



**SAPIENTIA
ALAPTIVÁNY**

Lektorok:

Standavid Beatrice (Marosvásárhely)
Benő Attila (Kolozsvár)
Fazakas Noémi (Kolozsvár)
Sárosi-Márdrosz Krisztina (Kolozsvár)
Kelemen Attila (Kolozsvár)

Sorozatorrító:

Miklósi Dénes



A szakmai felelősséget teljes mértékben a szerkesztők, illetve a szerzők vállalják.

Előő magyar nyelvű kiadás: 2015

© Scientia 2015

Minden jog fenntartva, beleértve a sokszorosítás, a nyilvános előadás, a rádió- és televízióadás, valamint a fordítás jogát, az egyes fejezeteket illetően is.

Descrierea CIP a Bibliotecii Naționale a României

NYELVI SOKSZÍNŰSÉG EURÓPÁBAN. Conferință națională (2013 : Targu-Mures)

Nyelvi sokszínűség Európában = Diversitatea lingvistică în Europa = Linguistic diversity in Europe = Sprachliche Vielfalt in Europa : conferință națională : Targu-Mures, 2013 / ed.: Pleșt Rita, Nagy Imola Katalin. - Cluj-Napoca : Scientia, 2015

Bibliogr.

ISBN 978-973-1970-87-5

I. Pleșt, Rita (ed.)

II. Nagy, Imola Katalin (ed.)

81(4)

**TARTALOM/CUPRINS/TABLE OF
CONTENTS/ INHALTSVERZEICHNIS**

Előszó	7
Prefață	8
Foreword	9
Vörwort	10
NYELVI ÉS KULTURÁLIS HATÁRVONALAK	11
Bence Erika – Németh Ferenc: A magyar történelem idegen kultúrtérben	13
Todor Imre: A fordítás mint alkudozás	25
Rajshi Ilona: A vajdasági magyarok utónévhasználata	35
Both Csaba Attila: Nyelvjárású változások Háromszéken	43
KULTURÁLIS MODELLEK REPREZENTÁCIÓJA A MÉDIÁBAN	51
Kos Anna: Globalizálódó tartalmak a többnyelvű Európában, avagy: Több nyelv – több kultúra?	53
Lechner Ilona: Politikai erkölcs – és ami mögötte van – a kárpátaljai magyar nyelvű sajtó anyagán vizsgálva	67
Suba Réka: A nyelvi norma és nyelvhasználat kérdései a romániai magyar audiovizuális médiában	77
ANYANYELV-PEDAGÓGIA, NYELVOKTATÁS	87
Pleșt Rita: A fogalmazástanítás az iskolai gyakorlat és az eredmények perspektívájából	89
Kovács Gabriella: Alkalmazott színház és dráma	99
Szaló Réka: A játék szerepe a nyelvoktatásban	109
CONTACTE LINGVISTICE ȘI CULTURALE	119
Anghelulă Mioara Adelina: Tipuri de evidențiale și rolul acestora în presa tabloaidă românească: cazul imperativului <i>vezi și a interjecției uife</i>	121
Buturcă Doina: Funcția fatică în limbajul jurnalistice	133
Jianu Maria-Magdalena: Efectele imprimaturilor lingvistice asupra limbii române actuale	143
Nagy Imola Katalin: Aspecte lingvistice ale nomenclaturii botanice	149
Suba Réka: Fenomene legate de acustica textelor în audiovizualul de limbă maghiară din România	163

MACHINE TRANSLATION AND CAT-TOOLS¹

Introduction

Nowadays, it is in fashion to measure everything from the point of view of profit; so, there is an increasing interest in language services of whatsoever type (translation activities, interpreting, etc.). This refers to machine translation (MT) and computer-assisted translation (CAT) as well. According to various European Union studies (e.g. *Study on the Size of the Language Industry in the EU, 2009*), the growth rate of the language industry is estimated at 10% minimum over the next few years.

The acquiescence regarding the laboriousness of translation resulted in a constant search for easing and boosting this type of work, facilitated by better and better computers, the spread of the Internet and the immense financial possibility of the business either as automated translation (MT) or computer-assisted translation (CAT).

Attitudes of experts (practising translators and interpreters) have changed considerably towards MT and CAT over the years, including rejection, contempt, fear and acceptance. The evergreen debate whether we translate *meaning* or *word* is irrelevant at this stage, although it is clear that MT focuses on words. At least, this is what the present MT results suggest, even if Reifler offered an answer in 1955: "MT is concerned primarily with meaning" (Reifler 2003: 22), delineating the task of MT as well: "The first concern of MT must always be the highest possible degree of source-target semantic agreement and intelligibility."

Machine translation constitutes one of the hottest topics regarding both the present and the future state of translation industry, bugging the mind of people hoping for an effective universal translator. The rapid technological developments resulted in deep and lasting impressions regarding this field as well, and we tend to believe that spectacular improvements are yet to come, after having celebrated its 60th birthday. Thinking about the rationale of the term 'machine translation', we fully agree with Somers, who accepts that it is "misleading", as nobody calls a computer a 'machine' today (Somers 2003: 1).

However, many agree that at present translators should be able to use dedicated resources (term bases and translation memories), adequate search engines and portals, specialist software as well as 'semi-automatic' translation

¹ The present article is based on two chapters from the author's *Traps of Translation* (Imre 2013).

systems. Gouadec is one of the ardent supporters of this "not so friendly revolution" (Gouadec 2007: 286), as it is rather painful to learn for translators not accustomed to these, let alone the financial aspect.

