A Study on Corpus and Translator Training in Power Sector

Wang Linyun
North China Electric Power University
NCEPU
Beijing, Changping
cyan_ly_wang@163.com

Abstract—**In the ever-expanding power market, the demand of technical translators in power sector enlarges. However, the imbalance between supply and demand in the power market causes problems. To solve these problems, this paper proposes to establish a power corpus used for cultivating translators majoring in both English and electricity-related subjects. After a brief introduction of corpus, this paper expounds the characteristics of and difficulties of electric power translation, presents the design of corpus according to these features, thus helping cultivate power translators with academic and practical skills. The corpus can be used by translators with electrical-related background to polish their languages and English majors to grasp basic knowledge of electric sector. The corpus can find solutions to most problems, but it still requires continuous improvements on the basis of user testing.**

Keywords—Power corpus; Translator; Characteristics; Difficulties; Design

I. INTRODUCTION

With the implementation of Belt and Road Initiative, China's power companies engage more in commodity export, overseas investment and international engineering contract.

According to China Language Service Industry Development Report 2016, the output value of the language service industry reached more than 300 billion yuan in 2017 [1], and professional division of translation companies is becoming increasingly obvious [2]. However, statistics from *China Regional Translation Industry Development Survey Report 2017* shows that translation companies specializing in the electronics, electrical and automation industries accounted for only 7.3% [3], which means that when professional translation is needed in fields like electricity and machinery, the customers have a relatively small selection range. It will be more difficult to complete in time when there is a heavy task.

To conclude, obtaining high-quality power translation has become a primary concern for many enterprises in overseas expansion.

II. POWER CORPUS

A. The Function of Power Corpus

The study on corpus has gradually developed in China since the 1980s. It has been widely used in linguistic research and translation teaching in recent years. Corpus has several advantages: (1) the materials stored in it are carefully selected and processed, (2) easy searching, (3) timely updating. In order to solve the problems mentioned, building a power corpus is of great benefit to the training of electric translators Maintaining the Integrity of the Specifications.

B. The Importance and Necessity of Building a Power Corpus

Nowadays, combining technology has become a trend for different fields to develop. Therefore, the use of corpus to assist translation is an initiative to keep up with the times.

1) Combining translation and technology is an irresistible trend:

Schäler believes that future translators must gradually resolve the conflict between traditional occupational value systems and new technologies [4]. In other words, with continuous development of science and technology and transformation of the training of translators, it is necessary to incorporate new technologies into translation learning.

Jiang Xiaolin also pointed out that language service talents must have professional background as well as technical capabilities [5]. Nowadays, BTI and MTI are aimed to improve the ability of students to control language quality. In the learning of corpus-aided translation, students can learn how to manage terminology, establish translation memory, and build a corpus for learning translation. This will not only improve the translation efficiency of students, but also plays an immeasurable role in their future translation career.

2) Corpus courses are offered in universities all over the world:

According to *MTI curriculum system under the technical service industry perspective*, the importance of corpus can be seen for CAT and translation learning, here are relevant courses in universities around the world [6]:

TABLE I RELEVANT COURSES IN UNIVERSITIES AROUND THE WORLD

Relevant Courses in Universities Around the World	
University	Courses related
Imperial College London	Lexicography, Terminology, Corpus, Machine Translation, Computer-aided Translation, Subtitle Translation, etc.
University of Leeds	Computer-aided Translation, Principle and Application of Machine Translation, Film Translation and Dubbing, etc.
Dublin City University	Translation Technology, Localization of Software, Audiovisual Translation, Corpus Linguistics, etc.
Université de Genève	Language Planning, Machine Translation, Corpus and Translation, Localization and Project Management, Computer-aided Translation, etc.
The Chinese University of Hong Kong	Translation and Language Technology, Computer-aided Translation, Corpus and Translation, Terminology Management, Workshop of Localization, etc.

III. FEATURES AND DIFFICULTIES OF POWER TRANSLATION

Being a branch of scientific translation, power translation has many characteristics of it, as well as some unique features.

