Translation Journal

Ilya Ulitkin is an Associate Professor of the Department of Linguistics at Moscow State Regional University and a freelance Russian-to-English translator and editor in the Quantum Electronics journal. In 1999, he defended his candidate's thesis and received the degree in Translation Studies from Moscow State Regional University, Russia. Since 2005 Ilya has been teaching translation at the same University. The areas of his interest include translation studies and computer linguistics. More importantly, he is interested in translator's electronic resources and web resources.

Ilya can be reached at ulitkin-ilya@yandex.ru.

Front Page

Select one of the previous 54 issues.

Select an issue:

Index 1997-2011

Computer-assisted Translation Tools:

A brief review

by Ilya Ulitkin

Two or three decades ago the work tools of a translator included a typewriter and a collection of printed dictionaries, which are not difficult to handle. However, as a result of incredibly rapid progress in the field of electronic hardware and computer software, nowadays an important component of any translator's professional competence is the technological one, which, first of all, assumes skills in handling electronic resources and tools [1].

Now that we hear the expression "translator's work tools," the first thing that comes to mind is a personal computer (a desktop or laptop computer, depending on one's personal preferences) and, surely, the Internet. Nobody translates the way they used to thirty or forty years ago because convenient electronic dictionaries, special translation software, and Internet resources are available, which allows us to keep up to date. This is especially important if we take into account the fact that we have entered the 21st century and virtually all translators use the Internet, the computer, and other electronic means in their work [2].

> Despite their efficiency and outlooks, the translation software and electronic means cannot replace the human translator and guarantee high-quality translations.

The idea of machine translation may be traced back to the 17th century. In 1629, René Descartes proposed a universal language, with equivalent ideas in different tongues sharing one symbol. Nevertheless, the development of machine translation started only in the 1940s when first electronic computers were designed. In March 1949 Warren Weaver, the director of natural sciences at the Rockefeller Foundation, intrigued by the way the British had used their

pioneering Colossus computer to crack the military codes produced by Germany's Enigma encryption machines, formulated the concept of machine translation. He wrote, "I have a text in front of me which is written in Russian but I am going to pretend that it is really written in English and that it has been coded in some strange symbols. All I need to do is strip off the code in order to retrieve the information contained in the text."

Weaver's memo attracted interest to the problem of machine translation, especially among the military. They were the first to actively support the development of the machine translation software. In the USA, much attention was paid to translation from Russian into English, while in the USSR, to translation from English into Russian.

Today, we can speak of three approaches to written translation: the first one is machine translation based on the rules of the source and target languages, the second approach involves statistical machine translation, and the third one is computer-assisted (or computer-aided) translation.

The earliest "translation engines" in machine-based translations were all based on the direct, so-called "transformer," approach. Input sentences of the source language were transformed directly into output sentences of the target language, using a simple form of parsing. The parser did a rough analysis of the source sentence, dividing it into subject, object, verb, etc. Source words were then replaced by target words selected from a dictionary, and their order rearranged so as to comply with the rules of the target language. This approach was used for a long time, only to be finally replaced by a less direct approach, which is called "linguistic knowledge." Modern computers, which have more processing power and more memory, can do what was impossible in the 1960s. Linguistic-knowledge translators have two sets of grammar rules: one for the source language, and the other for the target language. Besides, modern computers analyze not only grammar (morphological and syntactic structure) of the source language but also the semantic information. They also have information about the idiomatic differences between the languages, which prevent them from making silly mistakes. Nevertheless, much remains to be done.

The second approach is based on a statistical method: by analyzing a large amount of parallel texts (identical texts in the source and target languages), the program selects the variants that coincide most often and uses them in the translation. It does not apply grammatical rules, since its algorithms are based on statistical analysis rather than traditional rule-based analysis. Besides, the lexical units here are word combinations, rather than separate words. One of the well-known examples of this approach is "Google Translate," which is based on an approach called statistical machine

translation, and more specifically, on research by Franz-Josef Och who won the DARPA contest for speed machine translation in 2003. According to Och's keynote speech at the Machine Translation Summit in 2005, a solid base for developing a usable statistical machine translation system for a new pair of languages from scratch, would consist in having a bilingual text corpus (or parallel collection) of more than a million words and two monolingual corpora of each more than a billion words. Statistical models from this data are then used to translate between those languages. However, the translated sentences are sometimes so discordant that it is impossible to understand them.

Computer-assisted translation is quite a different approach. According to "Wikipedia," **computer-assisted translation**, **computer-aided translation**, or **CAT** is a form of translation where a human translator translates texts using computer software designed to support and facilitate the translation process [3].