Ardent defenders of machine translation usually highlight these aspects, benignantly overlooking the idiomatical, cultural, contextual or other knowledge parts. Thus, human "understanding" seems to be unrivalled in this regard, although huge improvements have been made in the field of machine translation. At present, it is questionable whether humans can programme a computer to create an "authentic" target text based on a source text, although there are different approaches. These may be of various or combined types (some even running on cross-platforms such as IBM):

1. *rule-based interlingual machine translation* (e.g. SYSTRAN, Eurotra, Apertium);

2. *rule-based transfer machine translation*:

a. *shallow-transfer machine translation* (no previous knowledge of the text)
 b. *deep-transfer machine translation* (presumes a comprehensible knowledge of the word (e.g. Matxin, Openlogos));

3. *rule-based dictionary machine translation* (starts from dictionary entries);

4. *direct machine translation*, words are translated directly;

5. *statistical machine translation* (SMT) works on bilingual text corpora;

6. *example-based machine translation* (EBMT) is in fact an analogy translation making use of a bilingual corpus;

7. *hybrid-based machine translation* (HMT) tries to balance the advantages and disadvantages of both rule-based and statistical MT (e.g. Wordlingo).

The list is probably not complete, but the basic idea is visible: rule-based methods parse the source text up to a certain (below-human) level, and due to this drawback the generated target text needs "human intervention." As Prószéký concludes, rule-based machine translation is characterized by low recall and high precision, whereas statistical systems try to handle each and every single case, but it is prone to error even in trivial cases (Prószéký 2005: 80).

Anyway, at present, we are less interested in the machine translation method than in its output: so, in the following, we will try to offer definitions of machine translation, looking into its history, offering both pros and cons. Furthermore, we are going to test it before enlisting some allegations and offering some concluding remarks.

Definition of MT

There are more possibilities to define machine translation. The simplest definition describes MT as a procedure by which an activated computer programme analyses the source text and produces a target text 'without further

human intervention.' Anyway, before the digital computers, this belonged to the realm of science fiction (Nirenburg 2003a). As Nirenburg explains, "[t]ranslation among languages was among the first non-numerical applications suggested", which was especially welcome during the wartime successes of cryptography.

However, there are more detailed approaches: *machine aids* for translators, *machine-aided* translation and *machine* translation (cf. Blatt's definition in Freigang 2001: 134). We knowingly avoid the term *computer-aided* translation, as this refers to something different today. Machine aids may include word processors, e-dictionaries, and various term banks without performing the translation task: Blatt defines machine-aided translation as a tool to help the translator, whereas machine translation is fully automatic. Naturally, human post-editing is more than desirable in this case. Although there are views (e.g. Somers) that *pre-editing* should be considered in order to obtain a better MT, we would still be inclined towards the seminal importance of *post-editing*.

Thus, our definition of machine translation sounds like this: a computer programme either separately installed on an operating system or accessed online, which is capable of reading a source text in a language and – without human intervention – is also capable of transforming it into (a)ny different language in a comprehensible way by a target speaker. However, the prerequisites include human intervention in the form of preparation (pre-editing):

- choosing a natural source language;
- preparing the source text in machine-readable form (e.g. scanned handwriting excluded);
- selecting the target language.

The *modus operandi* does not constitute the active part of the definition, and neither does the analysis/appreciation of the result. Nevertheless, at present, without post-editing, re-editing and correcting, one should not expect very good results, which may be explained by understanding how machine translation works. Bowker considers that "pre- and post-editing are accepted as the norm and some users are satisfied with output that simply allows them to get the gist of the source text" (Bowker 2002: 4), adding that modern trends highlight how MT can support human translation.

Albert explains that machines do not "translate", they do not search for equivalents or look for meanings, and they cannot "read between the lines." Instead, they recode a language system into another with formal equivalents, operating with word-meanings; hence, we can conclude that this is nothing else than mere code-switching (Albert 2011: 81) without the real problem of polysemy. However, when we would like to translate anything with machine translation (e.g. *Google Translate*), we are already offered possible variants.

Advantages of MT

It is true that the limited success of the second 'indirect' approach was partially due to the high expectations in Europe. However, if we focus on the positive aspects of MT, we can mention Somers' analysis of rough/raw input of the fully automatic MT (Somers 2001: 137–8). He claims that this "may still be useful, even though it may lack in style or even accuracy" when the source text is an 'exotic' language (cf. unfamiliar writing system) or "the consumer may be a (perhaps amateur) translator, or a subject specialist", as the 'quick and dirty' first draft offers a clue for the translator about the relevant parts or the scientist wants "to know only roughly what the article says."

If we want more than 'rough' translation, MT may be of great help in case of specialized translation:

"Technical writing is unfortunately not always straightforward and simple in style; but at least the problem of multiple meaning is enormously simpler. In mathematics, to take what is probably the easiest example, one can very neatly say that each word, within the general context of a mathematical article, has one and only one meaning." (Weaver 2003: 15)

Others correctly observe that even human translation is "subject to revision." (Somers 2001: 138); remarkable facts are that the revision of MT does not hurt feelings, and it is prone to commit recurrent mistakes which may be easily post-edited with further interactive tools or near synonyms by accessing dictionaries and thesauri. Indeed, proper terminology work and serious post-editing of MT can produce results in a shorter period of time. However, this is only a possibility by itself, and at present the tendency is to have an MT system integrated into a computer-assisted translation tool (cf. *MemoQ*). The main argument comes from Kay, who mentions that there is a real market for MT, simply because of "the large quantities of text to be translated" (Kay 2003: 223).

