Original Research Article

Analysis of cat tools and its technology trend research

Yuanfu Shen

The Hang Seng university of Hongkong Shatian, Hongkong sar ,999077, China

Abstract: This paper takes a famous article as English-Chinese translation sample by using online tool, thus analyze the tool's features. We also make a comparison of two online translation tools and result from this, proposal is made if we want to develop our own translation tool. Last part we analyze the technology trend of machine translation.

Keywords: Translation tools; Feature; Trend analysis

1. Introduction

We select a famous book (Three Days to See) and choose the most inspiring part to translate by using Youdao, DeepL and some AI tool. In general, AI tool's performance is the best but still have room to improve so we have to adjust sentence by sentence thus we get the following final translation. Actually we found more than ten Chinese versions of this book and we also checked one by one. All these translation versions are good enough but we still did not adopt any of them directly as we found some minor issues which we propose to revise. We choose the following paragraph from the case to check and evaluate the machine translation.

Source Text:

Friends who are near to me I know well, because through the months and years they reveal themselves to me in all their phases; but of casual friends I have only an incomplete impression, an impression gained from a handclasp, from spoken words which I take from their lips with my finger tips, or which they tap into the palm of my hand.

2. Tool evaluation of deepL

2.1. DeepL translation strong points

2.1.1. High translation quality

The section headings are in boldface capital and lowercase letters. Second level headings are typed as part of the succeeding paragraph (like the subsection heading of this paragraph). All manuscripts must be in English, also the table and figure texts, otherwise we cannot publish your paper. Please keep a second copy of your manuscript in your office. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use.

1) **Accuracy not inferior to human translation:**

Accurately conveying the original message. DeepL excels in translation accuracy, dealing well with complex grammatical structures and subtle semantic differences. For example: ‘ I know their personalities, of course,

through other means, through the thoughts they express to me, through whatever of their actions are revealed to me', DeepL translates it as " 当然 , 我可以通过其他方式了解他们的个性 , 通过他们向我表达的思想 , 通过他们的任何行为 ", which retains the precision and fluency at the same time.

2) **Advanced customization:**

Applying unique terminology in translations through the glossary function. DeepL also has some advanced features, such as the glossary function, which allows users to customize the way specific terms are translated. This is especially important for translating academic papers, research reports and other documents, and ensures that the terminology is translated consistently and accurately.

3) **Advanced customization:**

Wide range of language support: content can be translated into more than 30 languages.

2.1.2. Enterprise-level security

Data protection: no data will be stored without user consent.

Compliance and certification: DeepL complies with ISO 27001 and GDPR standards.

Secure access: using single sign-on (SSO) and centralized management to simplify user management and ensure their security.

2.1.3. Powerful translation functions

Document Translation: Translate the whole document, retaining the original format of PDF, Word files, etc.

Precise editing: use the editing mode to scrutinize the translation to ensure that the final result meets the user's requirements.

Flexible alternatives: choose the best translation solution according to the needs of the audience to improve the communication effect.

Although DeepL Translator does a great job in many ways, it still has room to improve.

2.2. Deepl translator weak points

2.2.1. Limited free access

The free version of DeepL has a limit of 1500 characters per task. In order to unleash the full potential of DeepL, including access to the API and higher character limits, as well as for translating large amounts of text, users need to subscribe to a paid plan.

2.2.2. Not applicable for all languages

DeepL may not meet the needs of users who need to translate niche languages. While DeepL performs well in many languages, it is not equally good for all language combinations, and some users may find differences in translation quality.

2.2.3. Lack of understanding of specialized areas

DeepL may not be as accurate as some customized translation services when dealing with specialized terminology and industry-specific content, and therefore may not be able to provide accurate translations in certain specialized areas or industry-specific terminology. In such cases, a human translator with specialized domain knowledge may still be required to provide an accurate translation. For example, 'Court records reveal every day how inaccurately "eyewitnesses" see', when translated using the DeepL translator will be ' 目击者 ' as "eyewitnesses", whereas in legal terminology it should be translated as " 目击证人 ".

2.3. Suggestions for feasible improvements

To address the issue of translation accuracy, DeepL can continue to strengthen the research of machine learning and deep learning technology to improve the accuracy and comprehensiveness of translation. At the same time, it can consider introducing more corpus resources, such as spoken language corpus, dialect slang corpus, etc., in order to improve the practicality of translation.

