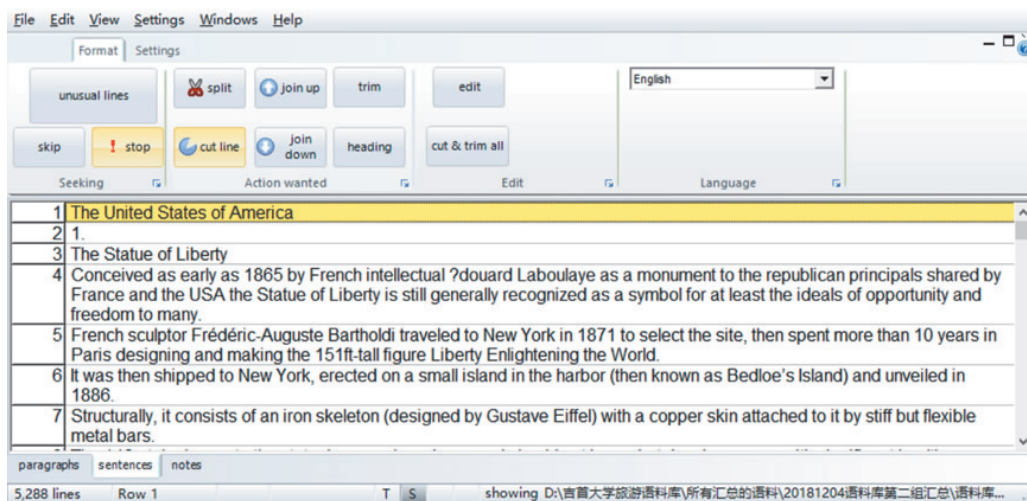

Figure3.6：*The Replacement of Wildcard "。^p"*

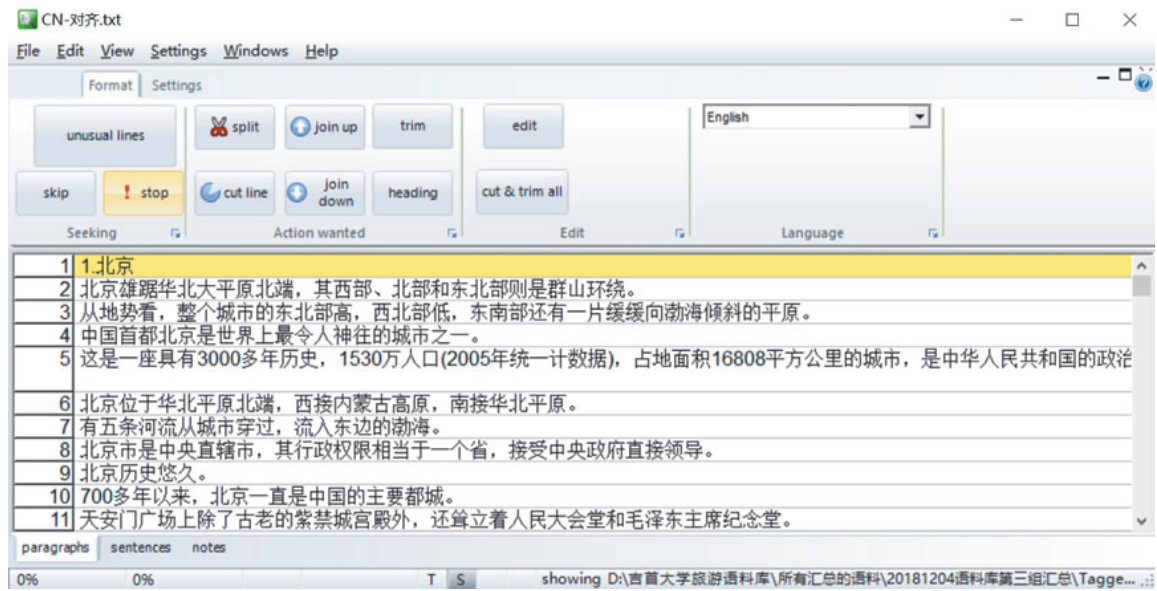**Figure3.7**：*Sentence Division in English*