A further real advantage of MT is "gist translations from languages you know nothing about. It allows users to identify the texts or fragments of interest, which they can then have translated by other means." (Biau Gil-Pym 2006: 16) They also mention that MT produces "high-quality translations in very restricted contexts", but in our opinion this is not a real-life situation. However, gist translation functions very well when in the 21st century people keep buying from eBay (technical gadgets or common, everyday products).

There are other real-world applications; in this respect, we can mention the customized version of SYSTRAN by the European Commission, the English-Danish *P4Trans* belonging to *Lingtech AS* (combined with *SDL Trados* commercial CAT tool), or the Spanish-Catalan MT system for the Spanish daily newspaper, *Peridico de Catalunya*.

Last but not least, supporters of MT mention the benefits in translation connected to terrorism and international warfare.

If humour is considered an advantage, then in particular cases MT systems should be awarded the special prize of the most "authentic" and funniest *mistranslation providers*.

Disadvantages of MT

From the outset, we would like to present the most negative evaluation of MT we have come across until now, which is Wikipedia's Hungarian page of what Wikipedia is not. It is explicitly stated (point 24) that Wikipedia does not collect machine translations, as texts translated by MT – in their experience – are simply not comprehensible in Hungarian, and they are not worth correcting either.² However, nothing similar is to be found on either the corresponding English or Romanian Wikipedia page.

Yorick Wilks mentions another very gloomy prediction regarding MT, referring to Martin Kay: "even if all the problems of syntax, morphology, and computational semantics had been individually solved, it might not improve MT" (Wilks 2003: 203) A further problem (initially raised by Yehoshua Bar-Hillel) is *word sense disambiguation* in case a word has more than one meaning. In the 1950s, this was an insurmountable problem, but today this may be (at least partially) solved with either the shallow (no knowledge of the text) or the deep approach (comprehensive knowledge of the word). As human (natural) language is contextual, seemingly little chance is given MT to "understand" it, but sometimes even humans misinterpret contexts (and not only in case of puns, jokes but everyday conversations as well).

A certain disadvantage of MT comes from proofreaders scared about the future possibilities of MT (to be discussed later). They tend to highlight all possible errors and they often approach the MT-produced target text subjectively. Hence the criticism that MT is only "gisting translation" (Thicke 2009), often requiring an experienced reader with good knowledge of both languages. Yet, the basic idea of gisting translation is to offer a rough translation of a language the reader has 'no idea about'. No wonder that Claude Piron concludes (Piron 1994): "MT, at its best, automates the easier part of a translator's job; the harder and more time-consuming part usually involves doing extensive research to resolve ambiguities in the source text."³

The translation of idiomatic expressions, metaphors, puns is extremely problematic for MT, often resulting in 'patent nonsense' and undeniably authentic humour. The explanation may stem from the stubbornness of MT supporters, relying on Bloomfield's remark, also mentioned by Nirenburg and

² Source: http://hu.wikipedia.org/wiki/Wikipedia:Mi_nem_valo_a_Wikipedia_ba?, 10.02.2013.

³ <http://phlendidichvien.com/fags/253-machine-translation-do-we-need-translators.html>, 10.02.2013.

Reifler (Nirenburg, 2003b: 6), according to which “whatever can be said in one language can doubtless be said in any other ... the difference will concern only the structure of the forms and their connotation.” Beaugrande and Dressler explain that “a computer working only with a grammar and lexicon (both virtual systems) was found unable to operate reliably, because it could not evaluate context” (Beaugrande–Dressler 1981, Chapter X.).

Machine translation is now available in many languages (cf. www.world2word.com), but it has become popular among people looking for gisting translation thanks to the Internet (*Babel Fish*, *Babylon* or *StarDict*). *Ectaco* produces pocket translation devices, which use MT (www.ectaco.com).

Google also uses MT (Google translation tools), which is getting more and more widespread. In order to test its efficiency, we selected a real conference programme in 2011 and “gave” it to Google Translator. The source text was in Hungarian and we asked for a Romanian translation.

Our first remark would be that although the “job” was to translate from Hungarian into Romanian, some words have been translated into English; this might mean that whatever language combination we choose, Google translator will first use English as the mediating language. And indeed, Jost O. Zetsche supports our observation (Zetsche 2009). When no corresponding word is found, the English intermediary word is preserved, and in case the initial Hungarian is preserved (rare/longer/compound words), then there is no English translation for that word. As of the beginning of 2013, there are more than 60 source and target languages available for Google translate,⁴ although Zetsche mentions more than 300 languages, Latin included. Allen differentiates MT with or without post-editing, stating that the primary reason for having Internet MT portals is gisting, which is in fact MT with no post-editing (Allen 2003: 301). And this is exactly what we ‘wanted’ from GT in our test.

