DIGITALISED DELIVERY OF LANGUAGE. AN APPROACH TOWARDS MACHINE TRANSLATION TECHNOLOGY

Ioan Beniamin **POP**
Technical University of Cluj-Napoca, Faculty of Letters,
North University Centre of Baia Mare, România
pbeniamin@yahoo.com

Keywords: machine translation, automated translation, built-in linguistic rules, statistics-based, rule-based

Abstract: *Unquestionably, in today's language professionals, the use of information and communication technology (ICT) is already taken for granted and electronic handling as well as digitalized delivery of language services are standard client facilities. The present paper, while taking into the discussion the progressive stages of machine-assisted human translation (MAHT) and human-assisted machine translation (HAMT), analyses the present-day realities of Machine Translation.*

1. INTRODUCTION

For a considerable amount of time, humans have been struggling to accurately translate a statement from one language into another. It may appear forthright at the outset since translation is the process of converting or transposing a text from a source-language into a target-language one. The process, however, is not as straightforward as it seems as it gives rise to a complex situation. In fact, several questions may appear at this initial stage. Some of them can be asked along the lines articulated by Poibeau (2017:13): 'What does it mean to "transpose a text"? How do we go from a source language to a target language? How does one find equivalent expressions between two languages? Should the translation be based on words, chunks of words, or even sentences? And, more fundamentally, how can one determine what the meaning of a text or an expression is? Does everybody have the same understanding of a text? If not, how can this issue be handled in the translation process?' [1].

2. ELECTRONIC TRANSLATION TOOLS

One of the most difficult subtasks is to choose the optimal translation from among numerous viable alternatives for each source word. Bilingual lexicons have been progressively replaced by modern electronic translation tools. However, the term ‘electronic translation tools’ does not always mean ‘machine translation’ (MT). There are many different types of translation tools, and MT systems are only one of them. Indeed, it is debatable whether MT systems produce translation at all since they ignore the communicative, cultural, and encyclopaedic components of translation. As Hutchins (1995:431) notes, ‘although the ideal goal of MT systems may be to produce high-quality translation, in practice the output is usually revised (post-edited)’ [2].

The widespread usage of current communication technology has triggered numerous changes in today’s communication processes. In the recent years, it has become increasingly evident that there exists tremendous expansion in information technology, which has brought with it benefits such as swiftness, simplicity of use, and, perhaps most of all, the almost astonishing cost-effectiveness when considering the advantages. Telecommunications, networking, and computer businesses have all worked together to convert industrial civilization into the Information Society we know at present. Simultaneously, as the global market develops, industry and commerce operate on a larger scale than ever before. This phenomenon in itself entails greater autonomy and flexibility in terms of product and service exchange. There is absolutely no soul nowadays who can imagine daily life without a form of a personal computer, since it has grown into a versatile communication and information-processing device that is utilised in all activities. The computer is turning into a universal standard for communication and collaboration. With its ubiquitous availability to knowledge and fast connection amongst users, people all over the world have been given physical and geographical flexibility that was before unthinkable, all due to the internet and the freedom it has generated for translators.

Nevertheless, computer technologies are no substitute for human translation. There is no computer-assisted process that can turn a bad translation into a good one. It has become evident over the years that the end product is far from a flawless translation, but the merit still remains: computer technologies are perfectly capable of producing translations, be they as imperfect as they are. When utilized correctly, however, the right tools may assist competent translators in improving the productivity of their work as well as, at the same time, making it more accurate. Undeniably, the key criterion in terms of evaluating electronic translation technologies is quality. This is still achievable when considering literary translations since they pose exceedingly significant challenges for MT to accurately identify the subtleties and intricacies or poetry, for instance. However, when it comes to commercial brochures, technical manuals, scientific texts exhibiting rather technical and specialized language, the computational analysis are well within the reach of MT systems.

The obvious reality of our society is that the usage of information and communication technologies in the lives of today's language professionals is a fact. It is no longer an issue of whether translators should employ computers and networks. Exposure to the correct techniques in utilizing electronic tools will result in bottom-up automation of the translator's workplace, allowing translators to considerably improve the quality and efficiency of their professional services. Language services are now handled electronically and delivered digitally as ordinary client services.

Computer concordances are other electronic tools which can be employed to handle texts for translation. They are especially practical when translating texts containing specialized materials, exhibiting vocabulary specific to technical contexts, for instance. Concordances therefore entail algorithms creating word-processing programs that generate all the specific occurrences within a certain corpus having as main purpose the identification of patterns which would be illegible to the human eye. They would thus give the translator more control over the text, regardless of it being short, long or complex, and, thus, providing terminological consistency.