Figure3.8：*Sentence Division in Chinese*

After the Chinese and English sentence division is completed，the divided version should be imported into the Tmxmall online alignment platform to export the text for the first alignment，and then import the text into the ABBYY Aligner software for secondary alignment. At this time，the accuracy of alignment has basically reached 85%. However，due to the differences between Chinese and English languages and cultures，the translation of tourist texts often does not fully comply with the principle of sentence-to-sentence. It will not be completely corresponding when the texts are imported into ParaConc software. Therefore，researchers need to use sentences as a unit to compare Chinese with English. That is，to add <seg> and </seg> at the beginning and end of the sentence respectively，so that a coded correspondence can be formed. At this point，import the coded text into ParaConc，and the text can be highly aligned.

Figure 3.9：*Corpus Coding*

| 北京位于华北平原北端，西接内蒙古高原，南接华北平原。 | Situated in northeast China, Beijing adjoins the inner Mongolian Highland to the northwest and the Great Northern Plain to the south. |
| 有五条河流从城市穿过，流入东边的渤海。 | Five rivers run through the city, connecting it to the eastern Bohai Sea. |

Figure 3.10：*Corpus Alignment*

The construction of corpus generally involves four major steps：data collecting，corpus cleaning，corpus tagging，and corpus alignment. In corpus processing，the software has been frequently used. The construction of the tourist corpus in this paper is no exception. However，it is worth mentioning that the rhetorical signs involved in this paper have not been specifically studied. Due to the special nature of the tagging，there is currently no software that can automatically make marks，and it needs to be executed manually. However，the initial realization of this tagging can provide novel data support and new research directions for training of MTI students and teaching practice.

4. The Application of Online Tourism Translation Corpus

Corpus research and English for Specific Purposes（ESP）started almost simultaneously in the late 1950s and early 1960s and developed rapidly in the 1980s. Interdisciplinary and multi-dimensional researches have been studies continuously since the 1990s（Feng Zhengbing and Wang Feng 2016）. With the development of the tourism industry，the importance of tourism translation has increased，and tourism translation has gradually bordered on ESP. The combination of corpus and ESP（tourism English）can be used for various researches. This paper intends to analyze the language，culture and modernization phenomena in tourism translation from three aspects：MTI curriculum design，researches on tourism translation and training for the tourism industry.

4.1 MTI curriculum design

Traditional translation teaching mode focuses on the teaching of translation knowledge and skills. Teachers usually play a leading role in the teaching process. Students' learning becomes largely passive. At the same time，the translation examples that students encountered are often limited in number，scattered and isolated in translation textbooks. For translation examples that lack contextual support，the evaluation of translation quality is almost entirely dependent on the intuitive judgment of teachers or students（Xiong Bing 2015：7）. Corpus-based curricula are normal in foreign countries，such as the learner-annotated corpus for a blended EFL reading course（Takeshi Okada 2018），the learner corpora with translation course（María Teresa Sánchez Nieto 2016），corpus in a 300-level Spanish grammar course（Carlos Benavides 2015），corpus-based EAP course for NNS doctoral students（David Lee & John Swales 2005），corpus for British student writing（Hilary Nesi et al. 2004），and so on. While for most practical majors such as MTI，traditional teaching methods cannot fully develop the potentials of students. Therefore，the MTI curriculum needs to be reformed according to the development of time and technology.

The "pedagogic processing" of corpora is a profound reflection on the application of corpus linguistics and language education by many scholars at home and abroad. It includes how to get the corpus into the setting of language syllabus，the compilation of teaching materials，the design of classroom teaching activities，implementation and effect evaluation，etc.. （Gui Shichun et al. 2010）At the same time，online courses represented by MOOCs have provided new ways for innovation and education. It becomes an important subject in the field of ed-

ucation（Liu Bing et al. 2016）. Therefore, in order to innovate the MTI teaching practice, corpus-based online MTI curriculum design can be described as the top priority. In view of this, based on the self-built tourism corpus, the following one-way and two-way tourism courses are designed.

