USING CAT SYSTEMS AND/OR CORPORA IN THE TRANSLATION PROFESSION

OANA TATU[1]

Abstract: Driven by the increasing requirement for efficiency at the workplace, professional translators worldwide resort to what is called "automated assistance", which allows for more or less control over the end product, while it attempts at improving, not always successfully, their work quality. The paper outlines the use of electronic translation tools by the Romanian professional translator. Thus, we analyze Computer-Aided Translation (CAT) Systems which prevail as assisting tools in the translators' work, and we consider the role of electronic parallel corpora, a rather marginal one, as it is. The paper highlights the obvious potential of electronic corpora, their advantages over traditional CAT systems, and asserts the need for an integrative approach to translations which should only be beneficial to the translated work as result of the translation profession.

Key-words: CAT systems, corpora, translation, professional.

1. INTRODUCTION

While the translation of a general text between one source language and one target language remains an actual challenge for both human and machine translators, the translation of special domain texts presupposes significant proficiency in handling both the content and the representation of that domain knowledge.

Indeed, the professional translator who deals with specialized texts on a daily basis is, just like his/her ancestors, a truly gifted person, he/she is a proficient linguist, but, given the pace of technological evolution nowadays, he/she must possess several extra traits among which one stands out as an umbrella term: the ability to be resourceful. Being resourceful means first and foremost being inventive and creative, but the semantics of the term also implies the capability of using resources which must be an inherent skill for a professional translator today. Several resources that all translators, the Romanian professional translator included, use are: dictionaries, in-house glossaries, reference materials, terminological databases, and very rarely, unfortunately, corpora. Along with these, translators need to improve and facilitate their work by using several tools, among

[1] *Transilvania* University of Brasov, oanatatu_brasov@yahoo.com.

RRL, **LVI**, *2*, p. 167–180, Bucureşti, 2011

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

which the most common are: the ordinary pen (gone almost extinct as it is), the computer, the printer, software programs, alignment tools, translation memories, and concordancers.

The reason why these tools aroused so much debate and stirred so many spirits is that among them, there is one which subordinates all others by becoming their support tool, and by turning into an almost indispensable instrument. The tool in point is the computer, which, in Martin Kay's words is *a device that can be used to magnify human productivity. Properly used, it does not dehumanize by imposing its own Orwellian stamp on the products of the human spirit and the dignity of human labour, but, by taking over what is mechanical and routine, it frees human beings for what is essentially human.* (1980:1). What Kay advocated here was, on the one hand, the necessity for translators to use a computer in their work, and, on the other hand, the control that professionals or practitioners must always possess upon the tools they employ.

Starting from here, the purpose of this paper is two-fold: firstly, we shall review the main tools that the Romanian professional translator currently uses, (machine translation, CAT-tools, dictionaries), those he should use (corpora), and show their actual interdependence; secondly, we shall proceed with a comparative analysis of several such tools (mainly translation memories and corpora) following the lines of a few organizing criteria. What actually underlies our research is a suggestion or a plea for the integration of these fine linguistic instruments we call corpora into the CAT tools used by the Romanian professional translator, given their advantages enhanced by their use along with other CAT tools.

2. THE TRANSLATOR'S TOOLS

2.1. Machine Translation (MT)

Although it may seem beyond the scope of our paper, we shall begin our analysis with machine translation (MT) proper, which was extensively used until recently in the Romanian professional setting, and also constitutes the basis upon which several computer assisted tools were developed or against which new methods were invented.

What machine translation basically did was to evict the human translator almost completely from the translation process, and this with the intention of turning the translation process more profitable both money-wise and time-wise.

Yehoshua Bar-Hillel, who was the first full-time MT researcher, was also the one who discovered the flaws of Fully Automatic High Quality Machine Translation (FAHQMT). Given the fact that languages are not mere lists of words, and that the process of translation involves more than just a simple matching and

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

substituting operation, going beyond to profound analysis and understanding of texts, MT proved quite impracticable and impractical as such. Thus, Bar-Hillel amended his orientation claiming that if the target text needs to be highly accurate (doesn't it always need to be so?) MT has to be given up in favour of mixed MT, namely a translation process in which the human takes part (1951:12). However, given the comfortable position the machine had acquired, it was very difficult to have it surrendered, and thus, humans were only suggested to intervene either before the translation process, or after it, but never in the middle of it. However, practice proved that pre- and post-editing procedures were both expensive and time consuming, so researchers and language practitioners had to turn to something different.