Of course, there were successful parts during the translation (Hu: *Szemelvények a virtudis világ történetéből*; Ro: *Extrase din istoria lumii virtuale*; En: *Extracts from the History of the Virtual World*), but we should also take into consideration that the text was very “easy”: titles and names of a conference, without elaborated sentences or longer paragraphs. Yet, the aforementioned verdict is still valid: at its best, MT is (only) good for gisting. For instance, Hungarian names (Haller *Piroska*, Tókécs *Györgyvári*) were translated as *Little Red Riding Hood* Haller and *Blood Pearl Malard*. The same test was repeated in 2012, and the results were much better: the Hungarian names (*Piroska*, *Györgyvári*) were not translated anymore. We can conclude that within a year considerable improvements have been carried out in three out of seven cases. The English imprint upon the Hungarian–Romanian translation is still visible, but proper names and word-boundaries were recognized, although not in all the cases. Still,

⁴ Source: <http://translate.google.com/>, 12.02.2013.

we consider that this is an encouraging result, at least at the level of lexicon, even if post-editing may take too much time occasionally.

Allegations

The first allegation regarding the problem of MT probably comes from 1954, when it was already predicted that by 1957 or 1959 the problem would have been “solved” forever, even if Yehoshua Bar-Hillel remarked that “RAHQ [Prilly automated, high-quality MT] is not really attainable in the near future” (Bar-Hillel 2003: 45). As a response, a new concept was born, the ‘computer-based tools for translators’ in the form of the translators’ workstation (cf. Somers 2003: 6).

Fifty years later (in 2004), MIT’s *Technology Review* stated that universal translation and interpretation would likely become available “within a decade”. The decade is over in 2014, and we are still confronted with the problem Reifler conceptualized a very long time ago: “The problem that faces us here is the minimum human intervention necessary or the maximum mechanization possible to make MT both feasible and practical.” (Reifler 2003: 27)

A recent prediction is signed by Raymond Kurzweil, an American author of books on health, artificial intelligence, transhumanism, technological singularity and futurism. He is also inventor, futurist and director of engineering at Google, attracting significant criticism from scientists and thinkers. Kurzweil predicted that, by 2012, machine translation would be powerful enough to dominate the field of translation.⁵

Has MT become so powerful by now? According to our two tests, MT is rather far from having solved the “problem of translation”, but let us check what the experts have to say about it. The ultimate question regarding MT may be verbalized this way: Who has to be satisfied? Somers mentions different users: end-user (i.e. the consumer of the translation), an intermediate agent, the translator, and the original author of the text to be translated (Somers 2001: 136), and we tend to believe that even the results of our brief tests are not satisfactory. Thus, our verdict is compatible with Biau Gil and Pym, who state that “machine translation systems are not replacing human mediators” (Biau Gil – Pym 2006: 13).

Shields mentions that dictation to a computer is already used and voice can be generated from a computer, seriously affecting interpretation as well. However, rule-based MT cannot be extremely successful, as Nagao sees the key in the quality of samples, which should result from real-life situations, not from extended rules (Prószyński 2005: 84). Biau Gil and Pym are more convinced that MTs are not replacing human translators as fully automated MT is not a viable solution; furthermore, quality MT needs “serious attention to controlling writing

⁵ Source: <http://www.axistranslations.com/translation-article/what-is-translation.html>, 10.02.2013.

of the input, which is an area that some translators may want to move into.” (Biau Gil – Pym 2006: 17) This means that the more texts we have, “the more texts will be processed, and the more work will be created for human translators.” (Biau Gil – Pym 2006: 13) They also repeat that at present less time is needed to translate from scratch than correcting MT errors.

Post-editing may result in more work on behalf of the proof-reader than a new translation, but we should not overlook the fact that the number of those who are satisfied with gisting translation is on the increase due to the presence of the Internet on a large scale (desktop computers, notebooks, smart phones, iPads, etc.). Official statements report that Google improved their translation capabilities by inputting approximately 200 billion words from United Nations materials to train their system; thus, the accuracy of the translation has improved.

Prószyński tries to extenuate the verdict by differentiating texts, and accepting that literary texts are non-translatable by MT (lack of cultural background), but MT may function better with specific/technical texts (Prószyński 2005: 79). He goes on and summarizes: our present MT systems should both translate and be tolerant with errors as the primary aim is understanding the content and definitely not to challenge the high-quality human translations (Prószyński 2005: 83). He also adds that the greatest challenge is in fact the internet with extremely many incorrect instances, which cannot be corrected. MT cannot decide whether it is faced with a new word/phrase or it is just a typo (Prószyński 2005: 82).

The basic-level MT simply substitutes words for words (and it is much faster and rather reliable in the case of nouns) and potential users may take it ‘as is’. Its speed will surely affect the translation industry, according to Mike Shields, presented in Anderman & Rogers (2003: 43).

However, in Piron’s estimation, only 25% may be automated, whereas the harder 75% is still done by a human translator. Boulton comments on the translation feature of Google Chrome: “not all of the translations will be clean, crisp and accurate. But as with everything else Google does, Translate is an iterative technology that will advance over time.” (Boulton 2010); Jost O. Zetsche is more concise: “neither the tool vendors nor translation agencies should be too worried” regarding MT (Zetsche 2009), although he remarks that whenever we upload an existing translation memory or perform any translation “the material will be used by Google for the training of its machine translation engine, even if you declare your translation memory to be »private«, and even after you »delete« it.” If true (and we see no reason why it should not be), certain conclusions may be formulated:

- a. MT will be better in time, including more and more languages;
- b. MT will be used by an increasing number of people;
- c. the more people will use MT (predominantly non-professionals), the less

6 <http://www.eweek.com/c/a/Web-Services-Web-20-and-SOA/Google-Chrome-Gets-Machine-Translation-New-Privacy-Features-839506/>, 13.02.2013.

trustful MT will be for professionals, which is not in contradiction with point a; in fact, the present-day ‘balance’ will be kept, as larger and larger databases (partially unchecked) enter the MT translation memory.