A. Demanding requirements about Terms

1) High verification ability:

Continuous verification is required in power translation. Conventional search tools include: paper dictionaries, electronic dictionaries, search engines, etc. However, these tools have their own shortcomings, for example: the contents of paper dictionaries are relatively fixed and the update period is quite long; the contents in electronic dictionaries are mostly from paper dictionaries; the search engine can be updated continuously but the technology is still not mature enough, which requires multiple searches for the inclusion of entries, and it is "unsupervised". It may take a long time to find usable information but the accuracy of these materials cannot be guaranteed.

2) Common words have scientific meanings:

Generally, word meaning varies in different contexts. In translation, terminology with solely scientific meaning is not the most difficult to verify, but words that apply to multiple fields, that is, when common words contain scientific meaning or a word can be used in multiple fields. Here is a case:

Example1: For the electrical wiring, there are two wires which connect the main breaker panel to the electrical meter outside, or there are breaker levers which are fused together.

Translation:布线方法有两种，一种是将主断路器面板与外部电表相连，另一种则是将其与断路器拉杆融到一起。

According to Oxford Advanced Learner's English-Chinese Dictionary，"wire" has four meanings as a noun: (1)金属丝，金属线；(2)电线，导线；(3)金属丝编制的栅栏，铁丝网；(4)电报. None of the four meanings are in complete correspondent with "布线". Therefore, translators are required to compare different meanings and combine it with professional knowledge.

3) Different forms of terminology:

Compared with Chinese classification of words, English uses diversified forms for words. So the same meaning may be expressed by multiple different forms, phrases or even sentences, as follows:

Example2：a transformer with inherent current limitation

Original：有电流限制的变压器

Improved：限流变压器

Example3：The transformers are dry.

Original：这些变压器是干的

Improved：干式变压器

From these two examples, we can see that the "XX Transformer" with uniform format in Chinese may be a phrase or even a sentence in English, which adds more difficulties to translation.

B. Translation requires adjustment in different contexts

Many people new to translation always stick to the superficial meaning of original expressions and cannot express hidden meanings. However, all words, phrases, and sentences should be subject to its context, so that the translation can be fluid and easy to understand.

Example 4: The electric power industry is commonly split up into four processes.

Translation: 电力行业的生产过程通常分为四个环节。

As in Example 4, if you want produce idiomatic translation, you need to study the expression of both English and Chinese.

C. Translators need some background knowledge

As far as ordinary translation practitioners are concerned, the reason why scientific texts are difficult to translate is because their knowledge of this field is not enough. Therefore, sometimes they cannot understand the original text or they can't express idiomatically even if they can understand it, which causes so-called translationese. It is safe to conclude that only when a translator has certain background knowledge can he better understand the original text and then produce idiomatic expression.

Example 4：For the most part, electrical energy from the grid cannot be effectively stored, meaning that production must follow demand.

Translation: 主要原因是，电网中的电能无法有效储存。这也就意味着即时产出的电量需要跟得上即时的需求。

Analysis: (1) "for the most part" is usually translated as "in most cases". But as long as a person has a common sense of electricity, he should know that electricity in the grid cannot be stored. We can find that literal translation may lead to factual mistakes, so post-translation checks are necessary. (2) For the second underlined part, the "production" is translated into "the instantaneous output of electricity". This translation is based on the understanding of the whole text and the reserve of power knowledge, and considers the readability of translation as well.

IV. THE DESIGN OF POWER CORPUS

According to the features and difficulties of power translation mentioned above, we designed a power corpus.

A. Basic Module

The index of our corpus consists of five modules, analytical processing tools, entrances to academic and pragmatic power materials, download links of materials, and channels for uploading materials and recent corpus-related news.

- Analytical processing tools: it includes some tagging tools, used for further analyzing the corpus from the perspective of linguistics.

- Entrances to academic and pragmatic power materials: materials stored in the academic part are mainly academic papers, conference papers, professional books, etc. while those in the pragmatic part are some texts often used in companies, such as instructions, business letters, etc.

- Download links of materials: this part includes information about materials stored in the corpus, such as: author, publication, etc. and download links, thus facilitating corpus users to learn from textual approach.

- Channels for uploading materials: Since the corpus uses point redemption scheme, it is easier for corpus developers to accumulate materials by opening the uploading channel. For the users, the accumulated points can be used as access to advanced functions.