Let us just observe what would happen if no human intervention were allowed and we let the Google translator (for instance) translate in our stead, and try to imagine what it would mean to post-edit such a translation. Doing it from scratch would take less time.

If trapped under debris

- *Do not light a match.*
- *Do not move about or kick up dust.*
- *Cover your mouth with a handkerchief or clothing.*
- *Tap on a pipe or wall so rescuers can locate you. Use a whistle if one is available. Shout only as a last resort. Shouting can cause you to inhale dangerous amounts of dust.*

(FEMA instructions *"What to do during an earthquake?"*)

Dacă prinse sub dărâmături

- *Nu lumina un meci.*
- *Nu mişca sau vor ridica praful.*
- *Acoperă gura cu o batistă sau îmbrăcăminte.*
- *Apăsați pe o țeavă sau de perete, astfel încât salvatori poti localiza. Folosiți un fluier dacă aceasta este disponibilă. Strigă numai ca o ultimă instanță. Strigând poate provoca să inhaleze cantități periculoase de praf.*

(The Google Translator − http://translate.google.com/)

In recent years, machine translation was rendered useful and practical once again as it came to be supported by corpora, that is, by collections of texts. Thus, corpus-based MT systems rely on translation rules that are learned as such from human translators. As Pierre Isabelle asserted in his *Machine Translation: Overview*, *"the crucial resource for corpus-based MT is a parallel corpus: a collection of documents that exist in both the SL and the TL, on which the system can be trained and tested."* (Isabelle, 2006:412). Naturally, the entire mechanism of MT relies in this case on the perfect alignment of sentences between two languages, and even then breakdowns might occur as exact matches might solely be found with simple and short sentences; thus, the corpus based method is rendered less precise. Also, the fact that it can only be applied within the domain of the training corpus is another drawback. For instance, supposing that the training

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

corpus were the Canadian Hansard (an English-French corpus of parliamentary proceedings) only within this specific field can MT be expected to provide good results. Otherwise, the chance of accurate translation lies, once again, with the computer.

To draw a partial conclusion here, MT has already changed the way professional translators work, but will not replace human beings. Today, MT can be used as a tool to supply professional translators with quick-on-the-fly versions that need thorough proof-reading. In light of all this, it is quite understandable why *the current share of everything that could be called MT, pure or mixed, is well below 1% of the total translation market* (Isabelle *et al*., 1993:206).

As the experience of passing the responsibility for accurate translation to the machine proved inconsistent and not feasible, the approach had to be reviewed. Instead of looking for human support for the machine (namely Human Aided Machine Translation − HAMT), why not reverse the entire process, and return to square one, that is to the translator, and look for ways to support him (Machine Aided Human Translation − MAHT). This is how the concept of CAT tools emerged to include all electronic support tools for translators. Actually, although the term is fairly recent, it echoes an older concept, dating back to 1980 when Martin Kay advanced the idea of a translator's workstation which he called "translator's amanuensis" – meaning the translator's manual labourer, a sort of factotum device. Although he imagined the entire thing before the advent of personal computers, Kay had a brilliant intuition placing at the heart of his amanuensis a sophisticated text-editor, assisted by a split-screen device, displaying two cells, both for the source and for the target languages.

What is interesting to notice here, besides the brilliance of the idea, is that Kay inferred the need for integrating several tools in order to assist the translator; this is what we shall eventually show in our paper: that there is neither absolutely perfect, nor fully flawed CAT tool, but rather the simultaneous use of several such tools, chosen in agreement with the translation task, should be of real help to the translator.

2.2. Translation Memory (TM)

The tool that is thought of as being the most reliable one, and that seems to best accommodate the Romanian professional translators' needs is the translation memory (TM) which is actually a localization tool that maintains a database of source and target language pairs, and retrieves automatically the translation of those units in a new text which occur in the database. The idea behind TM systems is to store the original source content and its translation produced by human translators; however, it does not store entire documents, as documents are broken down into manageable, usually sentence long units, which are archived as stand alone units. The most frequently used TM system in Romania is TRADOS with its Translator's Workbench.