The idea of computer-assisted translation appeared with the first computers: many translators were against machine translation, which was the object of many studies in computer linguistics, but actively supported the use of computers as a translator's workbench.

In fact the modern idea of computer-assisted translation was put forward by Martin Kay. His memorandum in 1980 combined a critique of the current approach to machine translation, namely the aim to produce systems which would essentially replace human translators or at best relegate them to post-editing and dictionary updating roles, and an argument for the development of translation tools which would actually be used by translators. Since this was before the development of microprocessors and personal computers, the context was a network of terminals linked to a mainframe computer. Kay's basic idea was that existing text-processing tools could be augmented incrementally with translation facilities. The basic need was a good multilingual text editor and a terminal with a split screen; to this would be added a facility to automatically look up any word or phrase in a dictionary, and the ability to refer to previous decisions by the translator to ensure consistency in translation; and finally to provide automatic translation of text segments, which the translator could opt to let the machine do without intervention and then post-edit the result, or which could be done interactively, i.e. the computer could ask the translator to resolve ambiguities [4].

Nowadays, the most wide-spread means of computer usage in translations is the software ensuring automation of the translation process, namely:

(i) Electronic dictionaries (iFinger, Abbyy Lingvo, Collins-Ultralingua, Mobile Systems, Paragon Software, etc) and corpora of national languages (British National

Corpus, American National Corpus, Russian National Corpus, English-Norwegian Parallel Corpus, etc.);

(ii) Computer-assisted translation tools (CAT-tools) or "translation memory" software (SLD Trados, MemoQ, Déjà vu, StarTransit, Wordfast etc.);

(ii) Editor software (SpellChecker, StyleWriter, etc.)

Electronic dictionaries can be monolingual, bilingual, and multilingual. Besides, they contain information on word forms, pronunciation (quite often voiced by professional speakers), and word collocations. They may also include dictionaries in particular fields of science (applied mathematics, physics, biology, medicine, religion, engineering, etc.), idioms, slang, etc.

All electronic dictionaries can be conventionally divided into online and offline dictionaries. Online dictionaries require access to the Internet, while offline dictionaries can be installed on a computer and used offline.

Dictionary software generally far exceeds the scope of hand-held dictionaries. Many publishers of traditional printed dictionaries such as Langenscheidt, Collins-Reverso, Oxford English Dictionary, Duden, American Heritage, and Hachette, offer their resources for use on desktop and laptop computers. These programs can either be downloaded or purchased on a CD-ROM and installed. Other dictionary software is available from specialized electronic dictionary publishers such as iFinger, Abbyy Lingvo, Collins-Ultralingua, Mobile Systems and Paragon Software. Some electronic dictionaries provide an online discussion forum moderated by the software developers and lexicographers [5].

An advantage of electronic dictionaries is their convenience, high speed of information processing, possibility to quickly import the equivalent of the searched for word into the text as well as compactness. However, some of the online dictionaries are based on the 'wiki' principle. It means that users themselves constantly upgrade it. An example of this is the Urban Dictionary [6]. However, as B. Osimo states, the advantage of continuous update has its reverse side: a printed text is usually more accurate; before investing time and money in paper publication, authors and publishers usually try to have an acceptable product. As is well-known, with the Internet anyone can publish their own site (providing they do not infringe national or international law), even without an advisor, editor, or publisher [7].

As a result, the attitude of many lexicographers to electronic dictionaries is quite skeptic because their use may degrade the quality of translation.

However, the advantages of electronic dictionaries are obvious. First, they are not as conservative as printed

dictionaries since they are constantly updated. This is especially important when we deal with promising and rapidly developing sciences such as telecommunication systems, nanotechnology, computers, etc. Printed dictionaries become outdated very soon, and the only way to keep up to date with scientific and technological progress is to use electronic dictionaries.

Second, electronic dictionaries provide an easy access to lexicographical resources and fast search for linguistic information not only in the dictionary but also on the Internet. For example, the printed version of the Oxford English Dictionary has 12 volumes, and to find the desired word will not be an easy task, while the time needed to look through the volumes in search of the necessary word will take seconds in its CD-ROM version.

Third, all electronic dictionaries have an option to add users' dictionaries, which helps expand the basic version.

Fourth, electronic bilingual and multilingual dictionaries make it possible to reverse the direction of translation [2].

Another way to improve your translation is the use of corpora of national languages. This is especially important when translating from a native language into a foreign language.