Instant access to relevant and trustworthy online and offline material has emerged as a critical issue in the retrieval of encyclopaedic and linguistic knowledge required for the activities at hand. The increasing need for high-quality translations of technical materials cannot be addressed without the use of computer-based technologies. There are various advantages to using electronic resources. The print culture has been gradually replaced by a new one, a screen culture which retrieves documents and information straightforwardly. All the data are promptly accessible and are well within reach. Due to the rapid production of domain knowledge in some creative subjects, most of the material of specialised tools which appear in paper form, such as dictionaries for instance, may have become obsolete even before the volumes find their place on the shelves or desks. Electronic dictionaries, on the other hand, - available in a variety of formats, including computer software, CD-ROMs, or through the Interned- may be published instantly online or offline, and they can be readily updated through the Internet. That is why many specialized encyclopaedias and scientific publications, which have long been useful resources for translators, are no longer printed and are only available online. Czulo and Hansen-Schirra (2017:3) note that 'translation scholars use corpora and strive for empirical models of the translation process (including translation strategies or specific properties of translated text). For professional translators, multilingual corpora serve as reference works that enable quick interactive access and information processing' [3].

The range of electronic translation tools includes a wide range of cutting-edge computer programmes. Spellcheckers as well as machine translation systems, word processing software, and terminological databases, electronic encyclopaedias or online dictionaries, HTML editors and software localization tools, they are all included. The range of computer translation technologies available may be as perplexing as it is impressive. Machine translation is the application of computer and language sciences to the construction of systems that meet practical demands, rather than a field of abstract intellectual investigation.

Human translation is often avoided because it is seen to be too costly. One of the possible reasons for it being rather expensive is in part due to the fact that a person's output is essentially restricted. The present reality has shifted rather dramatically, and the once-familiar antiquated notion of a solitary translator holding just a pen in his hand or aided by an old typewriter, while, at the same time, walled in by already outdated books, is no longer realistic. The need for translations is now unmet for several reasons, one of the possible explanations relying on a lack of human translators, or, if we consider the beneficiaries of the translated output, there are times when they do not view translation as a difficult action requiring a great degree of ability, and hence are unwilling to meet the financial demands on the part of the human translators. To that extent, nonetheless, the concept of an autonomously operating, always free translation machine is similarly impractical and will not become a reality for a long time, if ever. The field of translations where people and computers collaborate to improve the overall quality of professional translation is of particular interest. As Somers (2003:31) notes, ‘the idea is that the translator can consult a database of previous translations, usually on a sentence-by-sentence basis, looking for anything similar enough to the current sentence to be translated, and can then use the retrieved example as a model’ [4].

Depending on whether humans or machines do most of the translation, we refer to machine-assisted human translation (MAHT - translation assisted by text-processing software, terminology databases, or electronic dictionaries) or human-assisted machine translation (HAMT) (HAMT- translation memories, which come empty and initially have to be filled with translations from human translators, or MT systems that require extensive human pre- and post-editing). Both MAHT and HAMT are frequently used interchangeably and are generally coined computer-assisted translation (CAT) as they employ a plethora of tools to aid the translator perform his job swiftly and accurately.

3. MACHINE TRANSLATION

MT research has experienced some ups and downs throughout time. However, in the recent years it is undergone significant divergence of topics of interest. Machine translation (MT) fascinates people the most, the interest being highest particularly among non-translators. MT strives to compile all of the data required for translation into a single programme, allowing the translation of any text without the need for interaction from any human being. The public's opinion of MT oscillates between two extremes. Some believe that MT is completely worthless and a waste of time and money since the quality of output from an MT is often quite low, rendering it useless in practise.

A second view sees MT as a technique that will break down all language barriers, jeopardising translators' livelihoods. It makes use of the computer's calculation power to assess the structure of a text in the source language, divide it into pieces which are easy to translate, and then generate a text in the target language with the same structure. It has been posited that in only a few years, computer translations will be as excellent as human translations.

Reality, as is so frequently the case, lies somewhere in the middle, and the fact is that both ideas are incorrect. The entirely negative opinion shows an underestimation of MT's capabilities, whereas the entirely positive evaluation understates MT's limitations. The assertion that MT is ineffective in practise is definitely far from reality. The fact that multiple MT systems are used on a daily basis across the world refutes such arguments. The latest advancements highlight the fact that MT research and MT systems employ a large array of approaches to address the whole spectrum of linguistic phenomena. They include misspellings, grammatically unacceptable utterances, intricacy of constructions and terminology, to mention only a few. Other professional uses of MT comprise quality-improvement tactics such as pre- and post-editing, as well as dictionary update, which need a translator's human competence. In this regard, MT opens up new opportunities for translators as it makes use of the computer's calculation power to assess the structure of a statement or phrase in the source language, break it down into easily translatable pieces, and then generate a statement with the same structure in the destination language.