4.1.1   The one-way online tourism curriculum design

Online education, like on-site education（face-to-face education）, has its own model and method. From the perspective of the communication process and system, online education can be divided into four development stages：one-way online education model, two-way interactive online education model, cooperative learning online education model and development-oriented online education model. One-way education is mainly used by educators to pass information to the educated through several medias. It is an online one-way flow of educational information, with video lectures and lectures. Memoir, micro-video, and thematic web courses（Wang Weijun et al. 2016：68）.

Due to the tendency of one-way teaching, the teaching language in the classroom is usually teacher's language. Therefore, the author uses the Chinese monolingual tourism text corpus as the basis, and further improves MTI students' ability to control their mother tongue through the introduction of scenic spots and text interpretation in a one-way Chinese tourism course. At the same time, the corpus of English monolingual tourism text is extracted into TXT form, which is used as the extracurricular reading text for MTI students to improve reading ability and expand knowledge. In addition, the online teaching platform of university can also review teaching videos, and MTI students can also consolidate their knowledge by watching these videos.

4.1.2   The two-way online tourism translation curriculum design

The two-way interactive online education model is mainly based on the first generation of online education, adding interactive feedback links, considering the learning experience and learning effects of learners, and emphasizing the interactive nature of learning and the use of technology to transmit education and teaching content. At the same time, the design provides feedback with learners and organizes students to carry out exchanges and discussions. At the same time, it contains feedback with learners and organizes students to communicate and discuss. This kind of education model mainly focuses on three-screen courses, MOOCs and open video courses.（Wang Weijun et al. 2016：69）
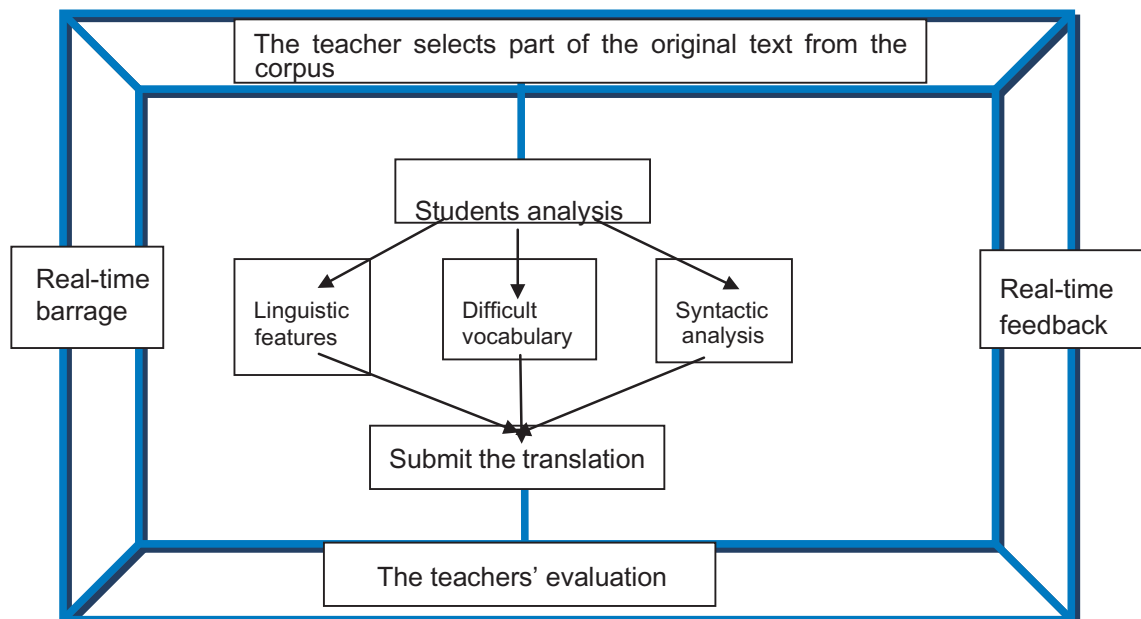