A linguistic corpus is a set of texts which are collected and accumulated in accordance with certain principles, tagged and parsed, and have a special search engine. The expediency of corpora for a translator is explained as follows:

(i) corpora are a collection of samples of written and spoken language from a wide range of sources (literature, scientific publications, magazines and journals, academic reports, and dialogues). They are useful when you are not sure whether or not the word can be used in a given context;

(ii) texts presented in corpora have representative data, since almost all modern corpora include more than 100 million words and word combinations;

(iii) corpora may help solve different linguistic problems including those in translation.

Nevertheless, despite large and accurate representative data, corpora cannot cover all spheres of human activities, which makes them inapplicable to some fields of science, for example.

The products based on the "translation memory" (TM) software are intended for professional translators and translation agencies.

Such CAT-tools as SDL Trados, Déjà vu, StarTransit, Wordfast, etc. are now an integral part of the modern

translation process. Their use is also expedient when some work should be done by a group of translators as it is necessary to provide consistent translation within one project. As a result, the translation is stored in one database available to all the participants of the translation process. The translators "see" the results of their work in real time. In this case, they can be connected to a single network locally or remotely. This is especially important for companies with branches in different countries.

Translation memories, also known as translation databases, are collections of entries where a source text is associated with its corresponding translation in one or more target languages. Typically, TMs are used in translation tools: the software divides the text into segments, which can be blocks, paragraphs, sentences, and even phrases. When the translator "opens" a segment, the application looks up the database for the equivalent source text. The result is a list of matches usually ranked with a score expressing the percentage of similarity between the source text in the document and in the TM [an exact match (100%) or fuzzy match (less than 100% match)].

Some of the advantages in using TMs include:

(i) The translation can be performed much faster: unnecessary retyping of existing translations is avoided, and at most only parts of text need to be changed.

(ii) TMs also allow a better quality control by offering translation candidates that have been already approved, with the correct terminology.

Translation memory is a powerful technology, which can help reduce the cost of localization. However, the use of TM needs to be weighted and all factors taken in account, since the application of the TM software is justified and effective in translating texts with a high degree of repetitions.

The final step in the translation processes is editing. Most translations benefit from careful editing to improve clarity and readability, since they can often be littered by wordy phrases, needless repetitions, clichés, trite expressions, vague terms, redundancy, pretentious language, illogical statements, homonym confusions, jargon, misspelled terms, incorrectly formed plurals and possessives, and other common errors. To improve the translation, one can use either human editors or editing software. Nowadays, almost all word processors have proofreading tools and even have options to pop up advice if some grammar or stylistic rules are violated. We may find information and instruction on how to deal with these problems in grammar handbooks, style guides, editing manuals, dictionaries of spelling and usage, and similar references. However, for an inexperienced translator, these resources have a nearly fatal flaw: they can help

improve the translation only if the translator knows what to look up. Besides, it is easy to overlook punctuation and spelling mistakes, word choice, phrasing, and style. Thus, a translator needs software that will combine both proofreading tools and style tools, which will correct and polish drafts at the word and phrase levels. Examples of editing software are Editor and StyleWriter programs, which search for thousands of writing faults, including complex words, jargon and abstract words, wordy phrases, hidden verbs, passive verbs, clichés and long sentences. After the analysis of the text, the program provides prompts showing how to edit each sentence. Initially, the editing software was developed to help professional groups (lawyers, office workers, etc.) improve the quality and clarity of written communication. However, translators can also use this software to write in an International English Style—a style that is clear, concise and readable.

Note in conclusion that use of electronic technologies is not a universal panacea for all the problems in translation. Despite their efficiency and outlooks, the translation software and electronic means cannot replace the human translator and guarantee high-quality translations. Their aim is to accelerate and facilitate the translation process, to help solve many problems appearing in the course of the process, and to minimize the time needed for translation [2].

A high-quality translation results from the combination of electronic technologies and the translator's skills, of good knowledge of a foreign language and theory of translation because programs and translation software will not replace humans even in the long-term future—at least not until actual high-performance artificial intelligence is created. Therefore, much depends on the translator's personality and his professional experience, while electronic systems are useful, necessary and sometimes required supplements.

References:

[1] Komissarov, V.N. (1997). *Теоретические основы методики обучения переводу* (Theoretical Foundations of the Method of Training of Translators), Moscow: Rema [in Russian].

[2] Shevchuk, V.N. (2010). *Электронные ресурсы переводчика* (Translator's Electronic Resources), Moscow: Librait, [in Russian].

[3] http://en.wikipedia.org/wiki/Computer-assisted_translation .

[4] Hutchins, J. (1998). *Machine Translation* 13 (4): 287-307.

[5] http://en.wikipedia.org/wiki/Electronic_dictionary.

[6] http://www.urbandictionary.com/.

[7] Osimo, B. *Translation Course* - http://courses.logos.it/EN/5_13.html.