The terms 'fully automatic high-quality translation' (FAHQT), 'fully automatic machine translation' (FAMT), 'human-assisted machine translation' (HAMT),'machine-assisted human translation' (MAHT), and traditional ‘human translation’ (HT) describe the degree of automation or human involvement in the translation process.

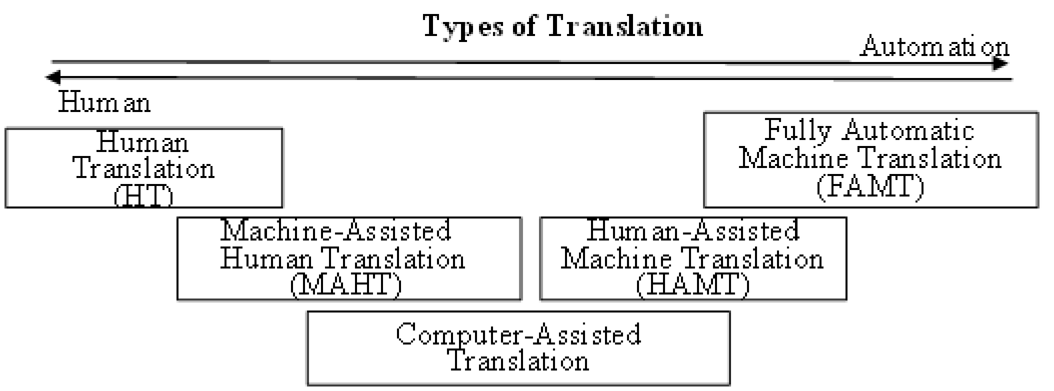

Fig. 1. Types of Translation. Human and Machine Translation

The radical notion of ‘fully automatic high-quality translation’ (FAHQT) was popular during the early stages of MT research, and it was predicated on the belief that MT systems could produce translations of human-like quality, but it was quickly abandoned. To accomplish high-quality human-assisted machine translation (HAMT), nearly all MT systems now in use rely on the aid of human operators. In this situation, the technology decodes and analyses the source text rather than the human operator, whose job is to help in the translation process. Human participation can occur before, during, or after the translation process. If teachers, scholars, scientists - who want to have their papers published in another language-, translators or other professionals are interested in the linguistic accuracy and the correct rendering of their publications, then they have to be aware that a significant amount of time should be allotted to the thorough and meticulous process of text preparation. Moreover, this undertaking in itself can rarely be completed in just one sitting and it generally entails several stages:

• Pre-editing, comprising the concept of getting the source text/input ready;

• Interactive mode, namely the interaction between the system and the human operator;

• Post-editing, i.e., correction and modification of the target text or output.

As seen above, it becomes obvious that the translations created by MT systems are neither intended for on-the-spot use nor are they proficient at generating instantaneously operational texts. If none of these options are practicable, the translation process will produce 'raw' (unrevised) translation output from systems with no limited or regulated input. 'Fully automated machine translation' (FAMT) is another term for this use of MT systems.

The employment of assistance such as electronic dictionaries or translation memory systems is referred to as ‘machine-assisted human translation’ (MAHT). Unlike HAMT and FAMT, the translator is responsible for decoding and analysing the original material due to the fact that languages are extremely dependent on context and the many words and word combinations which can be encountered in the text give rise to numerous denotations and connotations. Both HAMT and MAHT are frequently referred to as ‘computer-aided translation’ or ‘computer-assisted translation’ (CAT), as the computer provides a novel method to both source and target text processing. It is rather complex and it is supported by means of specific tools and technology which can be tailored to the translator’s needs.

The general organisation or abstract arrangement of an MT system's many processing units is known as the architecture. In time, there have been numerous approaches to MT, the most significant methods being essentially two: Statistics-based Machine Translation (SMT) and Rule-based Machine Translation (RBMT). The majority of RBMT systems in use today are based on one of three architectures: direct, transfer, or interlingua translation (cf. Hutchins 1995, Stein 2013, Poibeau 2017, Berndtsson 2015, etc.) [2] [5] [1] [6].

