
A Comparative Study of CAT Systems and its Educational Implications

Hao Zhang¹, Yi Chen²

¹ College of Journalism and Communication, Yangzhou University, Yangzhou 225009, China

² Chengxian College, Southeast University, Nanjing 210088, China

ABSTRACT

This paper puts forward the concept of CAT in its relatively broad sense and conducts a comparative study of the two major types of CAT systems presently available through their roles as assistants. Besides, it attaches great importance to its educational implication, esp. integrating CAT method into our traditional translation practice mode and holds that learners in the information age shall be skilled at CAT so as to meet new challenges.

Keywords: CAT (Computer Aided Translation); Machine Translation (MT); Translation Memory (TM); Human Translation (HT)

COPYRIGHT

Copyright © 2019 by author(s) and Frontier Scientific Publishing. This work is licensed under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). <http://creativecommons.org/licenses>

1. Introduction

Nowadays, in the growing trend towards globalization of economy and information, communication and cooperation among different countries have been multiplying, and with the rapid advancement of information technology as well as the wide exploitation of computer and Internet, localization of drastically increasing information has become ineluctable and pressing. This results in a much heavier demand of language translation. CAT has revolutionized the traditional translation method in which human translators employ only pens, sheets of paper and paper dictionaries, or sometimes computers but just for inputting the translated texts. CAT, also known as “E-translation” or “computerized translation” or “man-computer symbiosis”(Henderson, 2009), with all its types or subsystems, has attained a very vast expanse of use worldwide, thanks to its fast speed, outstanding consistency, great efficiency, easy operation, phenomenal cost-effectiveness, etc. Now, CAT has been one of the most important research domains for translatology and computational linguistics. This paper will make a comparative study of the major CAT systems and then probe into its educational implications.

2. Definition and Types of CAT

Computer Aided Translation (CAT) in its broadest sense means all the software and hardware that facilitate the act of translating source text of one language into target text of another, inclusive of disc drivers, scanners, word processors, OCR software, etc.; and in its narrowest sense it consists only of Translation Memory (TM) systems on which we will expound later in this paper. Here in this paper, CAT is considered in its comparatively broad sense, which refers to computer systems that are involved in the direct act of source-language-to-target-language translation or language rendering. Generally, there are two major types of CAT accessible to assist or automate the direct language translation process: (1) rule-based MT systems; (2) corpus-based MT systems. (Guerra, 2000) Besides, there are a few other types, such as Terminology

Management systems, Internet Support systems which feature Internet search engines and online encyclopedias and dictionaries, Electronic Dictionary systems (for single computer use) and so on.

3. Rule-based Machine Translation (MT) Systems

3.1 What is rule-based MT system?

Rule-based MT systems (or simply MT systems), sometimes called automatic translation system, refer to a system whereby a computer program through application of linguistic rules renders the text of a source language into that of target language automatically, and then displays it to receivers. It is intended to replace the work of human translators in some particular fields or palliate to some degree the workload of human translators. In MT systems, texts of the source language are algorithmically processed or analyzed into abstract representations of ‘meaning’, involving successive programs for identifying word structure (morphology) and sentence structure (syntax) and for resolving problems of ambiguity (semantics). (Baker M., Saldanha G., 2008) In this way, MT systems can parse sentences of source texts by identifying words and relationships, then select counterpart terms in target language, and finally put them in target language word order after inflection for the eventual product.

3.2 Weaknesses of MT Systems

In spite of the fact that many efforts have been invested in the of MT research and numerous rules covering a big portion of linguistic phenomena have been set down, it is still far from enough. In translations of short simple input texts, comparatively higher readability if not accuracy will usually be accomplished; however, in translations of long complex ones, readability often becomes a big problem. At lexical level, MT Systems often fail to deal successfully with polysemy, pronouns, proper names, abbreviations, etc.; at syntactic level, MT Systems usually fail to handle satisfactorily garden path sentences, elliptic sentences, equivocal sentences, partial negatives, and special negatives; besides, translations of metaphors, idioms, slang expressions and jokes often make no sense in the target language. Moreover, natural language is dynamic process of accumulation by its speech community through a very long period of time. As a result, new words and linguistic phenomena as well as fresh rules have been constantly arising and almost every existent rule has its exceptions, which make it harder for MT systems to fully describe their underlying rules. Thus, it is the lack of contextual or real-world knowledge, asymmetry between languages and inadequacy in grammatical rules that have induced MT systems’ failure to produce high-quality translations of source texts even when well written.

3.3 Two Major Applications of MT systems

There are generally two circumstances where MT systems are usually applied.