Aiming at the optimization of user interface and experience, DeepL can add shortcut keys, optimize the operation flow, improve the response speed and so on, in order to improve the user experience. At the same time, DeepL can consider adding some auxiliary functions, such as displaying real-time translation results, managing translation history, etc., to facilitate users' use and management.

In response to the demand of some users for more language options and higher-quality language pairs, DeepL should increase its R&D investment to improve the performance and accuracy of the translation engine to meet the needs of different users. At the same time, it can consider cooperating with language service providers around the world to add more high-quality language pair options.

3. Differences between Youdao Dictionary and DeepL in terms of functional features and translation quality

3.1. Differences in functional features

3.1.1. Youdao dictionary

Good multi-platform compatibility: When browsing an article on a mobile phone, the Youdao Dictionary APP can conveniently perform translations. And reading the web version of the article on a computer, the web plug-in of Youdao Dictionary can also quickly provide translation results.

Rich functions: Youdao Dictionary has a rich collection of example sentences. Moreover, its encyclopedia function enables users to obtain more relevant information when encountering unfamiliar words. The word list function is convenient for recording new words when translating articles.

3.1.2. Deepl

Support for multiple file formats: If an article exists in PDF format, DeepL can translate it well and maintain the original format, including paragraphs, punctuation marks, etc. This is very useful for handling complete documents.

Convenient browser extension: When searching for background information about this article or other relevant web content in the browser, using the DeepL browser extension can quickly translate the selected content if there is relevant English content on the web page.

3.2. Differences in translation quality

Example 1:

"I know my friends from the feel of their faces." This sentence is translated by Youdao Dictionary as " 我通过摸脸来认识我的朋友。", which is a rather plain expression. And DeepL translation is " 我通过抚摸朋友的脸庞来了解他们。", which is smoother and more natural.

Example 2:

"How much easier, how much more satisfying it is for you who can see to grasp quickly the essential qualities of another person by watching the subtleties of expression, the quiver of a muscle, the flutter of a hand."

It is translated by Youdao Dictionary as " 对于能看见的你们来说 , 通过观察表情的微妙之处、肌肉的

颤抖、手的挥动来快速把握另一个人的基本品质是多么容易、多么令人满意啊。”, the translation of phrases like “subtleties of expression” is not precise enough, failing to fully convey the delicate emotions and vivid imagery of the original text.

The sentence is translated by DeepL as “ 对于你们这些有视力的人来说 , 通过观察表情的微妙变化、肌肉的颤动、手的摆动来快速把握他人的基本品质是多么容易、多么令人满足啊。”, which can better reflect the vivid descriptions and delicate emotional expressions of the original text and has a more accurate understanding of complex sentence structures and vocabulary.

From these examples, when dealing with texts with strong literary and rich emotional expressions, DeepL is superior in translation quality, while the functional diversity of Youdao Dictionary also has its unique advantages in the process of reading and learning.[6,7]

A comparison between Youdao Dictionary and DeepL in terms of interface friendliness, word limit, charging standards, advertisement interference, typical errors, translation speed, and overall quality is presented in the following Table1

Table 1. Comparison of youdao and deepl.

Comparison Items	Youdao Dictionary	DeepL
Interface	The interface is simple, easy to operate, with clear function divisions, making it easy to get started. However, the advertisements in the free version may affect the visual experience.	The interface is simple and elegant, with a modern design. It focuses on translation functions, has less visual interference, and is easy to operate.
Limitation of Word	There is a certain limit on the free word count, which varies depending on the version and platform.	The free version has a word limit(1500), and processing long texts may be restricted.
Fees	There is a paid membership service. By purchasing it, users can enjoy offline functions, advanced pronunciation, ad-free experience, etc. The prices vary according to different packages.	There is a paid version called DeepL Pro. The fees can be checked on the official website. Paying unlocks more functions and a higher usage quota.
Advertisement Interference	There are quite a lot of advertisements in the free version.	fewer advertisements
Typical Errors	For complex sentence structures, professional terms, and content with cultural backgrounds, the translation may not be accurate, and there may be problems of Chinglish.	For extremely rare or content with specific cultural backgrounds, there may be occasional misunderstandings, but the overall error rate is relatively low.
Translation Speed	The translation speed is relatively fast, basically meeting the daily usage needs.	The translation speed is relatively slow, especially for long texts.
overall Quality	The performance in general fields is acceptable, but it is not as good as DeepL in professional fields. The naturalness and fluency of the translated text are sometimes not satisfactory.	The translation quality is high. It performs excellently in grammar, vocabulary, and context understanding, with the translated text being natural and accurate, and having an obvious advantage in professional fields.