Chronologically, direct architecture is the initial method employed in most first-generation MT systems. 'Direct' in this instance denotes that there are no intermediary steps in the translation process, therefore the words of the source text are more or less instantly substituted by their target language counterparts. In other words, as Stein (2013:8) notes, the direct translation merely replaces ‘words on a word-by-word basis and only rely on parallel dictionary – so they neither do analysis nor transfer or generation’ [5]. This is accomplished using morphological data, bilingual dictionaries, and target language reordering rules, all of which are centred on basic parsing techniques.

The system does not perform a comprehensive analysis of the source language sentence since no complicated linguistic theories or parsing procedures are employed. The source sentence's analysis is reduced to the bare minimum necessary for accessing a bilingual dictionary, such as identifying parts of speech, plurality or singularity, tenses, therefore involving only a minimum of linguistic theory. This method is based on a dual relationship source language- target language which has been predefined. According to this approach, each word in the source language is unidirectionally connected to a similar unit in the target language.

The direct method is gradually becoming outdated. The trend in MT research is to construct transfer systems, which is why MT was centred on direct architecture until the mid-90s. A new generation of translation software has arisen that has all the features of a full-scale transfer system. However, as Stein (2013:8) mentions, 'regarding the complexity of these rules there are no limits and tens of thousands of rules, combinations and exceptions may be coded. But in practice there seems to exist a point where higher complexity does not indicate better results anymore. Instead, internal conflicts and contradicting rules produce arbitrary new errors' [5]. The transfer approach, in contrast to the direct architecture, centres on the theoretical concept of 'level of representation'. It makes use of three stages: analysis, transfer, and synthesis or generation.

The first stage, namely the assessment one, employs a source language dictionary while linguistically evaluating the source material. The source language sentence is morphologically analysed, utilising the source language vocabulary and grammatical rules. The end outcome is portrayed as an internal structure that is abstract. The abstract source structure is then lexically and structurally converted into an abstract target language structure in the next stage, namely the transfer stage. It is the place where all the conclusions and results identified in the analysis stage are transformed and the linguistic and structural counterparts between the two languages are determined. This intermediate stage is the only one encompassing bilingual rules. Using a target language dictionary and grammar, the abstract target language structure is converted into a target language surface structure during generation. This constitutes the last stage in the transfer strategy. The generation step uses a target language dictionary to create a document in the target language based on the linguistic data of the source language.

Nevertheless, individual grammatical or syntactic rules are not the only ones employed in transfer systems which extensively employ comprehensive linguistic conceptions and theories. As a result, they outperform direct MT systems in terms of translation quality. The analysis and generation components can be utilised for other language pairings if they are rigorously separated and do not involve bilingual rules, and if the representation is abstract enough. Multilingual transfer systems may be designed thanks to the separation of the various modules.

Interlingua architecture, unlike the transfer technique, has only two steps, since the transfer stage is skipped. The source text is analysed and transformed into an interlingual or language-independent representation, from which the destination text is created directly. This technique has the benefit of being able to use the interlingual representation for any language, eliminating the requirement for language-specific transfer modules. In itself, the Interlingua 'universal language' is somewhat based on the idea of a neutral language which could render all meaningful information in every language.

Nonetheless, since it is difficult to create entirely language-independent representations devoid of aspects that are reliant on the source or destination language, the transfer approach is frequently preferred over the interlingua method. Another reason is that the analysis and generation grammars are intricate as the presentations are so dissimilar to the characteristics of the source and target languages.

While dealing with translation memories (TM) in machine translation (MT), Berndtsson (2015:10) [6] analyses the workflow of a convertus syllabus translator (CST) as in *figure 2* below:

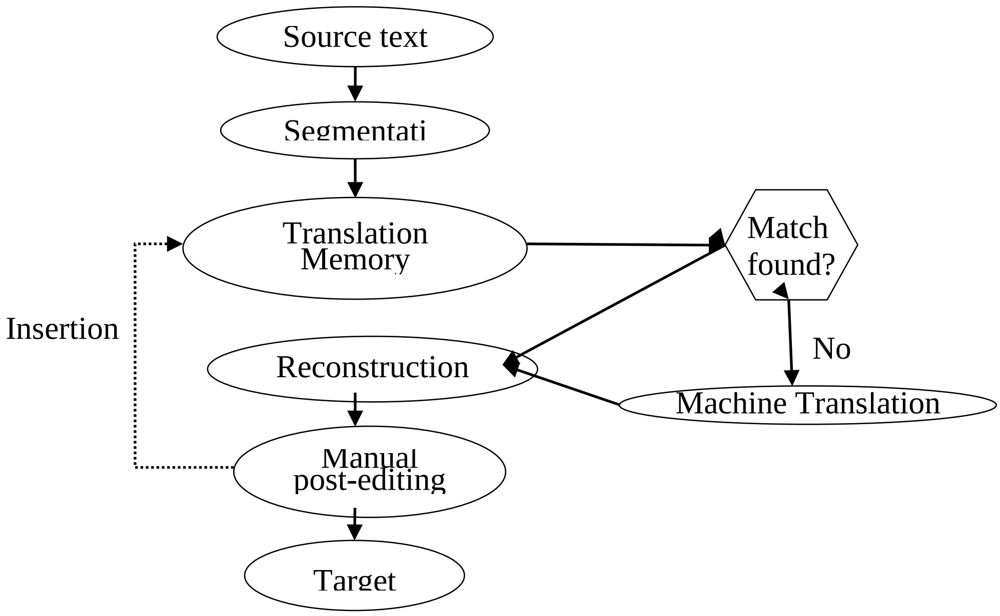

Fig. 2. Translator's work flow (adapted from Berndtsson 2015: 11) [6]

As seen in the figure above, the source text undergoes the process of segmentation, namely the text is investigated and then segmented into units corresponding to sentences, list items, etc. Within the process of segmentation, each segment is ascribed to the TM in order to look for a compatible match. When and if this process is achieved and a compatible translated segment exists in the TM, it is retrieved and sent 'directly to the reconstruction step, where all the translated segments are combined to form the full target text. If no TM match is found, the segment is translated by the MT system, and the target segment is automatically post-edited and sent to the reconstruction step' (Berndtsson 2015:20) [6]. These steps are then repeated for each and every segment, since the translation in itself has not reached the final stage and the target text has to undergo a technical human process carried out by human translators, namely the manual post-editing of the text. As Berndtsson (2015:10) further notes, 'after all corrections have been made and the translated text has been approved, all the source and target segment pairs are inserted into the TM, and the finished translated text is delivered to the user' [6].

On the other hand, Statistics-based Machine Translation (SMT) functions on different coordinates than RBMT, making use of substantial parallel corpora rather than employing the use of complex rule sets specific to RBMT. SMT implies considering all possible and impossible sentences in the target text as potential translations for the source text. For instance, in order to translate a statement from French, called 'f' for convenience' sake, into an English statement, dubbed 'e', all feasible and unfeasible English sentences 'e' may be perceived as prospective renderings of 'f'. Not all combinations would render the output as acceptable, nevertheless, the principle at core being that certain translations are more likely to be accepted as the end product than others. When analysing this possibility, it can be reduced to p(e|f), which can be rendered as the probability that 'e' is the proper translation of 'f'. On investigating the way in which SMT operates, Poibeau (2017:76) [1] as well as Stein (2013:9) [5] point out the fact that 'the concrete probabilities used by the computer are estimated with Bayes' Theorem:

$$Pr(e/f) = \frac{Pr(e)Pr(f|e)}{Pr(f)} \quad (1)$$

Therefore, the purpose of SMT can be described as focusing on the pattern of identifying the original sentence 'e' which is perhaps the most likely and probable translation. There is one flaw, however, in the previous statement, namely the inability of retrieving all sentences of a target language and an immediate solution had to be identified. The answer at hand was for the SMT to no longer try to access all sentences but instead to operate employing approximations or translation models. This process entails the concept of regarding every word as a possible translation of all the other words, but the actual probability of achieving the envisaged result lies with the words they are aligned to in the predefined bilingual corpus.

Bayes' Theorem as applied in 1) above can be further re-examined and, as Stein (2013:9) notes, the 'sentence can be reduced to the search of the maximum value of the terms Pr(e) ("Probability that e has been said by someone") and Pr(f|e) ("Probability that someone would translate e to f")' [5].

$$\hat{e} = argmax_e\, [Pr(e) * Pr(f\,|e)] \quad (2)$$

The language model comprises a supplementary corpus, a monolingual one in this instance, which purportedly encompasses the entirety not only of words but also word collocations and associations capable of producing grammatically and semantically acceptable sentences. The 'highest product of the values sentence validity (language model), word translation and word order (translation model)' (cf. Poibeau 2017:77) [1] are therefore established by the search algorithm which identifies the sentence as the SMT output.

4. TECHNIQUES FOR IMPROVING THE OUTPUT QUALITY OF MT

As Kremer et al. (2017:147) suggest, 'machine-supported human translation is an open field with ample potential for creative strategies to combine the complementary strengths of man and machine' [7]. To evaluate the effectiveness of MT systems, we must first admit that today's technology cannot attain the ultimate fully-automated high quality translation (FAHQT). Nonetheless, the user is given options for increasing the MT output quality wise. Pre-editing (including controlled language), updating the system's dictionaries, post-editing, and human-machine interaction are all ways for improving quality (interactive mode). These measures are not mutually exclusive and can be used in conjunction with one another. Pre-editing refers to the input text, whereas post-editing pertains to the output texts, respectively, whereas the additional methods have an impact on the program's operation.

