

# Research on the Establishment and Application of Petroleum English Terminology Corpus and Translation Memory Bank: A Case Study of Yangtze University

Linfeng Wu<sup>1\*</sup>

<sup>1</sup> School of Foreign Languages, Yangtze University, Jingzhou, China

## Email Address

1347329308@qq.com (Linfeng Wu)

\*Correspondence: 1347329308@qq.com

Received: 9 April 2021; Accepted: 21 April 2021; Published: 21 May 2021

---

## Abstract:

Yangtze University is a comprehensive university, and its petroleum major has a high status in the whole of China. In recent years, the Petroleum School of Yangtze University has more and more frequent foreign exchanges, and English is mostly used in foreign exchanges. Therefore, it is necessary to improve the accuracy of English terms and the efficiency of Chinese-English translation. The growing prosperity of import and export trade and the demand for translation has witnessed a huge increasing. The translation efficiency and quality put forward higher threshold, therefore the need for uniformity and accuracy of terminology translation [1] is of particular significance. At present, most petroleum enterprises use machine translation directly, but due to the lack of professional translation software for petroleum terms and the translation memory database, it is difficult to achieve the consistency in the overall content, style and terms of translation [3]. The purpose of this paper is to study the effects of the use of oil-related corpus and translation memory on the efficiency of translation, the reduction of repeated work, the demand of the market and the Times. At the same time, this paper is committed to serving the academic research and foreign exchange and cooperation related to the petroleum field of Yangtze University, and improve the accuracy and authority of the translation of Jiangnan Oilfield and other oil fields in the foreign cooperation to a certain extent, which has a certain academic value and practical value. This paper takes the computer aided translation software Trados as the carrier [5] to improve the accuracy and efficiency of petroleum translation by importing the terms collected and the translation memory database.

## Keywords:

Translation Memory Bank, Petroleum English Terms, Computer Aided Translation, Translation Efficiency, Post-Translation Edition, Trados

---

## 1. Introduction

Machine translation is playing an increasingly important role in today's world, but machine translation inevitably has fatal disadvantages. How to improve the accuracy of machine translation is a question worth thinking about [12]. The foreign exchanges of China's petroleum enterprises are more and more frequent, and China's import and export trade is increasingly prosperous. With the rapid growth of the demand for translation market, higher requirements have been put forward for the efficiency and quality of translation. Therefore, the accuracy and uniformity of the translation of relevant terms in the petroleum field need to be regulated. Efficient and accurate translation of petroleum terms is conducive to promoting the cooperation among the world's petroleum enterprises and realizing the effective allocation of global petroleum resources.

The construction of memory bank has the following functions: (a) It can realize the storage of translation resources of petroleum, so as to realize the sharing of translation resources among enterprises; (b) Liberate the translator from the tedious and repetitive work [11], and improve the efficiency and quality of translation; (c) To complete the translation of scientific and technological literature and related materials with the fastest speed and the best quality [8], so as to realize the standardization and efficiency of petroleum translation.

In order to improve the accuracy and speed of translation, the author used a computer aided translation software called Trados and created a term database and translation memory database on it. It is a daunting task to create a corpus and a translation memory bank. It took the author more than two months to collect more than 60,000 pieces of petroleum terms and create the memory bank.

In continuous practice, the translation memory database has been improving. In the process of creating a glossary, the accuracy and rigor of terms are very important, and post-translation editing is also an important part after translation [9]. The ultimate purpose of the translation memory bank is to serve the practice. The author will apply the translation memory bank to the foreign exchanges of Yangtze University and related enterprises in Jingzhou. At present, due to various constraints, there are still many things out of my reach. The author will strengthen the research on computer-aided translation and make the outcome available to more people.

## 2. Computer Aided Translation and Trados

Computer-aided translation technology has a relatively short history in China. In foreign countries, related research began in the 1980s. When it was first developed, due to the constraints of algorithms and other conditions, it can only provide a certain degree of understanding and reference for the translator, but can not produce a reasonable, smooth and ideal translation [3]. With the development of artificial intelligence, big data and other technologies, CAT technology has made continuous progress and achieved great achievements in recent years. Various translation websites and software based on CAT technology have greatly reduced the labor intensity and difficulty of traditional translation work. At the same time, with the development of computer storage function, cloud computing and big data, the capacity and content richness of translation memory database will be greatly improved. At that time, CAT technology will have a real development boom, and it will be possible to replace human translation completely.

Trados, a typical computer-aided translation software, has more than 200,000 customers worldwide. 90% of the global top 500 companies are using Trados for daily localized translation work. Because it is based on the principle of translation memory, Trados is currently the world's best professional translation software, which has become a professional tool in the field of translation standards, its background is a very powerful neural network database and information security guarantee system. Trados also has perfect auxiliary functions, such as time, metrics and form, automatic replacement fixed format, and can help customers to greatly improve their work efficiency.