4. MT history and trend

The proposal of machine translation: In 1946, the world’s first modern electronic computer, ENIAC, was born. Shortly after, Warren Weaver, a pioneer of information theory and an American scientist, proposed the idea

of using computers for automatic language translation in 1947. In 1949, Warren Weaver published the "Translation Memorandum" and formally proposed the idea of machine translation.

The pioneering period (1949 - 1964): In 1954, Georgetown University in the United States, in collaboration with IBM, used the IBM-701 computer to complete the first English-Russian machine translation experiment. The IBM701 computer automatically translated 60 Russian sentences into English for the first time ever, which kicked off the research on machine translation. During this period, the United States and the Soviet Union, due to military needs, invested a large amount of funds in machine translation. European countries also attached considerable importance to it due to economic needs, and there was a boom in machine translation.

The setback period (1964 - 1975): In 1964, the US Academy of Sciences established the Automatic Language Processing Advisory Committee. After two years of research, the committee released a report named "Language and the Machine" (abbreviated as the ALPAC report) in November 1966. This report completely denied the feasibility of machine translation and declared that "there is no hope of developing a practical machine translation system in the near future or the foreseeable future". Affected by this report, the number of various machine translation projects decreased sharply, and the research on machine translation experienced an unprecedented depression.

The recovery period (1975 - 1989): In the mid-to-late 1970s, with the development of computer technology and linguistics as well as the demand for social information services, machine translation began to recover and thrive. The industry developed a variety of translation systems, such as Weinder, Urpotraa, Taum-Meteo, etc. The Taum-Meteo system, jointly developed by the University of Montreal in Canada and the Translation Bureau of the Canadian Federal Government in 1976, was a milestone in the history of machine translation, marking the transition from recovery to prosperity.[2,3,4,13]

The development period (1993 - 2006): This period was mainly the development of statistical machine translation. In 1993, the translation model based on word alignment proposed by Brown and Della Pietra of IBM marked the birth of modern statistical machine translation methods. In 2003, Franzoch proposed the log-linear model and its weight training method, as well as the phrase-based translation model and the minimum error rate training method, marking the true rise of statistical machine translation.[11]

The prosperous period (2006 - present): In 2006, Google Translate was officially released as a free service, bringing a wave of enthusiasm for the research of statistical machine translation.

In 2013, Nalkalchbrenner and Phil Blunsom proposed a new end-to-end encoder-decoder architecture for machine translation, which can be regarded as the beginning of neural machine translation (NMT).

In 2014, Bengio proposed sequence-to-sequence learning based on the encoder-decoder architecture, where both the encoder and decoder are RNN structures using LSTM, and this architecture was also applied to Google's translation.

In 2015, the University of Montreal added attention weights to the previous encoder-decoder framework, making neural machine translation reach the practical stage.

In 2017, the Transformer architecture was born, abandoning the traditional CNN and RNN. The entire network structure is composed entirely of the attention mechanism, greatly accelerating the training process.

In 2018, BERT emerged. Essentially, it runs self-supervised methods on a large-scale corpus and learns a good feature representation, providing a better foundation for language understanding in machine translation.

Big Data, Deep Learning, AI, new and more advanced technology open a new door, but did not disclose the

way of Machine Translation. Maybe this just is the answer, no matter how technology develops, the key is still in human being's hand.[1,12]

Chart 1: Translation technology typology.