The dictionary component of an MT system can be customised to meet the demands of the user. In turn, it entails making changes to current entries or adding phraseological terms to the user's domain. These newly introduced phrases can be presented with morphological, semantic, and phraseological information if the dictionary's entry structure is sufficiently complex. Linguistic knowledge may be included into the algorithm, potentially improving translation quality. As Habash et al. (2011:133) note, 'the most important resource in the SMT approach is the corpus of paired source and target texts or parallel corpus. An initial step before a parallel text can be used involves cleaning it and pre-processing it to a representation that allow us to learn from it optimally' [8]. When analysing an MT system, one should always consider the complexity of the dictionary component. However, it is important to remember that the complexity of the entry structure is proportional to the expectations placed on the user: the more complicated the structure, the higher the demands.

The extent of how MT-friendly the input is dictates the quality of the output of MT. Thus, pre-editing entails identifying possible problems and therefore arranging a source text or an input to avoid issues from the start. As a result, the MT system may be affected and sometimes incapacitated by word omissions, i.e. ellipsis, idiomatic expressions, as well as structures which are syntactically rendered as too complex. The main idea, therefore, is to highlight and, whenever possible, eliminate or change beforehand any string of text which could pose difficulties to an MT system. Another factor to consider while pre-editing is the usage of simple and direct language to avoid semantic ambiguity. The combined process is supposed to produce better results in terms of both readability and translatability of the source text (ST). If there are, however, contexts where translatability and readability are not congruent with each other and do not entail synonymous relationships, ‘translatability will be given priority’ as far as MT is concerned (cf. Reuther 2003:129) [9].

According to its core definition, a Controlled Language (CL) entails a plethora of restrictions (both grammatical and lexical) whose sole purpose is to disambiguate the text when there is sufficient vacillation as to the intended meaning in the SL. A text exhibiting high levels of complexity is deemed undesirable from a MT perspective and in direct contrast to CL. Controlled language may be a source of frustration for technical writers, who believe it limits their originality. One of the principles of controlled language is that each word has only one meaning, that each word belongs to only one word class, and that complicated syntactic structures such as conditional clauses should be avoided. Predictably, source materials written in controlled languages frequently outperform those produced in uncontrolled languages when MT is used. Even if the source content was written in a controlled language and thus entailing pre-editing, the target text must be post-edited to ensure high-quality translation.

'Special languages,' which are employed in certain technological disciplines, are a more general manner of constrained input. In technical manuals, for example, instructional forms predominate. These kinds of languages share several attributes such as lexical items which are clearly specified and, at the same time, some grammatical constructions are employed to a greater extent in comparison to others. This means that, with specific syntax, semantics, and pragmatics, the MT system may be tailored to those structures.

Another stage in the overall translation process is post-editing. As the name itself implies, post-editing occurs after the translation has been carried out by the machine. It is a process which edits, changes, modifies, alters and at times even corrects the target text, or the raw output produced by the MT system. ‘This task poses specific problems as compared to purely human translations’, according to Čulo (2013:35), ‘as the post editors have to deal with output that can be erroneous on multiple levels: morphology, syntax, semantics and last but not least pragmatics. Also, the cognitive load is heightened with respect to focus: when postediting, translators have to focus on both the source text as well as the MT output’ [10]. The amount to which post-editing is performed and to which extent the target text undergoes ‘polishing’ is primarily determined by the quality demanded by the user. It can turn into a fairly delicate process entailing significant cognitive effort, being time consuming and inefficient especially when no expert assistance is employed. As Allen (2003:298) notes, ‘few benchmark tests have been conducted to estimate the productivity gain or loss of the post-editing process in comparison with the human translation process’ [11]. Pre- and post-editing activities should be delegated to professional translators who are not only conscious of MT limitations, but they are also aware of the cross-language transfer of concepts while mastering the technical resources necessary when context dictates.