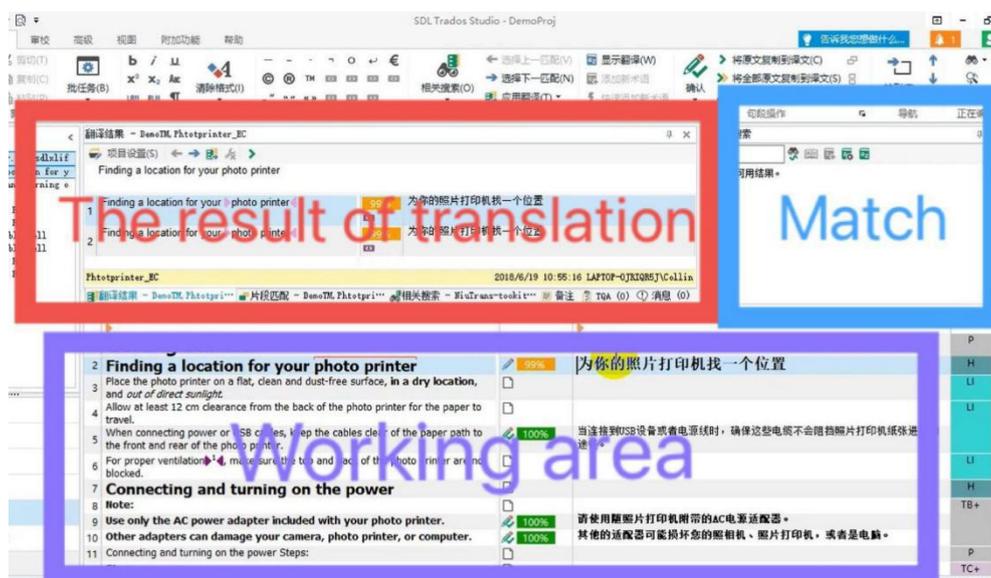


Figure 1. The working interface of Trados.

As can be seen from the picture, Trados has three working areas, among which the upper left is the translation memory area. The top right is the term matching section, and the bottom is the section for processing the translated text. As for the concept and function of each area, which will be introduced later.

### 3. Petroleum English Terminology Corpus and Translation Memory Bank

In Trados, users can create their own corpus and import the collected terms into the corpus. In every text translation, users can import their own corpus, which can effectively avoid the problem of inaccurate professional terms in simple machine translation.

At present, more than 60,000 petroleum terms have been collected, covering 40 fields related to oil. The collected terms are imported into Trados and carried out translation practice and application research. The terms collected are from authoritative websites in related fields at home and abroad. Petrowiki.com provides me with a wealth of petroleum jargon and detailed explanations, and Rigzone.com keeps me up to date with the latest industry news. The terms of petroleum industry can be easily searched on the Internet, but the terms and vocabularies in the petroleum industry standard cover the upper, middle and lower reaches of the petroleum industry, including exploration, development, refining, storage and transportation and other professional categories. At present, there is no independent terminology standard for 50% of the petroleum sector. Terms are scattered among various standards, which is

difficult to query and is prone to the phenomenon of repeated definitions of terms. Taking offshore petroleum engineering as an example, there is no term standard, but all terms in the standard reached 1,171. Because these terms are scattered in various standards, it is difficult to reach the consistency, and it is easy to repeat the definition of the same term, so the boundary between terms becomes blurred. For this reason, a month has been spent to collect relevant terms of petroleum, and carefully revised the collected terms to ensure the accuracy and rigor so as to facilitate later translation practice.



The image shows a screenshot of a spreadsheet with a list of 47 technical terms. The spreadsheet has a header row with '1' in the first column and '>>L<<English' in the second column. The terms are listed in the second column, numbered 1 through 47 in the first column. The terms are: abaxial, ablation, abnormal liquid, abrasion, abrasion hardness, abrasion test, abrasion tester, abrasives, abrupt perturbation, abscissa, absence of collision, absence of gravity, absence of stress, absence of vortices, absolute, absolute acceleration, absolute angular momentum, absolute ceiling, absolute coordinates, absolute deviation, absolute elastic body, absolute equilibrium, absolute extremum, absolute flying height, absolute frequency, absolute geopotential, absolute humidity, absolute instability, absolute motion, absolute path, absolute perturbation, absolute pressure, absolute rest, absolute rotation, absolute similarity, absolute stability margin, absolute temperature, absolute thermometer, absolute velocity, absolute viscosity, absolute vorticity, absolute weight, absolute zero, absolute zero point, absorbability, and absorbing agent.