- Recent corpus-related hotspots: lead language learners to follow up hotspots and then come up with new ideas.

B. Basic Module

1) Dynamic corpus: after completion of the corpus, developers still can add new materials. What's more, the corpus can be improved through feedback from users, so more people can participate in corpus improvement.

2) The purpose of power corpus: apply it to translation teaching, translator training, and power-related papers reading.

3) The source of materials: textbooks on electricity (criteria: wide range of use, authentic editorials, recommended materials from the Ministry of Education); papers published by core journals such as IE and IEEE; specifications, contracts, tenders, etc. of power companies (the need for the interpreters is not considered for the time being).

4) Terminology database classification: there are two main categories, namely academic and pragmatic database.

5) The proportion of databases: monolingual materials take the majority, which is supplemented by the bilingual materials, taking up about 30%.

6) Diachronic corpus: used to explore linguistic features of power translation in different periods and the hidden reasons, so as to provide technical guidance for translation.

7) Type of languages: In the early stage of construction, written language is dominant, and then some oral materials will be gradually added.

8) Quality of languages: First of all, it can be extracted from authoritative journals; secondly, it can be obtained from practice materials of power translation course, and use teachers' translation, whose advantage is that the manually processed translations are more accurate and readable. Finally, we can directly ask teachers with relevant experience whether there are any monolingual or bilingual materials available, which is the easiest and fastest way.

C. Design according to features of power translation

1) Analogy learning method: The essence of translation is to convey the information of source language into target language as much as possible. Therefore, there is a commonality for the translators to refer to, which can be used for most problems. Therefore, the learner can compare and contrast the content in the bilingual corpus to learn.

2) Corpus-aided learning method: Characteristics of different genres vary. Some unique problems can be solved by using corpus.

a) The difficulty of verification: In the respect of selecting proper nouns, corpus collection and processing can start from the following aspects: (1) Find electric power terms that meet the national standards, and keep updating (2) The corpus contains as many relevant papers as possible in authoritative journals (with English abstracts),

then readers can search articles by inputting keywords, find parallel texts, and then the developers can update the materials in the corpus according to word frequency. (3) Attach photos to some easily confused concepts, add some explanative words to help users to determine the object or the translation.

b) Common words have scientific meanings: In the web page, first list the meanings of the same word, and then classify the words according to their domain, such as: general knowledge, technology, etc. Then, by means of imagery thinking, place various meanings in the same block of different pages according to different fields, as shown in Figure 1.
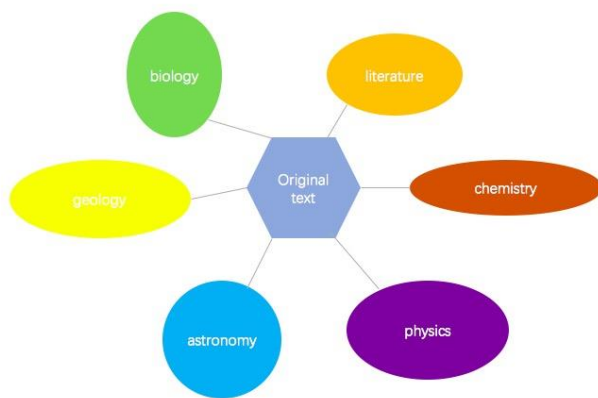

Fig. 1   Words divided into different fields

c)   Various expressions of terms: First, list the sample sentences of the word in different sentences, so that the learners can understand the difference of these words in each sample sentence; for the convenience of analysis and searching, words can be marked red or text highlighted.

d)   Sentence translation: We ought not to be constrained by the grammatical form of the original text, and not to be too "Europeanized". It is necessary to focus on the logical relationship between sense-groups of original text, and appropriately adjust the word order and perspective to make the expression more idiomatic.

e)   Textual translation: Translators need to pay attention to discourse, because the same concept may appear repeatedly in a text. The discourse can help us understand the text and choose the most appropriate wording. If we can't understand a sentence for the time being, we can continue reading and the following part may give some indications. Therefore, it is necessary to enter the author, journal, time and other information when collecting materials, which may facilitate users' learning and verification later.

f)   Background knowledge: According to different purposes, the classification of practical part has more explanations than the classification of academic part, such as a word (representing a loop) will be presented with a structural diagram next to it to indicate the cycle process, so that non-professional students can also learn some professional knowledge. Besides, the sources, and some pictures, related introductions, and video links (especially explanations for a certain process) will be added. In addition, users and developers can follow up hotspots, innovations, new inventions, etc. in related fields, and regularly help update the contents of the corpus.