Figure 4.1：*Flow Chart of Two-way Interactive Online Course*

As shown in the figure above, the design of the tourism translation course adds a real-time barrage function in addition to the test questions, assignments, and learning works included in the conventional two-way interactive education mode. According to the normal functions, MTI students can analyze and translate the original text from the parallel tourism translation corpus by functions in the corpus, such as concordance, clusters, collocates, word list and keyword list. The barrage video originates from a Japanese video website called "niconico". This website pioneers the technology of real-time commenting by netizens on the video playback page. A large number of comments may suddenly appear on the screen in a certain second and float across the screen like a horizontal version, this type of online comment is called "dàn mù"（dàn mù）by Japanese netizens（Zhang Qingfei, 2014）. Currently, barrage videos are mostly used in media fields such as movies and television. However, the potential value of barrage videos in education fields such as distance education and online learning has rarely attracted the attention and exploration of domestic and foreign researchers（Li Haifeng and Wang Wei 2015）. For traditional two-way interactive online education, the inability to obtain real-time feedback and interaction from students is a major disadvantage. The addition of real-time barrage enables MTI students to reflect and solve problems in time during the learning process. To a certain extent, the learning efficiency is improved, and the participation initiative of MTI students is enhanced. Compared with other majors, the MTI major emphasizes the active participation of students. Therefore, online teaching based on real-time barrage is more suitable for practical MTI teaching. The teaching content is informative and practical, which further promotes the innovation of MTI teaching.

4.2   Training for tourism industry

Based on the corpus, Tim John put forward the concept of Data-driven Learning（DDL）in the 1990s, which set off a wave of corpus application in teaching（Wang Xiaowen and Ge Shili 2015）. But teaching is not just limited to the classroom. In this era of big data, corpora can be applied to a wide range of fields depending on the corpus. For example, science and technology translation corpus can be applied to the field of science and technology research and development, and novel translation corpus can be applied to the field of novel compilation. The tourism translation corpus developed in this paper is also oriented to many aspects. Here, the preliminary conception is mainly based on the corpus-based tourism translation workshop and the tourism electronic textbook APP.

4.2.1   The design of corpus-based tourism translation workshops

With the strengthening of ties between countries, the requirements for translators have been further increased. The importance of translation workshops is self-evident for translators and MTI students. Regarding the translation workshop, it was first proposed by Gentzler. He believed that the translation workshop was a forum for two or more translators to cooperate in translation activities. Then Li Ming and Zhong Weihe redefined the translation workshop, that is, in the translation workshop, multiple translators work together to complete the same translation task and express their own opinions to seek a common recognized translation version.（Zou Yao *et al.* 2018：862）

But nowadays the professional ability of translators is no longer limited to a certain aspect. Only a full range of translators can meet the current requirements of the development. In tourism translation, not only translation quality, but also translation speed is required. The use of corpus can greatly improve the translation efficiency of translators. Since the existing high-quality translations, translators can search through the corpus to compare which vocabulary used in the translation is more in line with the context and which sentence structure is more in

line with the needs of the target readers. In view of this, a tourism translation workshop platform suitable for students and teachers can be constructed based on the three self-built tourism corpora. The specific construction is shown in the figure below:
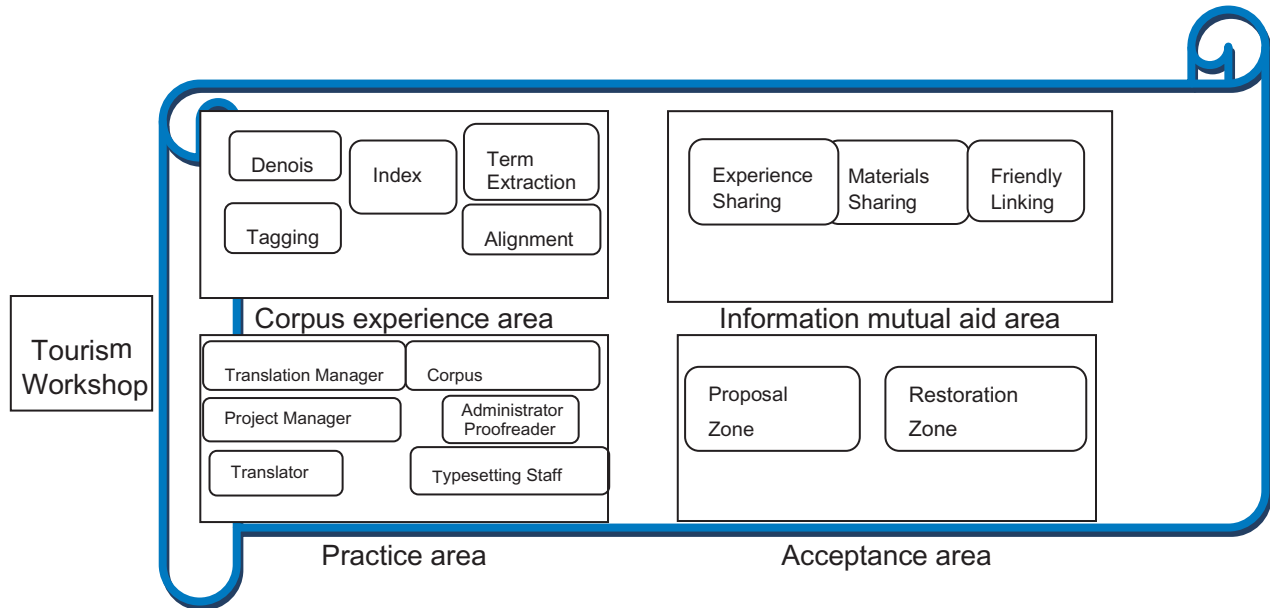