The first application circumstance is where an indicative translation is enough for intended use or where a high-quality translation is not necessary. An indicative translation is usually for the benefit of individuals to assimilate information rather than for publication. Though MT systems cannot produce high-quality translations, some users have found that they can extract what they need to know from the indicative output. (Newton, 1992) With the explosion of information and popularization of information technology, occasions demanding an indicative translation are increasing fast and considerably, especially among people from disparate language backgrounds who need translated version of bountiful texts to procure the substantial part of the targeted foreign materials, and in most cases, as soon as possible. Such demands have mostly resulted from the drastically increasing on-line cross-language communication, where quality may be less important than accessibility and usability. Hence, MT systems are notably desirable when their low quality is adequate for indicative ends. In such situations, fastness and convenience are usually preferred to grammaticality and readability as far as the limited usability exists or the output shed some light on the target text. Typical indicative MT tools include: LOGOS(US), METAL(US), ARIANE(France), SUSY(Germany), MTSU (Singapore), Oriental Express(China), King Express Trans(China), Tans Star System (China), Oriental Net Trans(China)

and so on.

Second, MT systems have been applied to particular-domain-specific texts in controlled language. Present MT systems can only yield satisfactory products when the source text is confined to a specific field of knowledge and comport with a certain controlled language which is not only limited in vocabulary and domain but in syntax and metaphor to reduce possible problems for MT systems. Only certain syntactic structures are allowed and metaphors must be of the frozen or fixed type rather than dynamic or changeable one. With a consequentially disambiguated input text, MT systems can guarantee a fast production of both grammatically and stylistically correct output (Carl M., Way A., 2003). In such cases, MT systems are designed to handle voluminous source texts in restricted field and style to obtain a quality translation.

Some renowned particular-domain-specific MT tools are as follows: Systran (which was first by US Air Force to render Russian military technical materials into English and now has been employed by US National Air Intelligence Center to provide nearly unedited translation service for many US government organizations), Météo (which is devised by Canadian meteorological department to translate weather forecasts between English and French), TITUS (a system used to translate abstracts of documents of textile industry between English, French, German, and Spanish) and ARGO (a system developed by CSK in Tokyo to translate Japanese stock market reports into English).

As we can see, MT systems have been successfully used by individuals, organizations, and even governmental and military departments. It is believed that more applications will show up as global communication increases and MT systems' practical usability comes to be known by more people.

4. Corpus-based TM Systems

4.1 What are Corpus-based TM Systems

Corpus-based TM Systems refer to computer programs through which previously stored data (translated texts in pair with source ones) can be automatically retrieved to avoid repeated translation of identical sentences or to render similar sentences with just a few modifications. TM systems depend basically on the availability of suitable large corpora of authoritative translations or indexed bi-texts. (Somers, 2003) When translating, TM systems will search for matches between source texts and data in its source-target-language-paired database and processes the result based on similarity percentage set by users. When an identical match occurs, TM systems can directly place retrieved translation of the sentence in the output; when similarity percentage is between 0% and 100%, TM systems will return the sentence to translators and translators can refer to the translation in stock then renew it with some modifications to match the source and simultaneously the data bank will be enlarged; when no match exists, the source texts will be manually translated and automatically added to the database. The larger the database, the faster the translation. Corpus-based TM systems can speed up the overall translation process by reducing retranslation, enhance the consistency and ensure the accuracy of the translated texts, making up for some inadequacies of rule-based MT systems.

4.2 Weak Points of TM systems

First, TM systems are predicated on the concept that source-target-language-paired sentence in previous translation can be reused. It is possible for translators to handle the text sentence by sentence mechanically, ignoring the fact that rendering of one sentence depends on that of surrounding ones. Translators are supposed to focus on translating and communicating the whole message of the text instead of just translating its individual sentences separately ignorant of those around it. Second, TM systems are not fit for texts with no or little repetition or unchanged parts. Although technical texts generally benefit TM systems best, literary, legal and other texts will be less becoming since there is often little repetition in language use. Third, the costs in training of users, inputting the previous translations into TM systems' data bank and so on, can mean a sizable investment. Forth, maintenance of TM system databases is mostly a manual job, and failure to maintain them can result in considerably diminished usability and quality of TM systems. Fifth, quality of

TM systems' data bank cannot always be guaranteed; if translations inside is inaccurate, the inaccurate one will be reused and the unacceptable inaccuracy be repeated when the same or a similar source text appears next time.