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

Typically, apart from the memory as a core element, a TM system also comes with several other tools such as: a terminology management program, an automatic term-recognition feature, a concordance tool, a multilingual editor and an alignment tool. This latter element, the alignment tool, helps create the TM as it compares ST and TT, matches corresponding elements, and pairs them up as units in the TM.

As said before, the TM database consists of sentence units which are searched for similarity. However, where there is need for smaller units correspondence (e.g. word correspondence) the terminology management tool pops in and offers its pre-stocked terms in order for the matching process to start.

As one can easily notice, there is a high degree of automation with TMs too. However, a TM is not to be mistaken for MT even though they are related up to a certain point, as TMs seem to have stemmed from a particular case of MT, namely the Example-Based Machine Translation (EBMT). Just like the EBMT, the TM looks for the retrieval of best matches for the sentences in the ST, from a bilingual archive containing sentence alignments of ST and TT units. Nevertheless, with TMs, the match is offered to the human translator, who is free to accept or to deny it, while the EBMT automatically builds up translations by recombining elements in the database.

Let us dwell a while longer on the sentence as the unit which is, by default, being stored in the memory. Given such circumstances, a TM will perfectly work for texts where sentence repetition is to be expected (such as technical manuals, alternative/new versions of older documents). But, even there, repetitions of full sentences are not expected to frequently occur. The developers of commercial translation memory software remedied this issue by introducing what is called the *fuzzy matching* procedure. What the fuzzy match does is to locate units that are either identical or similar to the source language unit to be translated, rank them according to the degree of similarity which the translator himself had adjusted beforehand, and finally suggest them to the translator. The problem with such a procedure is that *"most TM systems have difficulty with fuzzy matching – either too many irrelevant examples are extracted, or too many potentially useful examples are missed"* (Hutchins**,** 2003:11). We shall return to this idea further on.

For now, let us just say that TM systems have become an essential translation instrument for the Romanian professional as well, and have largely substituted machine translation, as they allow translators more control over their work while simultaneously relieve them from routine, mechanical work.

Naturally, there is always room for improvement, and TMs might certainly be improved by their acquiring several features specific to corpora, as we shall see in what follows.

2.3. Terminology Databases and Dictionaries

In order to preface our discussion of corpora, we shall briefly recall the lexical aids that a TM system comes with, namely the terminology management

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

programs. What these programs do is to actually manage a terminology database consisting of the lexical equivalents encountered or researched during one's personal work. These programs bring flexibility to the translator's work as he can easily access terms, thus go below the sentence level offered by the memory itself, and, moreover, can share with other translators the results of their terminological research.

However, such a database is bound to have limits as the translator's work is limited itself, so when a term is queried that is not present in the database, some TM systems would open the entry in one of their other lexical aids, namely a bilingual dictionary.

Now, the Romanian professional translator is quite used to employing bilingual dictionaries which are, what else, but repertoires of lexical equivalents (general dictionaries) or terms (specialized dictionaries or terminologies) (Zanettin**,** 2002: 10). Oftentimes, it will certainly be useful for the professional translator to turn to such a dictionary if all that he needs is a definition or a few instances of usage. Yet, chances are that a definition will not suffice; so, as things stand right now in the Romanian professional setting, the translator is either at a dead end, or will provide an inaccurate translation.

Still, there is always a third option, which is usually the best, and what we refer to here are the parallel corpora and the imperious need to integrate them among the CAT tools that the Romanian translator employs on a daily basis.

2.4. Parallel Corpora

What are parallel corpora if not repertoires of translation equivalents and of translation procedures deployed by previous translators? Thus, instead of getting a mere definition in a dictionary, the translator searching a parallel corpus will encounter translation solutions to specific translation issues. Moreover, as Zanettin (2002) points out, besides finding several equivalent translation strategies, the translator will also find instances of non-equivalence, namely evidence of past translators dealing with the lack of direct equivalence at word level.