Kay mentions the importance of modesty in developing MT systems, as it “will not run before it can walk.” And he is sure that an effective machine will be built and “[i]t is to be hoped that it will be built with taste by people who understand languages and computers well enough to know how little it is that they know.” (Kay 2003: 232)

In the meantime, less modest promoters and those involved in the development of MT will highlight only the positive parts. For instance, *Eclaco* promises that anything you say will be translated by their speech translator, although languages and quality are not mentioned.

Other noteworthy endeavours are Microsoft Office’s *Bing* translator, which offers instant translation from/to multiple languages without hearing any responsibility for the content and replaces with a single click the source text with the target text (built-in option of Revision tab), or the free *Glosbe Online Dictionary*, whose name may be misleading: although it is a ‘dictionary’, full sentences can be translated, even specifying the source (mainly the *Official Journal of the European Union* or *opensubtitles.org*). Allegedly, their translation memory is the largest online, containing more than 1 billion (10⁹) sentences in many languages and possibly many TMX files from volunteers.

Yet, Bar-Hillel’s remark 60 years ago still seems to be valid:

“... since we cannot have 100% automatic high-quality translation, let us be satisfied with a machine output which is complete and unique, i.e., a smooth text of the kind you will get from a human translator (though perhaps not quite as polished and idiomatic), but which has a less than 100% chance of being correct. I shall use the expression 95% for this purpose...” (Bar-Hillel 2003: 47)

Finally, it would not be nice to leave the patient reader without a prediction about the future. There are views that the frontier of MT will expand as far as speech translation, including speech recognition (similarly to text recognition, which is quite effective today), language translation and speech synthesis (Morimoto – Kurematsu 2003: 363). They estimate that we will be “able to speak freely without worrying about language differences at the beginning of the next century.” (Morimoto – Kurematsu 2003: 369).

Kis and Mohácsi-Gorove state that man cannot be fully replaced: the quality of machine translation will not reach the standard of a mediocre human translation for a very long time, so the bulk of translation is left for human beings. Legal, technical and other special texts are to be translated by human beings only, and even this has to be proofread (Kis – Mohácsi-Gorove 2008: 13). And we fully agree with that. Now let us take a look at CAT-tools.

7 Available at <http://www.microsofttranslator.com>, 12.02.2013.

Computer-Assisted Translation

At present, the most conspicuous products belong to the category of computer-assisted or computer-aided translation tools (CAT-tools), which should be clearly differentiated from machine translation.

The assistance of computers seems inevitable, as there are larger and larger translation jobs with shorter deadlines. There are more cases when a translation job is no longer the work of a single person but the contribution of a translation team (Kis – Mohácsi-Gorove 2008: 14), although freelance translators will still find their share in case they keep the pace with technology. Although Kay & Roschaisen observed in 1993 that “one of the most important sources of information to which a translator can have access is a large body of previous translations” (described in Bowker 2002: 92), this is not necessarily true today with the large number of closed corporate or free online databases.

Enhancing translation productivity has always been the concern of the European Union (EU), especially since new members were accepted. Minimizing translation costs became very important in case of legal documents in multiple languages, and the EU invested considerable sums in research on both CAT-tools and machine translation (cf. Kis – Lengyel 2005: 53–54).

As previously mentioned, machine translation produces a text “by itself”, awaiting human pre- and post-editing, whereas CAT-tools contribute to the quality translation by offering a full translation environment to human translators. Thus, we are going to use translation environment (TE) referring to these tools in the following. Present-day TEs provide a general framework for translation projects. Once the software is started, the translator has to create a project (name, source and target languages, client, domain, etc.), offering further possibilities: importing/adding one or more source texts, creating/using one or more translation memories, creating/using one or more term bases (standard options), but some TEs already offer very advanced word-processing (even in case there is no MS Office separately installed on the computer⁸), the possibility to use various online resources,⁹ including optional spell-check dictionaries (in multiple languages), term banks (as plug-ins), or even more than one machine translation aid (as separate add-ins). The TM part includes three important parts: an aligner where memory files are loaded, a format and tag management system, and a translation memory management system (Gouadec 2007: 272). Overall, statistics show that “65% of translators surveyed agree that using a CAT-tool allows them to work up to 40% faster.”¹⁰

As for the TB part, TEs can already retrieve terminology automatically, offering the translator copy and paste option, as well as pre-translate terms by identifying them in the term base and automatically inserting them into the corresponding

target segment (cf. Bowker 2003: 56–57). A distinct (“perennial”, cf. H. Somers 2003c: 15) problem of TBs is the management of diacritical marks; even if the UTF-8 encoding¹⁰ should solve all these problems (see, for instance the five Romanian diacritics: *ă, â, î, ș, ț*, or the Hungarian *á, é, í, ó, ő, ű, ú, ű*; the ones in bold are the most problematic), sometimes we still encounter problems, resulting in a string of weird signs and symbols (cf. archiving documents whose titles contain diacritics). During the entire translation project, the translation can follow project statistics: total number of segments, translated and/or accepted (confirmed) segments, number of repetitions in the source text, number of matches against a database of TBs and TMs with excellent filtering and sorting options. As Somers (2003c: 29) describes it:

“[...] the translator’s workstation represents the most cost-effective facility for the professional translator [as] it makes available to the translator at one terminal [...] a range of integrated facilities: multilingual word-processing, electronic transmission and receipt of documents, spelling and grammar checkers (and perhaps style checkers¹¹ or drafting aids), publication software, terminology management, text concordancing software, access to local or remote term banks, translation memory [...] and access to automatic translation software to give rough drafts. The combination of computer aids enables translators to have under their own control the production of high quality translations.”