Figure 4.2: *The Floor Plan of the Workshop on Tourism Translation Corpus*

The platform users are divided into three parts: 1) teachers; 2) students; 3) management personnel. In order to ensure the security of the platform, teachers and students use teacher work ID and student ID as login accounts, and management personnel need to register and authenticate according to their ID cards. The platform function currently has four areas, namely the corpus experience area, practice area, information mutual aid area and acceptance area. The corpus experience area is the block where platform personnel can practice denoising, tagging, index, term extraction and alignment, and finally, they can realize the true meaning of corpus-assisted translation. The practice area is mainly based on practical operation, which is mainly divided into the role play of project manager, translation manager, translator, project manager, proofreader, typesetting staff and corpus administrator. Platform personnel can conduct practical exercise according to their own needs and improve themselves during the exercise (Zou Yao *et al*. 2018). The information mutual aid area is relatively simple, in which platform security can share operation and translation experience with each other, as well as translation tools, corpus links and other learning materials. The acceptance area is the area where the management personnel are located. This area collects suggestions and comments from the platform personnel. Based on this, it continues to optimize the platform and fix problems further to improve the tourism translation workshop's operation.

4.2.2 The design of corpus-based travel electronic textbook APP

Textbooks, teachers, and students are the three basic requirements of classroom teaching activities, as well as the three basic elements for teaching quality. They have a decisive, fundamental and substantive impact on teaching activities and teaching quality from different perspectives and levels (Zhang Jiaqi and Cao Dianbo 2019: 39) Traditional textbooks are mainly printing books. With the continuous development of technology, textbooks have different presentation methods. According to their different characteristics, they are mainly divided into printing textbooks, audiovisual textbooks, and electronic textbooks (Zhang Jiaqi and Cao Dianbo 2019: 39). Information Technology (IT) resources and electronic corpus are two integrated tools in the evaluation method.

