

# Study on the Construction of the C-E Translation Corpus for the Public Service Areas of Dalian

Xiaoxuan Yang

School of Foreign Languages  
Dalian Neusoft University of Information  
Dalian, China

**Abstract**—In order to promote the urban internationalization progress under the Belt and Road initiative, the paper conducts a study on the construction of the C-E translation corpus for the public service areas of Dalian. With the support of *Guidelines for the Use of English in Public Service Areas (GB)*, the study integrates all the rules in the national standards with the local cultural and geographical characteristics of Dalian to design the framework of the corpus. The corpus construction process is also presented in the paper, including the four steps of corpora collection, corpora pre-edition, corpora annotation and corpora storage. The constructed corpus can be used as a library in the translation software for the public service areas of Dalian. Some problems and their countermeasures in the translation practice with the corpus are also proposed. The paper hopes to shed light on further studies on the C-E translation for public service areas and the parallel corpus construction.

**Keywords**—C-E translation; parallel corpus; public service; corpus construction

## I. INTRODUCTION

Under the Belt and Road initiative, the city of Dalian, one of the major gateway cities from China to the world, is carrying out more international communication and cooperation with countries and regions along the line, which continuously promotes the urban internationalization progress. Therefore, it is indispensable and imperative to provide standardized, professional and understandable C-E translation in the public service areas of transportation, accommodation & catering, tourism, culture & entertainment, commerce & finance, health & medicine, sports, education, etc.

In December 2013, General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China and Standardization Administration of the People's Republic of China jointly issued *Guidelines for the Use of English in Public Service Areas (GB)*. The national standards are of great significance in regulating Chinese-to-English translation, interpretation and writing in the various public services areas, including general rules, transportation, tourism, culture & entertainment, sports, education, health & medicine, post & telecommunications, accommodation & catering and commerce & finance, with unprecedented extensiveness, preciseness and demonstrativeness. It has been almost 5 years since the national standards were formally implemented on July 15, 2014. The unprecedentedly high level

of C-E translation in public service [1], as some experts and scholars anticipated, does not even come into existence. Instead, many mistakes and errors in C-E translation or expression are frequently found on the signs or boards in public areas, which can lead to the confusion and even misunderstanding for foreigners, with potentially damaging effect on the city's internationalization progress.

Arguably, *Guidelines for the Use of English in Public Service Areas (GB)* is standard, comprehensive, rigorous and precise, because it is established based on the guidelines for the use of English in public service areas of Beijing (2007), Shanghai (2009) and some other cities. Based on the national guidelines and combined with Dalian's cultural characteristics and geographical characteristics, this study aims to elaborate on the construction of the C-E translation corpus for the public service area, some key problems and their effective solutions. The study hopes to shed light on the design and construction of the C-E translation corpus for the public service areas in Dalian and provide some reference for the further study on the use of English in public service and the translation corpus construction.

## II. FRAMEWORK OF THE C-E TRANSLATION CORPUS FOR THE PUBLIC SERVICE AREAS OF DALIAN

The national standards are comprehensive, rigorous and precise to guide the C-E translation in public service areas, but there is still a lack of local standards in the translation in public service areas and the English translation is uneven, so it is necessary to be integrated into the overall standards [2]. The E-C translation corpus constructed and studied in this thesis encompasses the 9 public service areas of transportation, tourism, culture & entertainment, sports, education, health & medicine, post & telecommunications, accommodation & catering and commerce & finance, integrated with the general rules in the national guidelines, aiming to standardize the E-C translation for public service areas with local cultural characteristics and geographical characteristics. With this corpus, end users can directly implement standardized, professional and understandable C-E translation in the public service areas, saving the bother to check detailed items in the huge document of national guidelines and the local naming convention. The corpus can greatly promote the efficiency and accuracy of E-C translation for the public service areas in Dalian. In the study, the standards and rules in the national

Research project supported by Dalian Federation of Social Science Associations: Study on the Construction of the C-E Translation Corpus for the Public Service Areas of Dalian Based on National Standards [2018dlskyb238]

guidelines are effectively integrated with the local cultural characteristics and geographical characteristics, constituting the

following framework of the E-C translation corpus for the public service areas of Dalian.