As a result, TEs do not translate “instead of the humans”, as is the case of MT, which is an important difference. The cardinal question in the case of MT was/is the replacement of human translators, whereas in the case of TEs we can focus on how these tools can “support translators by helping them to work more efficiently” (Bowker 2002: 4) and increase translation productivity. Therefore, whereas MT systems try to replace translators, CAT-tools, TEs assist human translators. They provide full word-processing in the target-window, including the pre-editing possibility of the source text in case the translator would like to change something (i.e., typo mistakes, segmentation).

Even if they are extremely functional with many features, they are as user-friendly as possible, and regularly it takes less than a week to learn their basic handling. According to a 2013 ProZ survey, ease of use and ease of learning to use are of the highest priority when deciding upon a CAT-tool. Naturally, experience comes while translating, which is the case with TEs as well. They are designed in a way to be able to use them on almost all types of computers and notebooks, as their need for system resources has improved considerably. Some computers might open TEs more slowly, but once the project is started we should not experience any kind of delays during translation, except for exporting the newly translated document.

Many TEs have a trial version, which enables the interested ones to test their (usually) full functionality for a limited period of time (thirty, forty-five, sixty days).

⁸ Encyclopaedias, specialty web pages, other reference works.

⁹ <http://www.translationzone.com/en/translator-products/sdl-trados-studio-freelance/>, 18.04.2013.

¹⁰ <http://www.utf8-chartable.de/>, 16.04.2013.

¹¹ For instance, an activated list of forbidden terms (slang, taboo, swear etc.).

Our recommendation here would be to try during this period the possibility of using previous translations (including both TBs and TMs), saving options for newly created TBs and TMs, as well as back-up options, having in mind that “competing products are largely based on the same underlying principles” (Bowker 2002: 8).

Interoperability is also very important, as the chosen TE should handle all the document types leading TEs use. This refers to both TB and TM files, as well as to the format of the entire project. For the time being, the market leader is *SDL Trados Studio 2011*; thus, compatibility with this TE is definitely a must.

Last but not least we should check the retail price of leading¹² TEs and see what the price includes. Some TEs offer free upgrades “for ever”, others heavily charge on us for a newer version, whereas there are TEs with considerable discounts upon purchasing a newer version. Although price is only the sixth aspect to consider (cf. ProZ.com survey from 2013), we should mention that this is a global survey, and there are regions where price may be more important than the fullness of features, speed or ease of use.

In our opinion, the best idea to test their help would be to check a TE completely free of charge for a longer period of time. The most popular of freeware TEs is probably *OmegaT*,¹³ featuring multiple TMs and TBs, multiple-file projects, interface to *Google Translate*, glossaries, *MediaWiki*, around thirty file formats, non-Latin alphabets, spelling checker, and most importantly, compatibility with leading proprietary TEs (TMX, TTX, TXML, XLIFF, SDLXLIFF formats). Beginners should definitely try *OmegaT* to understand the basic idea behind TEs: the TM of any TE starts from scratch, that is, completely “empty”. A further advantage of *OmegaT* is that it runs on both *Microsoft Windows* and free operating systems (cross-platform software).

Kis and Lengyel (2005: 55) discuss the major benefits of empty TMs: they only “learn” from the translator such as the style or consistency. Bowker (2003: 58–63) dedicates long pages for the gain side of TBs, mentioning quality, speed, flexibility, shareability of information, term-extraction tools, and stop lists (ignore).

However, there are plenty of drawbacks as well: time-consuming, conservative, possible wrong terms, difficulty in joining TBs from different translators in case of large projects, let alone the conformation to the translators’ norm (B. Kis – Lengyel 2005: 55). This is the reason why large clients already offer TMs with database.

The segment matching feature of translation memories is probably the most interesting aspect of them while translating. Matches are based on the principles of ‘fuzzy logic’ (Baker 2001: 136), resulting in various types of matching in percentage. Of course, it often happens when there are no matches (pre-translation turns up no results), but in more fortunate cases we can differentiate various matches detailed below.

Exact matches refer to 100% identical segments (no further comment necessary), but strange as it may seem, there is also the possibility to surpass this match. For instance, *MemoQ* TE may also signal 101.9% matches, which means that the preceding/following segments also match. This may be considered as *context*, which leads us to Albert’s (2011: 171) observation: in his view, terms taken out of sentences or sentences out of a text leave them only a virtual semantic meaning, which may turn “live” or current when inserted back into a sentence or a text. Decontextualized terms or sentences do not add to the global meaning of a text, which may be the reason why TEs can offer two possible views while translating. *Full matches* differ from exact ones only in the so-called variable elements (“placeholders” or “named entities”), usually including numbers, dates, times, currencies, measurements, and various proper names; these elements require some kind of special treatment in a text (Baker 2001: 136).