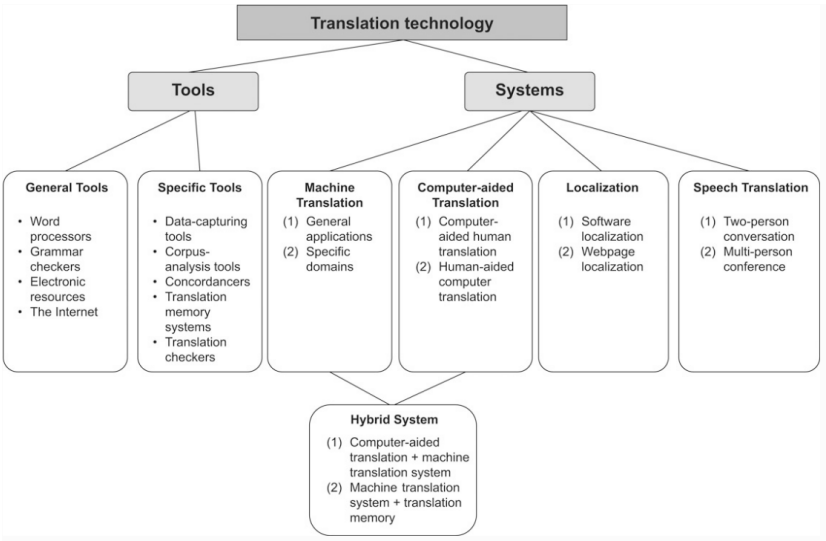
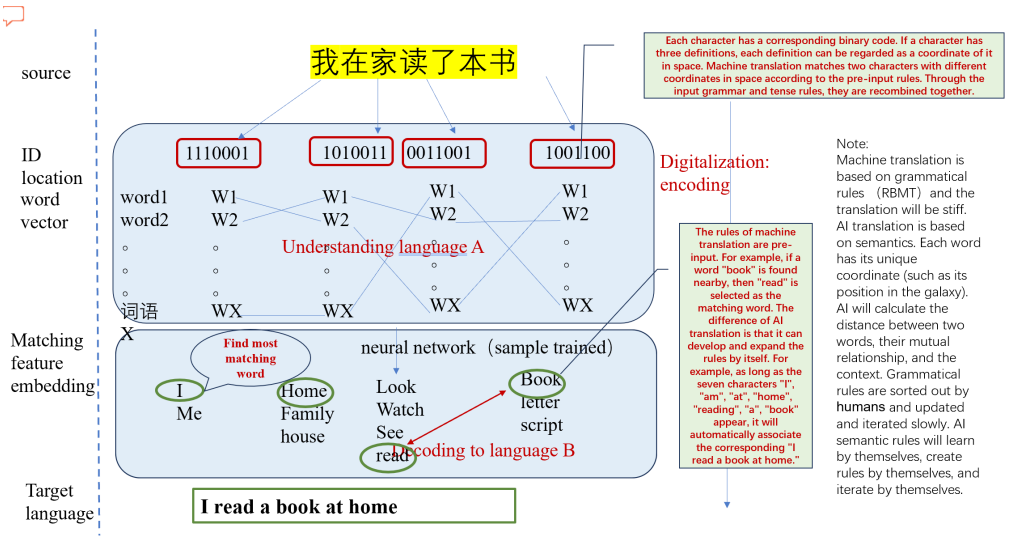


Chart 2. Mt mechanism.



5. New mt tool proposal

As mentioned before, existing MT tools all have their advantages and improving rooms. Generally speaking, current tools are already quite mature for daily communication and are playing quite important role in assisting professional translators. Since we have noticed their issues when we use them, we may have our own idea in developing a new tool with following features:

Continuous Learning and Updating Capability: Language belongs to culture domain, and culture changes with the development of the economy and society. Therefore, new words and new expressions keep emerging. Current translation software sometimes produces sentences that seem out of sync with the current era when dealing with some new words of the times and words with new meanings, such as some Internet slangs. They fail to accurately convey the current meaning. Therefore, the translation tool I develop needs to be frequently updated and have its model trained to adapt to these changes, so that it can keep up with the progress of the times despite

the constant changes in the semantics and usage of language.

Language Fluency and Naturalness: Due to different personal language habits and cultural backgrounds, many expressions and their implied meanings are difficult to directly correspond.

Therefore, when current translation software processes culturally specific expressions, slangs, idioms or idiomatic expressions during machine translation, it may seem unnatural, and there will also be problems with the fluency of translation. The fluency and naturalness are often inferior to those of human translation. Sometimes there are problems such as rigid grammatical structures, inappropriate word choices or unnatural sentence connections. Therefore, I want to develop a translation software that can select translation terms from books and film works. After all, most of the translations of film and television art works are done manually by experts. They will proofread the translation language when releasing film and television works.[10]

Regarding the above two points, my design ideas are as follows:

Nest an AI learning program, a user feedback window and a big data analysis back-end in the translation tool. The translation tool can obtain different results from the users' translation times and translation feedback. After conducting big data analysis and comparison, it will re-import them into the semantic library for expansion and comparison.