They are used to identify significant performance changes between learners who have managed such electronic resources and learners who have not（Curado Fuentes A 2004）. With the advancement of technology，most of today's electronic textbooks appear in the form of APPs，such as"Standard Japanese"and"Daily English Listening". With its interactive and comprehensive feature，electronic textbook APPs have gradually become the extracurricular learning for today's students. The design of the travel electronic textbook APP must contain a large amount of travel texts，videos，and audios. When collecting the data，in addition to the travel texts，the author also carried out video and audio interception，which provides a solid foundation for the design of the textbook APP.
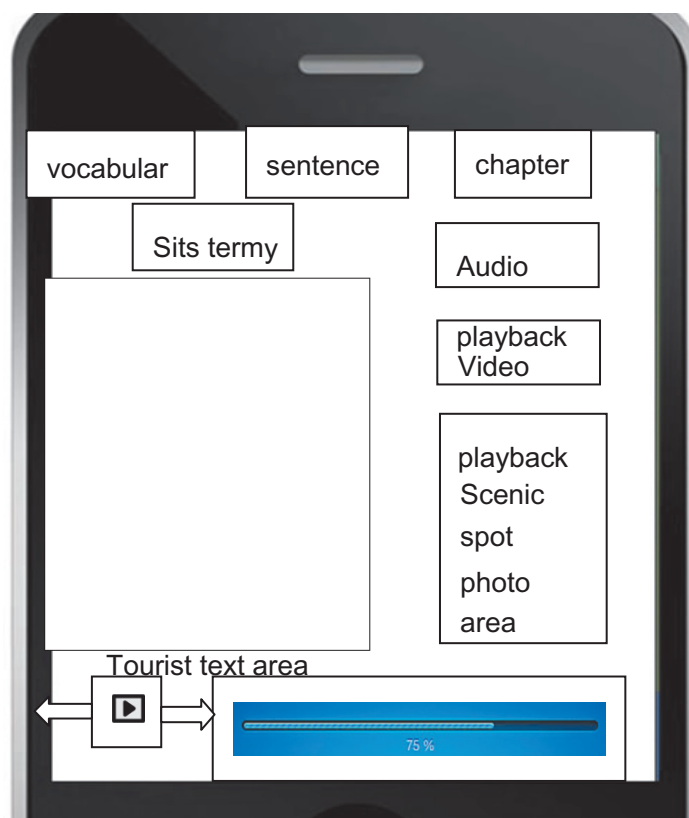

Figure 4.3：*The APP Page of Electronic Tourism Textbook*

Students will inevitably get bored under long-term learning. Therefore，based on the existing vocabulary，sentences and chapter part，the tourist electronic textbook APP has newly added bilingual audio and video links. At the same time，corresponding introduction pictures have been added to the scenic spot photo area. With the assistance of video，audio and pictures，whether it is an interpreter or an MTI major，the boring emotions produced by it can be well alleviated. Video and audio viewing can also exercise the students' audio-visual skills；and the display of pictures can also expand students' knowledge and experience；at the same time，the great convenience of the electronic APP is of great benefit to all MTI students who are still on campus or have graduated.

Whether it's the addition of the"barrage"function in online courses，the creation of terminological database and rhetorical analysis in researches on tourism translation，or the design of translation workshop platforms and APP concepts，these contents and ideas all aim at further promoting the innovative development of classroom teaching practice and cultivating translation talents.

5. Conclusion

With the introduction of corpus methods, translation studies have made remarkable progress in translation theory and application aspects such as translation items, translator style, case-based machine translation, and statistical machine translation. It has also attracted more and more attention from researchers and teachers (Zhu Yubin and Chen Jianlin 2015). Domestic research on corpora has gradually increased, such as English learner corpus (written and spoken), parallel corpus and special English corpus, but corpora on tourism translation texts are rare.

Besides data collecting, corpus cleaning, corpus tagging and corpus alignment, this article focuses on the training of translation talents and the design of teaching practice. However, the design is only the first step. How to effectively integrate the applied research with technology is still a major problem that cannot be solved by the current development. Although there are still some unresolved technical problems in the current research, the three self-built tourism translation corpora, terminological databases, and design blueprints still have important value. They can still lay a good foundation for researches in the future.

References

Baker, M. (1995). Corpora in translation studies: An overview and some suggestions for future research. *Target*, 7(2), 223-242.

BING, Xiong. (2015). Study on the English-Chinese parallel corpus based model of translation teaching. *Foreign Language World*, 4, 2-10.

Carlos, B. (2015). Using a corpus in a 300-level Spanish grammar course. *Foreign Language Annals*, 48(2), 218-235.

Curado, F. (2004). The use of corpora and IT in evaluating oral task competence for tourism English. *CALICO Journal*, 22(1), 5-24.

CAO, Xinyu. (2018). The origin, development and prospect of MTI education in China-Interview with Mr. Huang Youyi. *Chinese Translators Journal*, 5, 65-70.

David, G. F. (2015). Anglicisms in tourism language corpora 2.0. *Procedia - Social and Behavioral Sciences*, 198, 149-156.