Fuzzy matches may range between 99–1%, which is important to follow, as we can differentiate at least two types of threshold. Firstly, the translator may set a minimum sensitivity threshold to be shown by the TM (typically above 50%, around 75% or 90%); secondly, the client–translator agreement may also contain a sensitivity threshold of above 60 or 70%, leading to a differentiated price range for the segments translated.

Bowker further mentions *term matches* (when no exact or fuzzy matches are found, but there is a chance for translation equivalents for individual terms), *sub-segment matches* are situated between fuzzy and term matches (2002: 96–106).

Apparently, matches in TE triggered a modern Pandora’s Box: as translators are typically paid by the word, line, page, hour, or more recently, by the character. However, some clients take into consideration the match percentage not to pay translators full rate, arguing that re-using previous matches is not translation (Bowker 2002: 121). Translators fight back that building a TM providing good matches takes up a considerable time, and it improves consistency; furthermore, even 100% matches have to be proofread and, in fact, “customers should pay extra for this improved quality” (Bowker 2002: 122). As a final thought, we would like to present two pieces of statistics connected to CAT-tools. According to a 2006 survey,¹⁴ out of 874 translation professionals from 54 countries, 35% of translators use *Trados*, 17% *Wordfast*, 16% *Déjà Vu*, 15% *SDL Trados*, 4% *SDLX*, 3% *STAR Transit*, 3% *OmegaT* and 7% others. In 2013, we have fresh data from ProZ.com by Jared:¹⁵ *SDL Trados* has grown to 43.2%, *Wordfast* is still second, but with only 11.5%, whereas *memoQ* came third with 8% (it was not even listed in 2006!). *Déjà Vu*, *Across*, *SDLX* and *OmegaT* still share the pie, but their significance is minor.

12 If the chosen TE is not among the top ones, we will definitely encounter issues of compatibility regarding importing/exporting TB or TM.

13 <http://www.omegat.org/en/omegat.html>, 16.04.2013.

14 Imperial College London Translation Memories Survey, <http://www.webcitation.org/query?url=http%3A%2F%2Fwww3.imperial.ac.uk%2Fportal%2Fpjs%2Fportal/live%2Fdoc%2F1%2F7307707.PDF&date=2009-09-18,19.04.2013>.

15 <http://proz.com/blog.com/2013/03/28/cat-tool-use-by-translators-what-are-they-using/>, 17.04.2013.

Conclusions

We have covered a long way from a seemingly rational fear and threat against CAT-tools or translation environments. Bowker talks about these attitudes as belonging to the past (2002: 120–121), even offering a psychological explanation:

“No translator wants to have the feeling that he or she is going to be replaced by a computer or be reduced to someone who merely has to click on the ‘OK’ button to accept a ready-made translation. Clients need to be made aware of the limitations of the tools.”

Now we have high hopes that this fear is at least reduced, and the majority of translators and clients are able to understand what lies behind. Yet, a constant education of both translators and clients is more than necessary due to the technological changes, but there is no room for complaints: countless courses dealing with all conceivable aspects of modern translation are organized worldwide, which are extremely easy to track and follow on the Internet. A good clue might be to check the official web page of national translators’ and interpreters’ associations, and become a member. Membership will result in various advantages, but we would only highlight here the knowing of many important events and organizations connected to improving our technical knowledge.

However, not only clients but also translators are interested in prices of translations, which may be influenced by or connected with the purchase price of TEs. In this respect, Haynes recommends that TE providers should offer their products “freely available to translator-training institutes”,¹⁶ having in mind that “these developers may well benefit financially when the students graduate and are in a position to influence corporate purchasing decisions regarding translation technology” (Bowker 2002: 136). Proprietary vendors may also benefit from this relationship from the very beginning, as beta products (before the final release) can be tested with the help of a number of students.

One may wonder if there is any more room left for development in translation technology. Indeed, it is hard to imagine what way this field may evolve, but experts offer a few clues. Somers predicts spoken-language translation (SLT), which combines speech understanding and translation (Somers 2003a: 7), or linguistically more “sophisticated tools”, which may be able to globally change the style of a text (*purchase to buy*).

The proper search and correction of “false friends” in translation may be another improvement. Somers discusses the case of the French *librairie*, which does not match the English *library* (2003c: 16). A bilingual concordance of a corpus makes it possible to check false friends even today, although this is not yet an add-in feature of TEs. In the near future, we may be offered suggestions to inspect the English

library versus *bookshop* (consistency-enhancing feature), the Romanian *librarie* (bookshop) and *biblioteca* (library), *roman* (novel) versus *nuvelă* (short-story) or the Hungarian *helyiség* (settlement) versus *helyiség* (room). To a certain extent, this is already present as corpus research is more and more effective. As a result, large parallel corpora are at the reach of the hand with correctly aligned pairs.

Launching new websites is almost inconceivable today without multilanguage support, especially when the site accompanies product release, called simultaneous shipment (cf. “simship” in Bowker 2002: 12–13). Although large companies produce high-quality website content in multiple languages, the content of smaller ones is often questionable.

As a result, if we tend to remain competitive in the field of translation, we should consider how MT and CAT may contribute to our success.

