

Translation and Technology: CAT tools and MT



Sabir Hasan Rasul
University of Human Development



CAT Tools

The term Computer-Aided Translation (CAT) refers to a translation operation in which human translation (HT) is aided by computer applications.



CAT Tools

A key characteristic of CAT is that a human translator takes control of the translation process and technology is used to facilitate, rather than replace, HT.



CAT Tools

CAT has become the predominant mode of translation in scientific and technical translation, where technology is employed to increase productivity and cost-effectiveness as well as to improve quality.



What do CAT tools refer to?

These include 'any type of computerized tool that translators use to help them do their job, ranging from general-purpose applications such as word processors and Internet search engines, etc., to more translation-oriented tools such as multilingual electronic dictionaries, corpus analysis tools, terminology extraction and terminology management systems.



CAT Tools

The main CAT tool is Translation Memory (TM), which emerged as one of the earliest translation technologies in the 1970s and commercialized in the mid 1990s.



What is TM?

TM computerized linguistic database that allows the translator to store translations in a database and 'recycle' them in a new translation by automatically retrieving matched segments for reuse.



How does TM work?

All previous translations are accumulated within the translation memory (in source and target language pairs called translation units) and reused so that you never have to translate the same sentence twice. The more you build up your translation memory, the faster you can translate subsequent translations, enabling you to take on more projects and increase your revenue.



How does TM work?

So, the TM database consists of source text and target text segment pairs which form so-called translation units (TUs). After dividing a new ST into segments, the system compares each successive ST segment against the ST segments stored in the translation database. When a new ST segment matches an ST segment in the database, the relevant TU is retrieved.



Different types of matches

There are three types of matches:

- 1) 'exact matches'
- 2) 'full matches'
- 3) 'fuzzy matches'



Exact match

1) An exact match means that the ST segment currently being translated is identical, including formatting style, to a segment stored in the memory.



Full match

2) A full match means that the ST segment matches one stored in the memory with differences only in 'variable' elements such as numbers, dates, time, etc.



Fuzzy match

3) A fuzzy match is one where there is a great degree of similarity between the ST segment and a segment in the memory, which can be reused with some editing.



Segmentation

TM technology relies on text **segmentation** and alignment. Segmentation is the process of splitting a text into smaller units, such as words or sentences. Most TM systems use the sentence as the main unit, but also recognize as segments other common standalone units such as headings, lists, table cells or bullet points.



Segmentation

On the basis of segmentation, the process of **alignment** explicitly links corresponding segments in the source and target texts to make up TUs.

Alignment algorithms are usually based on 'anchor points' such as punctuation, numbers, formatting, names and dates, in addition to the length of a segment as a measure for correspondence.



Segmentation

When a memory is created in interactive mode, alignment is verified by the translator. However, when automatic alignment is used to create memories retrospectively from past translations, known as 'legacy data', misalignments may occur.



Development of TM

In a relatively short timespan, TM technology has evolved from a first generation ‘sentencebased memory’, only able to search exact matches on the level of the full sentence, to a second generation where fuzzy matches can also be retrieved. A third generation of TM technology is now emerging where repetitions below sentence level – subsentential matches – are exploited (Gotti *et al.* 2005).



What counts for the best TU?

Translation researchers have discussed the disadvantages of using the sentence as the key processing unit from the viewpoint of translator productivity (e.g. Schärer 2001). More efficient approaches to identifying useful matches for the translator have been explored (Bowker 2002; Macklovitch and Russell 2000), but an ideal translation unit which optimizes precision and recall of matches, while facilitating but not interfering with the human translator's cognitive process, is still to be identified.



Machine Translation (MT)

MT involves the use of very large databases and statistical models to translate text from one language into another. MT technologies have improved exponentially since research began in the 1950s, with the goal of producing an infallible, universal translation tool. Despite constant improvements, researchers and practitioners recognize the concrete and measurable limitations of using MT. Rather than MT replacing human translation, the effect these advanced technologies have had has been to enhance the speed of human translation. MT is a human translation accelerator, not a replacement.

American Translators Association



History of MT

The Georgetown experiment in 1950s

- ◆ The experiment involved successful fully automatic translation of more than sixty Russian sentences into English.
- ◆ Researchers of the Georgetown experiment asserted their belief that machine translation would be a solved problem within three to five years.
- ◆ In 1966, a report by the Automatic Language Processing Advisory Committee (ALPAC) indicated that ten years of research had not fulfilled the expectations of the Georgetown experiment.



Types of MT system

Unassisted Machine Translation:

Unassisted MT takes pieces of text and translates them into output for immediate use with no human involvement.

The result is unpolished text and gives only a gist of the source.



Types of MT system

Assisted machine Translation:

Assisted MT uses a human translator to clean up after, and sometimes before, translation in order to get better quality results.



Types of MT system

In unassisted or fully automatic MT, the whole text is translated by computers without the intervention of human operators. These systems are sometimes called 'batch' systems since the whole text is processed as one task. Assisted MT is split into human assisted MT (HAMT) and machine assisted human translation (MAHT).



Types of MT system

In HAMT (also known as interactive MT), human translators intervene to resolve problems of ambiguity in the ST or to choose the most appropriate TL word or phrase for output. In MAHT, computer programs are utilized to help human translators carry out the translation. An increasingly popular form of MAHT is computer aided translation (CAT) (Baker & Saldanha, 2009).



translate



All

Images

Videos

News

Books

More

Settings

Tools

About 203,000,000 results (0.38 seconds)

English - detected



The men climbed the hill, and they were tired when they reached the top. But there they saw a crystal shop that offered refreshing mint tea. They went in to drink the tea, which was served in beautiful crystal glasses. "My wife never thought of this," said one, and he bought some crystal—he was entertaining guests that evening, and the guests would be impressed by the beauty of the glassware. [Edit](#)

Kurdish



Gava çiyê çiyayî çûn û gava ew gihîştin serî de ew teng bûn. Lê wir ew dikana kirkêşek ku dîtina çayê tepê kir. Ew diçû çê vexwarinê, ku di çarçoveya cilên cilên xweş de dixebitin. "Jina mîrê min qet qet nefikirin," yek got, û ew hinek cilan kirîn-ew şahiya mêvanan dixwînin, û mêvan wê bi bedewiya kêşan.