David L. & John, S. (2005). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora. *English for Specific Purposes*, 25(1), 56-75.

FENG, Zhengbin & WANG, Feng. (2016). The conception and teaching application of business English news corpus. *Computer-assisted Foreign Language Education*, 2, 54-58+39.

FENG, Xiaofei. (2018). A feasibility study on the construction of tourism translation analogical corpus: A case study of Jiangmen tourism publicity translation corpus. *Journal of Kaifeng Institute of Education*, 6, 61-62.

Giampieri, P. (2019). Corpus-based translation in the tourism sector: A case study with final year bachelor students. *MediAzioni*, 24, 1-32.

GUI, Shichun & FENG, Zhiwei. (2010). Corpus linguistics and foreign language teaching in China. *Modern Foreign Languages*, 4, 419-426.

Hilary, N. & Gerard, S. (2004). Student papers across the curriculum: Designing and developing a corpus of British student writing. *Computers and Composition*, 4, 439-450.

HUANG, Libo & ZHU, Zhiyu. (2012). Corpus-based translation studies: The object of research method. *Foreign Languages in China*, 6, 28-36.

HUANG, Libo. (2017). An overview of translation studies based on corpus of LSP. *Journal of Beijing International Studies University*, 2, 70-82+136.

Johnson, M. (2019). An exploratory survey of the linguistic landscape of Lake Toya. *The Study of Hokkaido Language Culture*, 17, 69-79.

Kennedy, G. (1998). *An introduction to corpus linguistics*. London: Longman, 60.

LIU, Bin & ZHANG, Wenlan. (2016). Online course learning experience: Connotation, development and influencing factors. *China Educational Technology*, 10, 90-96.

LI, Dechao & WANG, Kefei. (2010). Development and application of bilingual corpora of tourism texts: A new approach. *Modern Foreign Language*, 1, 46-54+109.

LU, Zhi & HU, Jian. (2018). An overview of modern corpus development in the era of big data. *Foreign Languages and Translation*, 4, 39-44+98.

LI, Haifeng & WANG Wei. (2015). Barrage Video: A new approach to interactive learning in online video. *Modern Educational Technology*, 6, 12-17.

María, T. S. N. (2016). Learner corpora, corpora of professional translations and creative writing in a course on translation of general texts: An action research project. *Cadernos de Tradução*, 36(1), 143-173.

Takeshi, O. (2018). A blended EFL reading course based on the idea of the learner-annotated corpus. *Japan e-Learning Association Journal*, 18, 19-28.

TANG, Fang & Li, Dechao. (2016). A corpus-based study of the lexical characteristics of "Translation Inherent Types" in Chinese tourism texts. *Journal of PLA University of Foreign Languages*, 3, 117-124.

WANG, Xiaowen & GE, Shili. (2015). Research on the data-driven teaching model of college English. *Foreign Language Research*, 1, 47-52+112.

WANG, Weijun & JIANG, Shuangshuang. (2016). Analyzing students stroke order errors in Chinese characters writing using automatic tracking systems. *e-Education Research*, 2, 68-74+101.

WU, Wei & CUI, Gang. (2016). The practice research of process genre approach in English literature review writing teaching. *Journal of PLA University of Foreign Languages*, 2, 73-81.

ZHANG, Jiaqi & CAO, Dianbo. (2019). The enlightenment of electronic textbooks to the development of new stereoscopic textbooks. *The Chinese Journal of ICT in Education*, 9, 39-42.

ZHANG, Qingfei. (2014). The present and future development of barrage movies. *Movie Literature*, 12, 9-10.

ZHU, Xiaoming. (2011). On the employment of COCA and CCL in translation teaching. *Foreign Language Learning Theory and Practice*, 1, 32-37.

ZHU, Yubin & CHEN, Jianlin. (2015). The theoretical conception and engineering practice of a corpus-based computer-assisted translation teaching platform. *Technology Enhanced Foreign Language Education*, 4, 52-57.

ZOU, Yao & ZHENG, Weitao. (2018). Construction of teaching and training platform based on multi-modal sports corpus "Translation Workshop". *Journal of Central China Normal University* (Natural Sciences), 6, 862-870.

(Editor: LU Chunying)