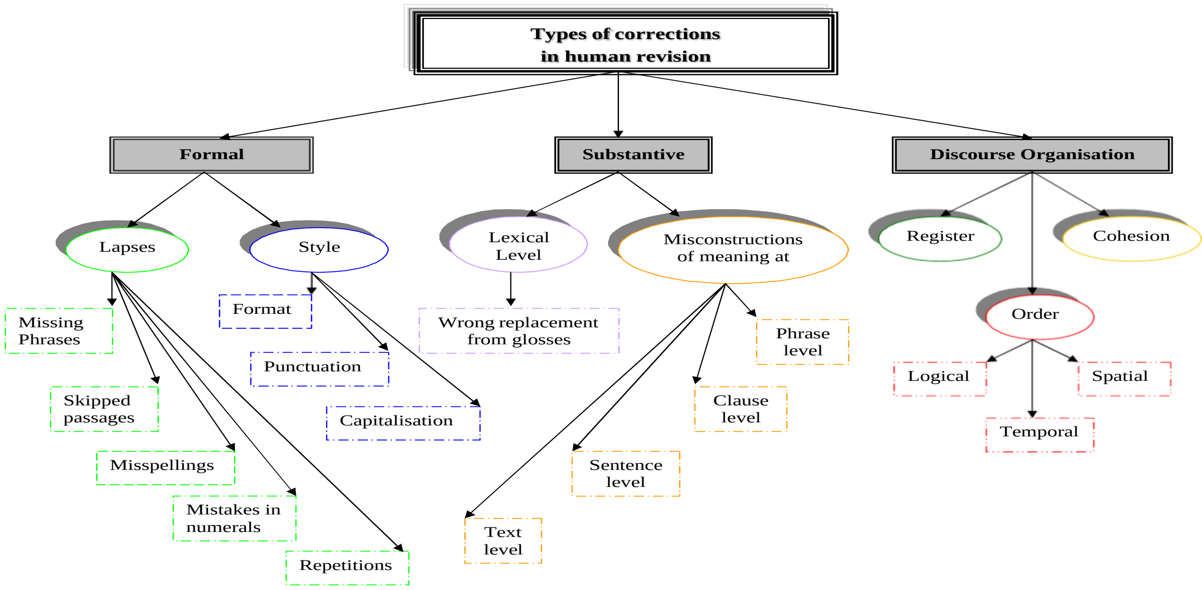

Fig. 3. Types of corrections in human revision (adapted from Martinez (2003:26)) [12]

5. ADDRESSING THE MOST COMMON ISSUES IN MT

Word order may at times pose problems in terms of Machine Translation. Syntactically, the order of elements within the sentence can be problematic especially when the object or subject of the sentence does not assume the standardised position and thus generates several readability challenges. Passive voice might also sparingly raise numerous issues when considering general stylistic recommendations, therefore the universal norm would be to keep to standard, formal English in which grammatical relationships are conveyed clearly. Employing passive constructions may contribute to a style of writing promoting ambiguity and vagueness. Active voice, on the other hand, eliminates confusion. If ambiguity is not removed, the machine translation engine will struggle especially when there is duality of meaning within the sentence.

Individual words are subject to lexical ambiguity, which happens when a single word can have more than one meaning. Polysemy and homography are key concepts in this setting. As they belong to separate grammatical groups, homographs are frequently mistranslated. This is not an issue for the MT system in brief syntactic contexts, but it can become one if the context is complicated or incomplete. Specific approaches such as spell-checkers are not always the solution since the words are not misspelled as such, but rather miswritten in the given context. Other contexts where they would be perfectly acceptable can be easily constructed. However, as Somers (2003:96) notes, 'for a spell-checker to correct this type of error would require sophisticated computational linguistics software that would analyse and, in some sense, "understand" the text it was checking' [13]. The disambiguation of homographs, on the other hand, is usually easier to achieve than that of polysemes. The latter refers to two or more words that are classified as belonging to the same grammatical category yet having diverse meanings. When the translation of a word varies depending on its meaning, polysemes become important in MT.

Unlike lexical ambiguity, structural ambiguity concerns sentence syntactic structures and representations. Confusing constructions have a detrimental impact on the clarity of the text and the output, and, when this happens, the reading and translation processes are rather complex. This problem is significant in any CL use case. Ambiguities may arise from the fact that phrases, usually prepositional ones, can be assigned more than one location in a sentence.

Another noteworthy aspect involves pronoun usage. Since they are referential, ellipsis of pronouns when preceding verbs, or, for that matter, omission of relative pronouns leads to the system generating erroneous structures, affecting not only the translatability of the text but also posing a challenge in terms of the comprehensibility of the output. This principle applies for all elliptical constructions. As Jurafsky and Martin (2021) as well as Sejnowski (2018) note, when the machine encounters such situations, the mechanisms assigned for parsing processes attempt to recreate the missing components. Whereas human analysis may solve these kinds of issues substantially successfully, machine translation achieves solutions resulting in failed parses in no insignificant situations [14], [15].