For instance, a sentence such as *To say the Moldavian monasteries are pure art would be an understatement* (tourist brochure) will represent a translational challenge given the already famous untranslatable term *understatement*. However, translators should make recourse to different strategies and deal with the lack of direct word correspondence:

A spune că mănăstirile din Moldova sunt artă pură nu este îndeajuns.

Este o afirmație modestă aceea conform căreia mănăstirile din Moldova reprezintă arta în stare pură.

Cuvintele sunt prea sărace pentru a exprima calitatea artistică a mănăstirilor din Moldova.

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

To provide a proper definition, parallel corpora are collections of original texts in one language aligned to their translations in another language or in several other languages.

The process of alignment principally follows the same procedure as with a TM, with a few essential differences: while with the TM alignment is achieved at sentence level, following the isolation of each sentence from the neighbouring ones, with the corpus the alignment is achieved first at section level, then at paragraph level, and finally at sentence level, with sentences never having to be broken up from their contextual environment.

It is a fact that one cannot translate a word or a sentence out of context, and this is the essential asset that a corpus brings to the work of a professional translator.

The software tool that a parallel corpus relies on is called a bilingual concordancer. The concordancer works by searching for the query in the database, and providing all the occurrences it finds, each in its full sentential context together with their aligned translations. From among these, the translator is free to choose the optimal translation variant.

Actually, the TM could be seen as a particular type of parallel corpus, or, better said, the parallel corpus could be a particular type of translation memory, as it is a computerized repository of past full text translations which promote translation recycling.

3. THE S.M.A.R.T. ANALYSIS

Once again, the purpose of this paper is not to determine which of these CAT tools are better suited for the translation work; it is rather a combination of all the above mentioned instruments that ensures the achievement of high-quality translation.

The truth is that it is only nowadays that corpora analysis expands its scope beyond language engineers or linguists in the academic setting to language practitioners in the professional setting. In what follows we intend to highlight several advantages that the integration of corpora among the most reliable of CAT tools might have for the Romanian translator as professional but, more importantly for the output of his work – the translation.

Our analysis is based on five criteria, namely **segmentation**, **matching**, **adaptability**, **reusability**, and **translator's input**, which are actually the most often debated upon issues when it comes to the integration of parallel corpora along with TMs in the translator's toolkit. As one can easily notice, the initial letters in the five criteria add up to the acronym S.M.A.R.T., which is a well-known guideline for successfully managing business projects. Although we have altered the original words to suit our purpose, we strived to preserve the acronym

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

itself given the fact that every professional translator is the manager of his own translation project, and for it to have a positive outcome, several SMART decisions need to be made.

3.1. S – SEGMENTATION

The issue of segmentation (or parsing, as it is called) has long been debated upon, and this is not surprising if we are only to think about the first piece of advice that we offer to our translation students, future translation professionals: *Do not translate out of the context! Read the entire piece first!!* The reason for our legitimate requirement is that for translation equivalence to be achieved between ST and TT it is not enough to sum up mathematically the SL corresponding equivalent units in the TL. Actually, that is exactly why FAHQMT failed as absolute translation method. Languages structure differently the information about reality, so that it is not fit for a machine to decide upon translation options, or to just replace one word in the SL for what it labels as accurate correspondent in the TL. It simply could not do that acceptably, for it lacks the ability to contextualize and the background knowledge or the common sense often required in translation tasks.

For instance, how would a machine translate an ambiguous sentence such as *Mind the doctor! – Ascultați sfatul doctorului!* or *Feriți-vă de doctor!* What about a *criminal lawyer*? Is he a murderer himself, or just a defender of murderers? Or, if we say *I'm dead serious* what is the machine to do? Pronounce me either dead or serious, for it certainly could not detect the collocational intensifying value of *dead*.

In a nutshell, (even here *într-o coajă de nucă*) the principle of word segmentation that machine translation applies to the ST will more often than not fail to yield satisfactory translation results, and, as a result, it was almost fully abandoned in favour of a different strategy.

The strategy we are referring to is the one employed by TM systems. As already mentioned, TM systems usually operate at sentence level which means that a new text is segmented into sentence-long units, these are extracted from context and exported to the database in isolation. Their origin, the document they used to be part of, and their place in it do not seem to be of any importance. The sentence is by default the main unit in a TM because this is the easiest way to build large databases of past translations. While such a degree of automation may work well for updated documents, for instance, where sentence repetition is a must, for other text types its limitations are striking.