References

- ALBERT, S.
2011 “A fővényre épített ház”. *A fordításelméletek tudomány-és nyelvfilozófiai alapjai*. Budapest, Áron Kiadó.
- ALLEN, J.
2003 Chapter 16. Post-editing. In: H. L. Somers (ed.), *Computers and Translation: A Translator’s Guide*. John Benjamins Publishing Company. 297–317.
- ANDERMAN, G. M.–ROGERS, M. (eds.).
2003 *Translation Today: Trends and Perspectives*. Multilingual Matters.
- BAKER, M.
2001 *Routledge Encyclopedia of Translation Studies*. Routledge.
- BAR-HILLEL, Y.
2003 The Present Status of Automatic Translation of Languages. In: S. Nirenburg, H. L. Somers–Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 45–76.
- BEAUGRANDE, R. D.–DRESSLER, W. U.
1981 *Introduction to Text Linguistics*. Longman.
- BIAU Gil, J. R.–PYM, A.
2006 Technology and Translation. A Pedagogical Overview. In: *Translation Technology and Its Teaching (with much mention of localization)*. Tarragona, Intercultural Studies Group. 5–19.
- BOULTON, C.
2010, March 2. Google Chrome Gets Machine Translation, New Privacy Features. *eWeek.com*. Retrieved from: <http://www.eweek.com/ea/Web-Services-Web-20-and-SOA/Google-Chrome-Gets-Machine-Translation-New-Privacy-Features-839506/>

16 In this respect, I wish to express my gratitude and thanks to *Klgray Translation Technologies* for supporting the Sapientia University, Faculty of Technical and Human Sciences in Targu-Mures, Translators and Interpreters specialization since 2009.

- BOWKER, L.
2002 *Computer-Aided Translation Technology: A Practical Introduction*. University of Ottawa Press.
- FREIGANG, K.-H.
2001 *Machine-Aided Translation*. In: M. Baker (ed.), *Routledge Encyclopedia of Translation Studies*. London and New York, Routledge, 134-136.
- GOUADEC, D.
2007 *Translation as a Profession*. John Benjamins Publishing.
- IMRE, A.
2013 *Traps of Translation*. Braşov, Editura Universităţii "Transilvania."
- KAY, M.
2003 The Proper Place of Men and Machines in Language Translation. In: S. Nirenburg, H. L. Somers Y.-A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 221-232.
- KIS, B.-LENGYEL, I.
2005 A fordítás számítógépes eszközei. In: "Mindent fordítunk, és mindenki fordít". *Értékek teremtése és közvetítése a nyelvészetben*. Bicske, Szak Kiadó. 53-59.
- KIS, B.-MOHÁCSI-GOROVE, A.
2008 *A fordító számítógépe*. Bicske, Szak Kiadó.
- MORIMOTO, T.-KUREMATSU, A.
2003 Automatic Speech Translation at ATR. In: S. Nirenburg, H. L. Somers-Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 363-370.
- NIRENBURG, S.
2003 Introduction. In: S. Nirenburg, H. L. Somers-Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. XI-XV.
- PIRON, C.
1994 *Le défi des langues - Du gâchis au bon sens*. Paris, L'Harmattan.
- PRÓSZÉKY, G.
2005 Egy újelvtű gépi fordítási rendszer felé. In "Mindent fordítunk, és mindenki fordít". *Értékek teremtése és közvetítése a nyelvészetben*. Bicske, Szak Kiadó. 79-83.
- REIFLER, E.
2003 The Mechanical Determination of Meaning. In: S. Nirenburg, H. L. Somers-Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 21-36.
- SOMERS, H.
2003 Chapter 1. Introduction. In: H. L. Somers (ed.), *Computers and Translation: A Translator's Guide*. John Benjamins Publishing Company. 1-12.

- SOMERS, H. L.
2001 *Machine Translation Applications*. In: M. Baker (ed.), *Routledge Encyclopedia of Translation Studies*. London and New York, Routledge, 136-139.
- WEAVER, W.
2003 Translation. In: S. Nirenburg, H. L. Somers-Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 13-17.
- WILKS, Y.
2003 Introduction. In: S. Nirenburg, H. L. Somers-Y. A. Wilks (eds.), *Readings in Machine Translation*. A Bradford Book. 203-206
- ZETZSCHE, J. O.
2009 Let's Talk: Trados and the Google Translator Toolkit. *The ATA Chronicle*, 38(10) 18-21.
- Online resources**
- [http://hu.wikipedia.org/wiki/Wikipédia: Mi nem való a Wikipédiába?](http://hu.wikipedia.org/wiki/Wikipédia:_Mi_nem_való_a_Wikipédiába?), 10.02.2013.
- <http://phiendichvien.com/fags/253-machine-translation-do-we-need-translators.html>, 10.02.2013.
- <http://prozomblog.com/2013/03/28/cat-tool-use-by-translators-what-are-they-using/>, 17.04.2013.
- <http://translate.google.com/>, 12.02.2013.
- <http://www.axistranslations.com/translation-article/what-is-translation.html>, 10.02.2013.
- <http://www.eweek.com/c/a/Web-Services-Web-20-and-SOA/Google-Chrome-Gets-Machine-Translation-New-Privacy-Features-639506/>, 13.02.2013.
- <http://www.microsofttranslator.com>, 12.02.2013.
- <http://www.omegat.org/en/omegat.html>, 16.04.2013.
- <http://www.translationzone.com/en/translator-products/sdl-trados-studio-freeance/>, 18.04.2013.
- <http://www.utf8-chartable.de/>, 16.04.2013.
- <http://www.webcitation.org/query?url=http%3A%2F%2Fwww3.imperial.ac.uk%2Fportal%2Fpls%2Fportallive%2Fdocs%2F1%2F7307707.PDF&date=2009-09-18>, 19.04.2013.