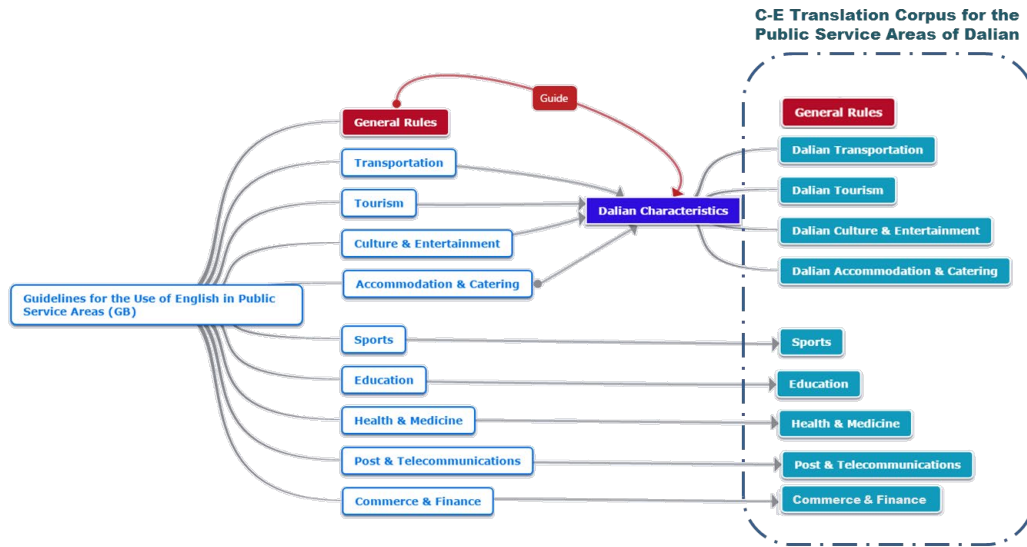


Fig. 1. The E-C translation corpus for the public service areas of Dalian

### III. CONSTRUCTION OF THE C-E TRANSLATION CORPUS FOR THE PUBLIC SERVICE AREAS OF DALIAN

#### A. Construction Basis

Different methods are applied in the construction of different types of corpus, so it is critical to identify the purpose and application field beforehand [3]. The C-E translation corpus for the public service areas of Dalian is designed to provide standardized, professional and understandable C-E translation in the public service areas, saving the bother to check detailed items in the huge document of national guidelines and the local naming convention. Therefore, a Chinese-English bilingual parallel corpus in the field of public service is constructed, while linguistic data are paralleled at word level and sentence level.

In order to meet the requirements in practical Chinese-English translation in public service areas of Dalian, the linguistic data of sports, education, health & medicine, post & telecommunications, and commerce & finance covered in the national standards are strictly selected and imported into the corpus. A survey on the practical Chinese-English translation signs and other forms of text in public service is carried out to extensively collect the linguistic data and achieve the local cultural and geographical characteristics of Dalian. Under the guidance of the general rules in the national standards, necessary combination and modification is conducted on the linguistic data of transportation, tourism, culture & entertainment, and accommodation & catering. The processed data are imported into the corpus, so as to better adapt to the application in the local public service of Dalian.

#### B. Construction Process

At its most basic, a corpus is a digital table extracting and collecting the linguistic data in accordance with certain criteria [4]. Since all the current linguistic data in the national standards are presented in PDF documents, the preliminary construction step of digitalization can be skipped. Therefore, the construction of the corpus for the public service areas of Dalian is conducted in the following 4 steps.

##### 1) Corpora collection

In accordance with related corpus extraction rules, all the linguistic data in *Guidelines for the Use of English in Public Service Areas (GB)* are extracted and converted into editable data by using some applications with PDF transforming and optical character recognition functions. The extracted data of general rules, sports, education, health & medicine, post & telecommunications, and commerce & finance are already prepared for further processing, while the data of transportation, tourism, culture & entertainment, and accommodation & catering are integrated with the local cultural and geographical characteristics of Dalian, generating the data suitable for the local public service areas in the city of Dalian. All the collected data are stored in Microsoft Word documents (with the file name extension of .docx).

##### 2) Corpora pre-Editon

Generally, corpora pre-edition is conducted through the three sub-steps of text denoising, text splitting and text alignment [5]. In the sub-step of text denoising, all the raw linguistic data in MS Word are saved as plain text (TXT documents with Unicode), removing all the redundant format and non-text contents. ICTCLAS (Institute of Computing Technology, Chinese Lexical Analysis System) is applied to split Chinese text, while Wordsmith is used to split English text

into terms. The plug-in of WinAlign in Trados is used to align text at word level.