Embed video recognition and text extraction functions in the back-end program of the translation tool. First, identify a large number of classic Chinese and English bilingual film and television art works, record the obtained Chinese and English usages into the corpus, and use the AI program for matching at the same time. Meanwhile, design a display window on the front end of the software to show the source of translation.

6. Trend analysis and tactics

We noticed during an interview to the Chairman of China Translation Association and when he was asked that if the MT or AI will take place of human translation, his reply is no. According to him, some crucial issue still can not be solved in the foreseeable future such as comparing with so called "official translation" although it may not be so accurate but more popular. There is also another finding that Translation Memory may have different size especially some political sensitive words as different countries share different values. This may influence the Translation Memory and translation result.[5,9]

6.1. What is the impact of CAT to those translators or professional language service providers?

6.1.1. Positive impacts

a) Improve translation efficiency: CAT can automatically store the content translated by users. When encountering the same or similar sentences and fragments, the system will automatically prompt users to use the similar translation method, which saves time. For texts with high repetition rate, such as technical documents, product instructions, etc., the effect is particularly obvious.

b) Translation quality and consistency: CAT software can also automatically scan the translated text to identify wrong translations, simple spelling mistakes or incorrect punctuation marks, thus improving the translation quality.

c) Easy for team collaboration: In large-scale translation projects, multiple translators can use the same translation memory at the same time to realize the sharing and collaborative work of translation resources. This reduces communication costs in translation.

d) Promote knowledge accumulation and management: For professional language service providers,

the accumulated corpus is the company's language asset, which helps to improve the core competitiveness of enterprises.

6.1.2. Negative impacts

a) Over-reliance on technology: Some translators may be over-reliant CAT software leads to the lack of exercise and improvement of artificial translation ability and language literacy. This improvement in translation level will greatly be the improvement of software rather than the improvement of translators.

b) Challenges brought by technological updates: Translators need to not only learn how to translate, but also consider how to update translation software and use them, which undoubtedly increases their learning costs and work pressure.

c) Data security and privacy issues: In the process of using CAT software, translators and language service providers need to store a large amount of text data in the system. Since too much reliance on electronic tools generally leads to data security and loss problems.

6.2. Strategies to upgrade the skills of language professionals in the context of rapid development of Artificial Intelligence

6.2.1. Learning to use tools

In addition to CAT tools, language professionals should also master other related translation technologies and software, such as machine translation, terminology management tools, corpus management tools, etc. The progress of technology always depends on people themselves, so we must learn not only how to use it, but also how to update the tools.

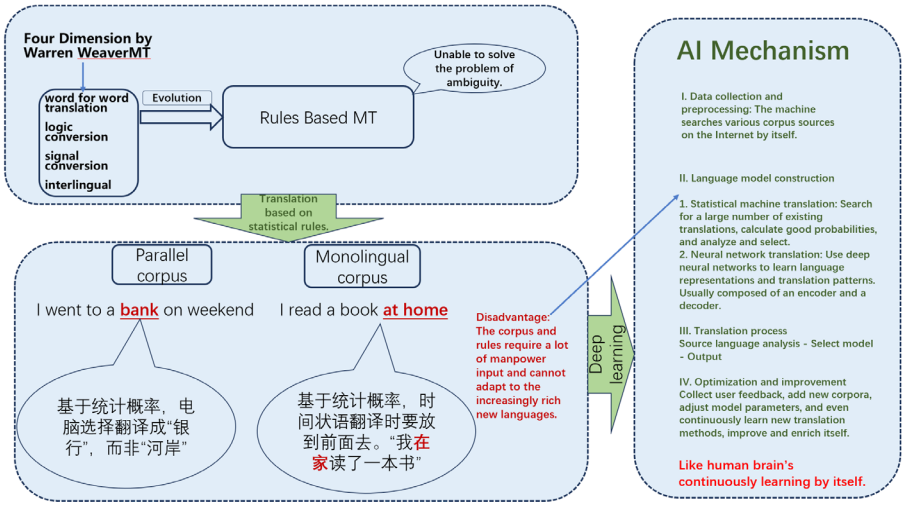
6.2.2. Strengthen language skills

Language is the foundation of translation, and language professionals should constantly improve their language level, including grammar, vocabulary and rhetoric. Through reading, writing, translation and other practices, we can enhance our sensitivity and ability to control language.