D. Other functions

- The frequency of search items will be ranked from high to low, as a reference to added materials later.

- The teacher-student interactive platform is established for students to upload assignments, of which scoring can be assigned to the platform scoring system or the teacher. The platform can not only serve for teachers to give feedback of students' translation, but also can act as a first-hand material for teachers and linguistics to study students' mistakes in translation.

- On the basis of the second point, translation correction procedure is introduced. The translation and correction process can be divided into two modules: one for students to remark on each other's assignment and the other for teacher to correct students' assignment. The first module is open to all users, allowing everyone to learn from mutual remarks. The reward mechanism will encourage people to actively participate: multiple proofreading or question points will award points, which can be used for advanced functions of other modules in the corpus. In the long run, the mechanism will not only guarantee the correctness of the materials in the corpus, but also improve users' enthusiasm and participation.

V. THE CONSTRUCTION OF CORPUS

A. Collection and Processing of the Materials

Materials in the corpus are divided into two different parts, academic and pragmatic. Take the academic part as an example: firstly, we select the designated textbooks for electrical major. The steps are as follows: (1) Take photos of the materials and convert the paper version into an electronic version (this step is more convenient if there is an electronic version of the material); (2) Convert the photo/PDF or other forms of documents into word format with ABBYY; 3. Organize the text and correct errors according to the original text; 4. Clean the text, and remove the punctuation in the original text; 5. Align the text, and use Aligner to align

Chinese and English; 6. Tag the materials and store them in the terminology database.

B. Build a corpus:

The corpus includes a terminology database, a multimedia database, a material database (my SQL, oracle, etc.), and then an entity relationship can be established among the three databases

VI. STRENGTH AND WEAKNESS

The design proposed in this paper has solved the following problems:

- People new to translation may find it difficult to get started;

- The difficulty in finding meanings of certain professional items;

- The lack of materials for teachers' after-school exercises and student feedbacks and platform of teacher-student interaction;

- The disparity between different requirements for searching data;

- The materials in the database is not so authentic or professional;

- Understanding of and access to the frontier professional articles.

But there are a lot to improve:

- The materials in this corpus is suitable for translation learning, but not so useful for interpreting learning;

- The corpus is more focused on basic construction, emphasizing the collation and classification of existing materials, but lack detailed research on more extended functions;

- Some steps are cumbersome in the converting part during the construction of corpus, more convenient ways are needed to improve the efficiency.

VII. CONCLUSION

Combining corpus to cultivate professional translators in power sector has become a trend, but it is also necessary to consider the characteristics of this specialty and solve these problems and prevent potential risks. The corpus proposed in this paper solves most of the problems, and there are a lot to be improved. Being a dynamic corpus, it can cultivate power translators through constantly updating, and perfecting. It is hoped that it will contribute to China's power development in the future.

REFERENCES

[1] http://www.gov.cn/xinwen/2016-12/23/content_5152184.htm.

[2] Wang Bin. "Let every fiery heart beat with you - the 2007 Chinese translation market year-end inventory". 2008:361-363. (In Chinese)

[3] Huang Tianliang, "SUNTHER Translation Company Competitive Strategy Research". Shandong University, 2011. (In Chinese)

[4] Sch äler, R. The Problem with Machine Translation [A]. InL. Bowker et al. (eds.). Unity in Diversity? Current Trends in Translation Studies[C]. Manchester: St. Jerome Publishing, 1998.

[5] Jiang Xiaolin. Quality Requirements for Language Service and Translation Talents [A]. Proceedings of the First National Symposium on Translation Masters and Translation Industry [C]. Beijing: School of Foreign Languages, Peking University, 2009. (In Chinese)

[6] Wang Huashu. "The construction of MTI technology course system under the technical field of language service industry [J]. Chinese Translation, 2013, 34(06): 23-28+128. (In Chinese)