4.3 Application of TM Systems

TM systems work best on source texts with abundant repetitions, like technical documents, user manuals, software localization and so on, in which the vast majority often keeps the same and the work can simply be automatic substitution by TM systems through the fast retrieval of preceding translation from the continuously expanding large-scale database, and only the small altered parts need human efforts. In such cases, TM systems assist translators to spot and re-exploit previously rendered parts. In situations where no precise equivalent is available, the presented results may be re-utilized with some modifications. Therefore, TM systems can augment tremendously the efficiency and productivity of translation, and ensure a good quality suitable for publication or circulation that is usually anticipated from human translators. In addition, TM systems are of greater cost-effectiveness and usefulness when it comes to the translation of materials restricted in a certain domain, like medical, legal, sports, agricultural, even literary and art. A recent survey by the author shows that multitudinous domestic and international companies and organizations (especially those engaged in some foreign businesses) are using TM systems. Besides, for language professionals and advanced language learners 39 out of 50 have confirmed the use of TM systems. The current typical TM system tools include ones such as IBM TM/2, Star Transit, SDLX, Déjà Vu, Wordfast, Eurolang Optimizer, Star Transit, Trados, DéjàVu, WordFisher, SDLX, AlpNet's proprietary TSS/Joust, Dr. Eye and Yaxin CAT. What's more, Most TM systems are used in conjunction with a terminology management system, a multilingual e-dictionary, and a MT system.

5. Educational Implications of the Study

The above part has made a comparative analysis of the two major types of CAT, MT and TM systems from their different definitions, weaknesses and applications, which in turn has given rise to the following educational implications for second language learners and users.

First, for language majors and professionals, the skills and knowledge about CAT systems, especially TM, are quite essential in that they can deal with the translation work more easily when encountering some big translation tasks (like highly-repetitive technical and professional texts such as localization manuals) that are usually inevitable, instead of toiling at consulting paper dictionaries, since no one can be so well-informed of all subjects, no dictionary be all-inclusive, and no one complete a huge task of translating voluminous texts in a short time on his own in a society with explosively growing information. It is very necessary to integrate such CAT method into our traditional translation practice pattern.

Second, for users not majoring in language or with low language proficiency, when they want to get some information from a certain language, CAT systems, especially MT, can be drawn on to have a useful glimpse and instrumental understanding of the targeted material in the fastest manner by automating the translation process. For instance, to get a sketchy view of some e-texts like Web pages, e-mails, etc., that may be browsed by people who just hope to catch the general information and who will not be impeded by the poor style and grammaticality, ignoring whether all is comprehensible. MT systems are the only feasible solution here, with no rivalry in terms of speed and cost-effectiveness. Language learners and users in the information age shall be skilled at such CAT so as to meet new challenges.

Last, it sheds some light on the important role human play in TM and MT systems and the relation between them. In the cases where TM and MT systems are quite useful and necessary, they are still human-centered and the final work must be done by human, even in some very restricted domains where TM results are used directly without any editing, human work in programming the systems and inputting controlled language is essential and indispensable. There is no prospect that MT or TM systems can take the place of human intelligence.

6. Conclusion

This cyber age of specialization and globalization has witnessed the increasing use of CAT. Sciences are increasingly branching out into new fields of study, bringing about more terminology. Markets have been internationalized and businesses are expanding all over the world. In such a global context, CAT has benefited from and also conduced to the global development. It can be seen that the linguistic capabilities of MT systems have not been very much improved within the last two decades, and TM systems have been widely used by language professionals and advanced language learners to handle specialized texts that are often repetitive. Many people are now finding benefits in using CAT tools when communicating across languages, particularly when a translator is not available and the user has limited multilingual skills. Language professionals and advanced language learners are finding CAT tools highly effective and indispensable in improving their translation productivity and quality for work that lends itself to the use of the different CAT tools. It can be predicted that CAT productivity tools especially TM and MT systems will in the future be fundamental to the success of language users.

References

1. Henderson H. (2009) Encyclopedia of computer science and technology. Infobase Publishing, New York City, 278.
2. Baker M., Saldanha G. (2008) Routledge encyclopedia of translation studies. Routledge, London, 163.
3. Guerra A. F. (2000) Machine translation: capabilities and limitations. Universitat de València Publications, València, 66-67.
4. Newton J. (1992) Computers in translation: a practical appraisal. Routledge, London, 148.
5. Carl M., Way A. (2003): Recent Advances in Example-Based Machine Translation. Springer, Germany, 96-99.
6. Somers H. L. (2003) Computers and translation: a translator's guide. John Benjamins Publishing Company, Philadelphia, 167.