### 3) Corpora annotation

Corpus annotation is the practice of adding interpretative linguistic information to a corpus. No matter adopting automatic or manual annotation, errors cannot be completely eliminated [6]. The tags of type, purpose, properties and part of speech are added to each individual word or term, providing evidence and criteria for data searching, inquiring and filtering.

### 4) Corpora storage

Finally, the corpora are inputted and stored in Tmxmall, generating a complete parallel corpus for the Chinese-English Translation for the public service areas of Dalian, which can be used as a library in Trados, supporting and facilitating the Chinese-to-English translation practice for the public service areas of Dalian.

## IV. PROBLEMS AND COUNTERMEASURES

Although all the linguistic data and corpora are processed and tagged in various aspects, some translation problems and errors are still found in practical application of the corpus. More attention should be paid to the specific context in practical translation.

### A. Subject and Its Properties

In practical English-Chinese translation, it is critical to consider the relationship between the subject and its properties. For example, the subject of “大连北站 (a railway station)” coheres with its properties as a railway station, so the correct translation of Dalian North Railway Station can be achieved from the corpus. As to “青泥洼桥 (a commercial area)”, there is not a real bridge around the place and its subject is irrelevant to its properties, so it should not be translated into *Qingniwa Bridge*, but *Qingniwaqiao*. Therefore, the tag of Properties should be added on the corpus in the corpora annotation step, which can effectively solve this problem and avoid mistranslation.

### B. Different Purposes

The name of the same facility should be translated differently according to its different purposes in practical application. For example, “残疾人卫生间” (Accessible Toilet) is quite common in public area. It should be translated into *Disabled Only* when it is used as a sign to direct people. Therefore, the tag of Purpose should be added on the corpus in the corpora annotation step, which can effectively solve this problem and avoid mistranslation. All the translation should be offered to users and tagged explicitly with different purposes.

### C. Complexity in Nature

Integrated with the local cultural and geographical characteristics, the Chinese-English translation in the public service areas of Dalian can be more accurate and precise, but more complicated as well. Some errors and mistakes in the translation are inevitable with the corpus, so it is critical to offer regular maintenance and improvement to the corpus, in order to continuously refine the corpus and improve the Chinese-English translation of the public service areas of Dalian.

## V. CONCLUSION

With the comprehensive and rigorous Guidelines for the Use of English in Public Service Areas (GB), the Chinese-to-English translation has been seriously and prevised guided. Thanks to the great achievements in previous studies and practice, the parallel C-E translation corpus for the public service areas of Dalian has been constructed with sufficient reference and experience. The corpus framework is designed in accordance with the specific rules in the national standards and the local cultural & geographical characteristics of Dalian. The C-E translation corpus for the public service areas of Dalian has been constructed through the 4 steps of corpora collection, corpora pre-edition, corpora annotation and corpora storage. The corpus constructed in this study can be used as a library to support the development of the translation application for the C-E translation for public service areas of Dalian. In order to improve the E-C translation and avoid mistranslation, more specific tags should be effectively applied in the corpora annotation step.

## REFERENCES

- [1] LIN Kenan. The Necessity of Considering Reader Reaction: A Proposal for Improving Guidelines for the Use of English in Public Service Areas. *East Journal of Translation*. 2018. 2
- [2] ZENG Xiaojuan, Hu Libing, Zeng Ying. Research on the Standardization of English Translation of Public Signs in Wuhan City: A Study on the Local Promotion and Application of the National Standards of “Guidelines for the Use of English in Public Service Areas”. *The Science Education Article Collects*. 2018. 1
- [3] LI Jiakun, GUO Xinzhu, JIANG Zhiwei. Study on the Construction of Architecture Bilingual Parallel Corpora and Cultivation of Students' Critical Thinking Ability in MTI Teaching. *Journal of Shenyang Jianzhu University (Social Science)*. Vol. 20, No. 5. 2018.10
- [4] LIANG Maocheng, LI Wenzhong, XU Jiajin. *Using Copora: A Practical Coursebook*. Foreign Language Teaching and Research Press. 2018. 7
- [5] HUANG Yong. *Construction and Application of Parallel Corpus for Translation Criticism*. Overseas English. 2014. 8
- [6] DOU Zhi. Construction of the English-Chinese Parallel Corpus of American Drama Subtitle. *Management & Technology of SME*. 2018.7