This is indeed a high risk that the translator runs, as he tends to focus too much on isolated sentences, disregarding the context which yielded the sentences. The effect will be that textual features which are cross-sentence or above-sentence level are quite impossible to notice. The following examples of input sentences are purposefully excerpted from technical texts:

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

This *can be accessed via the Options menu.* ***This*** = acesta, aceasta; ***accessed*** = accesat, accesată.

It *is displayed in the Additionals area*. ***It*** = acesta, aceasta.

Unless the translator can access the larger context, the translation is 50% likely to fail.

A secondary effect of translating isolated sentences is that the current user working for a translation agency, and having access to an in-house TM system, will necessarily recycle previous translations by his fellow translators, which will turn his end product into a *"sentence salad"* (Bédard**,** 2000:45), that is a collection of different sentence translations by different translators.

Naturally, steps can be taken to prevent such situations from happening, and the answer comes from parallel corpora which possess a semi-automatic concordancing software. This allows the user to specify the length of units to be aligned, and operate adjustments, as appropriate.

In the instances provided above, the user would be granted direct access to both the single sentence unit and to the larger context (or the entire text) and can decide with certainty upon a translation variant. Thus, the larger disambiguating contexts of the above instances are:

Observe the route ***change****.* ***This*** *can be accessed via the Options menu.* (***aceasta***)

To open the list of traffic messages, select ***Information***. ***It*** *is displayed in the Additionals area*. (***acestea***)

Even though this advantage of corpora has not yet been exploited by TM manufacturers, and corpora are still being considered inefficient tools for fault of being less automatic, we suggest the implementation in the Romanian professional setting of this optional tool, perhaps as part of a TM system, so that situations such as those described above should no longer occur. Not only will corpora enable the translator to use a more flexible search tool (ranging from words to paragraphs), but it will enhance translation quality, as corpora will become themselves translation memories.

3.2. M – MATCHING

Although when pronounced, this word seems to suggest a mathematical procedure, the process of matching has not so much to do with exact sciences as one might expect. Rather, the process of matching units is based on establishing degrees of similarity between units.

As established before, matching at word level following a formal similarity between units could not credit machine translation as a reliable translation tool.

Above the word level, the TM is the one that automatically determines similarity between sentence units and operates the match. However, there are several issues which have aroused controversies regarding the kind of SL segments that qualify as exact matches of sentences in the TM database. For a TM system, an

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

exact match is a sentence which is repeated within the same document (disregarding the context), or over several documents.

When no such match is found, the TM employs what they call the *fuzzy matching* technique, by means of which similar matches are displayed on screen. The important thing to underline here is that the degree of similarity is judged in point of number of characters not of semantics. Thus, wherever morphological variance is involved, like inflection or compounding, TMs fail to accomplish their duty.

If our sentence to be translated is ***Open the file*** (*Deschideți documentul*), a TM system is more likely to provide as suitable match a sentence as ***Oven the film*** (*Introduceți filmul la cuptor*) – given the high degree of formal similarity between the two sentences, but for two characters – instead of matching it with *He is opening the file* (*El deschide documentul*) as closer match given the semantic similarity between our sentences.

Also, the professional translator using a TM is expected to adjust the degree of similarity in percents. However, if the search is set at a high degree (say 90%), possible useful but slightly different matches will be overlooked (a process known as "silence"). On the contrary, if the degree of similarity is set too low (like 30%) too many fuzzy matches will be excerpted (a phenomenon known as "noise").

There is one enlightening example offered by Elliott Macklovitch and Graham Russell (2000) showing that if our input sentence consists of 20 words and shares a 5-word sequence with a stored sentence made up also of 20 words, the TM programme will not be able to retrieve it as match given the low degree of similarity, namely 25% match (unless, of course, the degree of similarity is set that low, which is quite improbable).