1	>>L<<English
2	abaxial
3	ablation
4	abnormal liquid
5	abrasion
6	abrasion hardness
7	abrasion test
8	abrasion tester
9	abrasives
10	abrupt perturbation
11	abscissa
12	absence of collision
13	absence of gravity
14	absence of stress
15	absence of vortices
16	absolute
17	absolute acceleration
18	absolute angular momentum
19	absolute ceiling
20	absolute coordinates
21	absolute deviation
22	absolute elastic body
23	absolute equilibrium
24	absolute extremum
25	absolute flying height
26	absolute frequency
27	absolute geopotential
28	absolute humidity
29	absolute instability
30	absolute motion
31	absolute path
32	absolute perturbation
33	absolute pressure
34	absolute rest
35	absolute rotation
36	absolute similarity
37	absolute stability margin
38	absolute temperature
39	absolute thermometer
40	absolute velocity
41	absolute viscosity
42	absolute vorticity
43	absolute weight
44	absolute zero
45	absolute zero point
46	absorbability
47	absorbing agent

*Figure 2. A screenshot of the terms the author collected.*

Translation memorization is a computer aided translation technique to avoid repeated translation and improve translation efficiency [4].

Specifically, The translation memory bank can dynamically store the sentences translated by the translator [11]. In the next translation, if the sentence with a high repetition rate is encountered, the translation memory will automatically prompt for

previously translated sentences. Translator may use a previously translated sentence in part or in whole, depending on the degree of similarity.

Then, the translated text is guided to the translation work platform. Based on different software, sometimes it is necessary to deal with terms extraction, file format conversion and other work in advance, and then start the translation. Terms can be generated manually or automatically during translation, which can be used to refer to the memory bank, or the relevant glossary can be imported before the translation project, and the translation can be exported and generated in the corresponding format. At present, most of the translation software commonly used in the market belong to the category of machine translation. Due to the complexity of language grammar, semantics and style, the accuracy and practicability of its translation is far from human translation. In this case, computer-aided translation makes use of the memory function to reuse the translated resources, so that the translator can avoid translating repeated information, thus improving the efficiency and the quality of the translation.

In a word, translation memory bank plays an increasingly role in translation and has become a trend of translation.



*Figure 3. The sources of petroleum terms.*

### **3.1. The Importance of Post-Translation Editing**

In recent years, due to the development of software technology and the emergence of new algorithms, machine translation technology has made great progress [1]. Computer-aided translation has been recognized by both the translation industry and clients as an important part of an efficient translation process. However, the quality of a machine translation is far from comparable to that of a human translation. Post-translation editing can effectively achieve the balance between translation and efficiency and give full play to the advantages of human-computer interaction.

Translation errors in machine translation mainly come from logic and structural errors of complex sentence patterns, such as ellipsis, reference, syntactic structure conversion errors, and wrong forms. Specifically, a high quality post-translation editor

needs to pay attention to whether the translation contains errors in semantics, grammar, pragmatics, terminology, spelling, punctuation, symbols, numbers, format, addition, omission, ambiguity, consistency, cultural conflict, and so on.

Trados can handle post-translation editing efficiently. After each translation, one can modify the translated paper to make the translated text more accurate. This process of modifying the translated text is called post-translation editing. Trados can import the modified text into the translation memory database, which can reduce the workload of translation and improve the efficiency when translating similar text next time.

### ***3.2. The Establishment of Corpus and TM Bank***

How do you create a corpus? Firstly, create a blank term database on Trados, import all the collected terms into the corpus, and finally name the corpus.

Then the alignment of the corpus is carried out. Other computer-aided translation software, such as the snowman CAT, also has the alignment function, but also requires a series of formats to be converted into usable TXT text. To avoid such a series of operations, the current online corpus big data and intelligent translation platform can provide faster operations. For example, the intelligent alignment algorithm—WinAlign which is a built-in function of Trados can quickly and automatically align the original text with the translation, automatically identify the complex sentence correspondence relationship, and greatly improve the efficiency and accuracy of alignment. The interface of adjusting alignment is intuitive and easy to operate. Users can easily preview the alignment corpus online. The platform supports the import of 36 mainstream formats such as DOCX, XLSX, PPTX, PDF and TXT, and can directly export four formats such as TMX, XLSX, TXT and DOCX. With the preparatory work done, the glossary is ready to be put into use.

The establishment of translation memory bank is similar to that of corpus, so it will not be explained in detail here. It is worth mentioning that the retrieval and matching of the translation memory bank can help the translator to match the previously translated texts before translation, so as to find the same or similar sentence patterns in these texts for reference [4]. The outstanding advantage of the translation memory bank is that it can increase efficiency and reduce repetitive work.