6.2.3. Be critical

Based on the results of artificial intelligence translation, language professionals should have critical thinking, be able to analyze and evaluate the results of machine translation, and judge their accuracy and rationality. For inaccurate or inappropriate translations, be able to revise and improve them.

Chart 3: Trend analysis[8].



7. Conclusion

The evolution of true machine translation has actually been no more than 40 years to date. During these 40 years, we have witnessed the rapid development of science and technology, which has driven the machine translation technology into fast track. This convenience gives people an illusion that eliminating language barriers for humanity seems to be within reach. In fact, many manufacturers and technology companies are also vigorously promoting their products and technologies in this way. Indeed, technological advancements, such as the emergence of functions like simultaneous translation headphones, smart phones, and speech recognition, have made daily communication between people of different languages more convenient.

However, from a philosophical perspective, the development of things is determined by their internal principal contradiction and the principal aspect of the principal contradiction. Therefore, when looking at the development prospects of a new thing, we need to find what its principal contradiction is. From the perspective of language barriers, the principal contradiction in the current global society is that language barriers still negatively affect the exchange of economy, politics, technology, and culture. The development of translation technology is essentially to promote these exchanges between different countries and different ethnic groups. To put it simp, it is to sign contracts more easily, share technologies, and eliminate more of the so-called "official translations" (in fact, many of those are not translated well or need to be updated). This requires societies with different values to reach a consensus in thought. Only then will there be more and more accurate translation entries that corpora, translation memories, and AI can search for and learn, and the translation will be more and more accurate.

Unfortunately, we are currently facing some different voices against globalization, which makes the development of translation technology not so optimistic in the short term.

Acknowledgements

Dr. SIU Sai Cheong, THE HANG SENG UNIVERSITY OF HONGKONG

References

[1] Person, Chan, & Sin-wai. (2014, November 18). Routledge Encyclopedia of Translation Technology: Chan Sin-wai: Tayl. Taylor & Francis.

[2] Slocum, J. (1985). Machine Translation. Computers and the Humanities, 19(2), 109–116.

[3] Duoxiu, Q. (2018). Translation Technology and Its Practical Applications. In C. Sin-wai (Ed.), An Encyclopedia of Practical Translation and Interpreting (pp. 391–414). The Chinese University of Hong Kong Press.

[4] Nirenburg, S. (1989). Knowledge-Based Machine Translation. Machine Translation, 4(1), 5–24.

[5] Kadiu, S. (2019). Human vs. Machine Translation: Henri Meschonnic's poetics of translating. In Reflexive Translation Studies: Translation as Critical Reflection (pp. 71–94). UCL Press.

[6] Specia, L., Raj, D., & Turchi, M. (2010). Machine translation evaluation versus quality estimation. Machine Translation, 24(1), 39–50.

[7] Popović, M., Avramidis, E., Burchardt, A., Hunsicker, S., Schmeier, S., Tscherwinka, C., Vilar, D., & Uszkoreit, H. (2014). Involving language professional in the evaluation of machine translation. Language Resources and Evaluation, 48(4), 541–559.

[8] LAKI, L. J., & YANG, Z. G. (2022). Neural machine translation for Hungarian. Acta Linguistica Academica, 69(4), 501–520.

[9] Baron, N. S. (1986). Language, Sublanguage, and the Promise of Machine Translation. Computers and Translation, 1(1), 3–19.

[10] Giménez, J., & Márquez, L. (2010). Linguistic measures for automatic machine translation evaluation. Machine Translation, 24(3/4), 209–240.

[11] Wikimedia Foundation. (2024b, November 5). History of machine translation. Wikipedia.

[12] Richbourg, R. F. (2018). Deep Learning: Measure Twice, Cut Once. Institute for Defense Analyses.

[13] Tong, L.-C. (1987). The Engineering of a Translator Workstation. Computers and Translation, 2(4), 263–273.