By using a parallel corpus at this point, the translation performance will certainly be improved. Some might argue that it is tedious and time consuming to manually insert such extensive strings into the search bar, and eventually retrieve translation variants. This may be valid criticism but the quality of the end result is worth the effort. As far as morphological variation is concerned, the parallel corpus together with its concordancing tool will be able to override the issue successfully due to a particular device it employs, namely the part-of-speech tagging system which enables the translator to retrieve all instances of usage for a certain word, function of it being a verb, a noun, an adjective, a.s.o., without having to account for a different inflectional form. So, *file* and *is re-filing* would be matches as they both belong to the verb class.

3.3. A – ADAPTABILITY

Indeed, adaptability is a particular feature that all CAT tools should share at a given point, as it pertains to a higher idea, that of improvement, which should govern all man-made tools.

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

In point of adaptability, TM systems are quite limited as they can successfully be used only for limited applications (manuals, materials with repetitive sequences, or updated documents). It follows that TMs are fairly prescriptive in nature and quite task-bound, thus not allowing for much flexibility of use. From a different angle, they can be considered prescriptive also by the very nature of the matching procedure: they only suggest the highest-rank match, that is the match judged by a software program as being the most similar one to the input sentence. However, as Fifer concluded, the highest-rank match is not always the best translation, and 40% of translators enquired deemed the second ranked match as the most useful one. But, for several other matches to be available, the translator needs to follow more clicking steps, thus extra effort is involved (2007: 13).

Viewed from the adaptability viewpoint, the parallel corpora approach is quite descriptive, as it displays all matches at once, and the translator is invited to consult them, to check the strategies of other translators, and to select or not a part or an entire translation variant. Therefore, parallel corpora are more adaptable not only as they apply to a larger category of texts, but mainly because they allow the translator the freedom to choose, and to avoid being engulfed in a rather automatic decision making process.

We would thus favour an integrated approach where the search procedure in corpora is adapted to fit the speedy automatic retrieval in a TM, where inflected forms are treated as alternative word forms, not different ones.

For such a mixed approach to be possible at all, the TM system should become host for entire documents, and should be able to process entire passages rather than sentences in isolation. Impossible situations would thus be avoided, such as the translation of one sentence in the SL which is distributed over several sentences in the TL, or the other way around, situations which would not occur at all if the translator could himself adjust the alignment levels.

e.g. *You go ahead and do what you want. You will anyway.* – two sentences in the SL should be translated as one sentence in the TL:

Tu oricum faci întotdeauna ce vrei.

Such an integrated system should then override the formal approach to similarity, and adopt a semantic approach where the translator is required to make decisions and judge similarity with other than automatic tools (thus, *child* would match *children*, *close* would match *closing*, a.s.o.).

3.4. R – REUSABILITY

The truth is that the very basic idea of TM is fundamented on the concept of repetition of information allowing for recycled material to be used anew. However, as established above, very few text types lend themselves to such a degree of sentential repeatability that would justify the extensive and exclusive use of TMs within translation agencies. More valuable, unexploited information could certainly be found at levels above and below sentence, and only by combining the

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

capabilities of TMs with those of corpora/concordancers could the entire potential of previous translations databases be of real use.

Although the supporters of TM systems claim repeatability as their top advantage since, according to what they say, a translator does not need to translate the same sentence twice, the upholders of corpora search will retort that it is not so bad to translate the same sentence twice and in different manners, because once the context changes, the translation should logically change too.

It is one thing to translate a sentence such as

Remember! Push the red button first!

− for an electronics specialist:

Atenție! Apăsați întâi butonul roşu!

Or,

− for a child in a parent note:

Nu uita să apeşi întâi butonul roşu!

Besides the limited applicability of TM systems, and following from it, TM systems are also said to be proprietory (Zannettin, 2002) which means that they are created around specific projects within the professional setting, and are indeed helpful when used for translation and localization of updates, but are of little help when a new project is involved, displaying a different text-type or topic. It is here that the archive of different texts that a corpus represents may come in quite handy.

Finally, for texts to be recycled, one needs to make sure translation errors are corrected, and the memory does not surface as match a flawed sentence which has not been updated. It might be easier and more cost-effective to update entire texts than just sentences, so once again the recycling procedure of TMs should adopt a more flexible, corpus-like approach to it.

3.5. T – TRANSLATOR'S INPUT

This criterion actually translates as the professional's degree of control over the task to be fulfilled. And if we were to put in a nutshell, whatever explanations may follow henceforth, we would say that the overwhelming majority of CAT tools use an automatic computer search program, while with parallel corpora, the human end user is the one initiating the search.