## **4. Current Progress and Existing Problems**

At present, the author has used the corpus and translation memory database to translate more than 200 petroleum literatures. Meanwhile, the author constantly expanding the capacity of the corpus to make the scope of the corpus larger. At present, the research results are still in the early stage, and some relatively simple translation experiments are carried out with satisfactory results. With the continuous development of corpus and translation memory database, translation will be involved in a wider range of fields. Recently, the author has been using the corpus and memory bank to translate the petroleum literature of Yangtze University, and achieved good fruition. Computer aided translation saves a lot of time and improves the efficiency of translation. The complex literature has not been covered at present, and will be as the research progresses.

There are still some challenges. Even though the author found 60,000 petroleum terms, the number and scope of terms collected so far is relatively small. The collection of terms, the creation of corpus and translation memory database, and the practical application of Trados were all completed by the author himself. Due to

limited time and energy, there are still many inadequacies in this study that need to be adjusted and improved gradually.

## 5. Future Goals

The author will enhance terminology collection and collect terminology from multiple sources. Checking the collected terms to make sure they are correct.

To strengthen post-translation editing, the author needs to have a positive attitude to understand and agree with the translation methods of machine translation output, figure out the basic principles of machine translation, straighten out the error types of machine translation, and get familiar with the projects, fields and terms involved, as well as the translation environment [12]. Through constant practice, you will improve your ability to read quickly, to identify major machine translation errors that affect the reader's understanding, and to modify translations that are contrary to the original author's intent [2].

According to the quality objectives and personnel skills, the post-translation editing workload should be reasonably set. Post-translation editing of machine translation is a professional profession. With the rapid increase of translation market demand, the shortening of translation delivery time and the continuous progress of translation technology, post-translation editing of machine translation will play an increasingly important role in the practice of language service. In order to give full play to the high efficiency of machine translation, the high quality of human translation and provide more professional language services [11], the combination of machine translation, human translation post editing and translation project management will be the trend of translation development in the future. To this end, while learning professional knowledge, the author also needs to understand the knowledge of post-translation editors to create a rigorous translation memory database.

## Conflicts of Interest

The author declares that there is no conflict of interest regarding the publication of this article.

## Funding

This work was supported by the Project Fund of the Innovation and Entrepreneurship Project for College Students in Hubei Province (2020) for the research of Research on the Establishment and Application of the Translation Memory Database for Petroleum and Mechanical Equipment Terms. Grant number: S202010489080.

## References

- [1] Chiocchetti, E.; Wissik, T.; Lušický, V.; Wetzel, M. Quality assurance in multilingual legal terminological databases. *The Journal of Specialised Translation*, 2017, 27, 164-188.
- [2] Gural, S.K.; Chemezov, Y.R. Analysis of Efficiency of Translation Quality Assurance Tools. *Procedia - Social and Behavioral Sciences*, 2014, 154, 360-363.

- [3] Seljan, S.; Dunđer, I. Machine Translation and Automatic Evaluation of English/Russian-Croatian. In Proceedings of the “Corpus Linguistics 2015” Conference at the St.Petersburg State University, 2015a; pp. 72-79.
- [4] Seljan, S.; Pavuna, D. Translation Memory Database in the Translation Process. In Proceedings of the 17th International Conference on Information and Intelligent Systems IIS 2006; pp. 327-332.
- [5] Yang, J.; Ciobanu, D.; Reiss, C.; Secară, A. Using Computer Assisted Translation Tools’ Translation Quality Assessment functionalities to assess students’ translations. *The Language Scholar Journal*, 2017.
- [6] Zhang, L. A Study on the Efficiency Influence of Computer-Assisted Translation in Technical Translation. *Journal of Chemical and Pharmaceutical Research*, 2016, 8(4), 801-808.
- [7] Garcia, I. Power shifts in web-based translation memory. *Machine Translation*, 2007, 21, 55-68.
- [8] Christensen, T.P.; Schjoldager, A. Translation-memory (TM) research: What do we know and how do we know it? *Hermes – The Journal of Language and Communication*, 2010, 44, 1-13.
- [9] Aziz, W.; Koponen, M.; Specia, L. Sub-sentence level analysis of machine translation post-editing effort. 2014.
- [10] In S. O’Brien, L. W. Balling, M. Carl, M. Simard, & L. Specia (Eds.), *Postediting of machine translation: Processes and applications* (pp. 170–199). Newcastle: Cambridge Scholars Publishing.
- [11] Banerjee, S.; Lavie, A. METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. Paper presented at the Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan.
- [12] Liu, Q.; Zhang, X. Machine translation: General. In S. W. Chan (Ed.), *Routledge encyclopedia of translation technology*. London: Routledge. 2015; pp. 105-119.



© 2021 by the author(s); licensee International Technology and Science Publications (ITS), this work for open access publication is under the Creative Commons Attribution International License (CC BY 4.0). (<http://creativecommons.org/licenses/by/4.0/>)