In fact, all four criteria analysed above relate to this final one, and bring their own arguments as to why it appears to be best for a translation community to adopt an integrative approach to the matter of professional translation.

Thus, as far as segmentation is concerned, TM systems have by default the isolated sentence as translation unit, so the human translator has no say here. With corpora, the translator is free to decide upon the extent of the search unit, to view it in context or in the original document.

In point of matching, with TMs the translator is able to insert the degree of similarity needed for the text-type, he then can accept or refuse the proposed

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

match, but the process of acceptance or denial of match will never occur in full conscience, given the absence of context. With corpora, however, the level of segmentation is adjusted by the translator, several matches are provided and they only serve as reference material, so to say, for they can be consulted alone if not actually used.

As for adaptability, the translator should ideally feel as the adaptor of his own work; thus he should never be confined within the limits of a prescriptive, automatic pattern, because**,** eventually he will perceive this as a frustration coming counter to the creative enthusiasm a translator should always possess.

In point of reusability, as mentioned above, consistency in translation is not a must as long as the message is rendered accurately, and it is even undesirable if different text types, clients or translators are involved.

4. CONCLUSION

In the end, everything boils down to a simple conclusion: the professional translator everywhere in this world is the ultimate authority in a translation agency, and while he may feel comfortable for a while to be fully assisted by a repetitive device, soon he will remember to use his SMART human capabilities, those he has been trained to use: judgement, semantic competence, playfulness, creativity, originality.

REFERENCES

Bar-Hillel, Y., 1951, "The State of Machine Translation in 1950", *American Documentation,* vol. 2, 229−237.

Bédard, C., 2000, "Mémoire de traduction cherche traducteur de phrases", *Traduire, Societé Française de Traducteurs, Paris: INIST-CNRS,* Cote INIST, 186, 41−49.

Bowker, L., M. Barlow, 2008, „A Comparative Evaluation of Bilingual Concordancers and Translation Memory Systems", in: *Topics in Language Resources for Translation and Localisation*, Amsterdam, John Benjamins, 1−22.

Fifer, M., 2007, *The Fuzzy Factor: An Empirical Investigation of Fuzzy Matching in the Context of Translation Memory Systems*, University of Ottawa, Canada, MA dissertation.

Hutchins, J., 2003, "Machine Translation: General Overview", in: R. Mitkov (ed), *The Oxford Book of Computational Linguistics,* Oxford, Oxford University Press.

Isabelle, P., 1993, "Machine-Aided Human Translation and the Paradigm Shift", *Proceedings of the Fourth Machine Translation Summit***,** Kobe, 177−179.

Isabelle, P., 1992, "Bi-Textual Aids for Translators", *Proceedings of the Eight Annual Conference of the UW Centre for the New OED and Text Research,* University of Waterloo, Canada, 76−89.

Isabelle, P., G. Foster, 2006, "Machine Translation: Overview", in: *Encyclopedia of Language and Linguistics,* 2-nd edition, Elsevier: Gatineau, 404−422.

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)

Kay, M., 1980, *The Proper Place of Men and Machines in Language Translation*, CSL-80-11, Xerox PARC California.
Macklovitch, E., 2006, *Machine-Aided Translation: Methods*, University of Montreal, Elsevier Press.
Macklovitch, E., G. Russell, 2000, *What's Been Forgotten in Translation Memory*, in: J. S. White (ed.), "Envisioning Machine Translation in the Information Future", 4th Conference of the Association for Machine Translation in the Americas, AMTA 2000, J.S., Berlin: Springer Verlag.
Reinke, U., 2006, Translation Memories, Amsterdam, Elsevier Press.
Zanettin, F., 2002, *Corpora in Translation Practice*, in: E. Yuste-Rodrigo (ed.), *Language Resources for Translation Work and Research*, LREC 2002 Workshop Proceedings, Las Palmas de Gran Canaria, 10−14.

BDD-A373 © 2011 Editura Academiei
Provided by Diacronia.ro for IP 216.73.217.42 (2026-11-16 16:58:07 UTC)