Avoid splitting separable English verbs since idioms, idiomatic phrases and colloquial constructions are difficult to manage in an MT system because their meaning cannot be fully comprehended from the individual meanings of their components. Idioms must be treated as single units of translation in MT systems, which is generally accomplished by adding them in the system's dictionaries. As Lopez (2015:25) points out, 'if the idiom is in the bilingual dictionary of the system, the machine translation will be better than the translation by a professional who, for not having understood the context, translated it literally or too freely' [16]. Idioms, fortunately, are uncommon in special language texts, which constitute the vast bulk of texts submitted to MT. However, it is safer and less time consuming to avoid using idioms entirely, which is why this is one of the key principles in the pre-editing stage.

During MT analysis, complex syntactic structures are a typical cause of error. Complexity causes issues not only for human but also for machine analysis, affecting the source text's readability and translatability. If the system is unable to fully analyse the construction and assign the appropriate grammatical categories, it must resort to the robust mode, which involves translating word-by-word and duplicating the structures of the source sentence (cf. Jurafsky and Martin 2021 [17]). The more complicated a structure, the less probable a precise match will be discovered in the translation memory of the system. The human translator, on the other hand, should encounter less challenges in the process, provided they have grasped the meaning of the text, according to Munday (2016) and Ionescu [18] [19]. Although advanced technology can handle extensive structures, sentence construction should thus be kept clear, straightforward and explicit.

6. CONCLUSION

The MT system is essentially still short of perfection and there are heated debates amongst researchers, scientists, and translators who at times are apparently rather dissatisfied with the output. Nevertheless, while they are still deliberating, an increasing number of consumers find the flaws, if not palatable, at least acceptable, and exhibit no frowning on welcoming the technology.

The construction of a functioning MT system is unavoidably a long-term 'engineering' endeavour requiring the use of well-known, dependable, and time-tested methodologies. The best concluding remark, however, can be drawn along the following lines: ‘while academics debate linguistic and statistical approaches to MT, organisations in the public and private sector are putting it to work […]. The breakthrough is market-driven rather than technical: MT is not perfect, but it has become an economic necessity. We must learn how to use it and how to optimise its benefits in practical environments’ (Van der Meer 2003) [20].

REFERENCES

[1] T. Poibeau, *Machine Translation*, Massachusetts: The MIT Press, 2017.

[2] W. J. Hutchins, *Machine Translation: A Brief History*, Concise history of the language sciences: from the Sumerians to the cognitivists, E. F. K. Koerner and R. E. Asher (eds.), Oxford, Pergamon Press, pp. 431-445,1995.

[3] O. Czulo and S. Hansen-Schirra, *Introduction* in *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation*: *TC3 II* (Translation and Multilingual Natural Language Processing 4), O. Czulo and S. Hansen-Schirra (eds.), Berlin, Language Science Press, 2017.

[4] H. Somers, *Translation technologies and minority languages* in *Computers and Translation: A translator's guide*, vol. 35, H. Somers (ed.), Amsterdam, John Benjamins Publishing Company, pp. 87-104, 2003.

[5] D. Stein, *Machine Translation. Past, Present and Future* in *Translation: Computation, Corpora, Cognition. Special Issue on Language Technologies for a Multilingual Europe*. G. Rehm, F. Sasaki, D. Stein, and A. Witt (eds.), vol. 3, no. 1. 2013.

[6] J. Berndtsson, *Variable Insertion and Search in a Translation Memory*, Uppsala University, 2015.

[7] G. Kremer, M. Hartung, S. Padó and S. Riezler, *Statistical machine translation support improves human adjective translation* in *Crossroads between Contrastive Linguistics, Translation Studies and Machine Translation*: *TC3 II* (Translation and Multilingual Natural Language Processing 4). O. Czulo and S. Hansen-Schirra (eds.), Berlin, Language Science Press, pp. 121-152, 2017.

[8] N. Habash, J. Olive, C. Christianson and J. McCary, *Machine Translation from Text* in *Handbook of Natural Language Processing and Machine Translation*, J. Olive, C. Christianson, J. McCary (Eds.), New York, Springer, pp. 133-398, 2011.

[9] U. Reuther, *Two in One- Can it Work? Readability and Translatability by means of Controlled Language*, in *European Association for Machine Translation*, pp.124-132, 2003.

[10] O. Čulo, *From Translation Machine Theory to Machine Translation Theory – some initial considerations* in *The Future of Information Sciences. Information Governance*, A. Gilliland, S. McKemmish, H. Stančić, S. Seljan, J. Lasić-Lazić (eds.), Zagreb, Intering, pp. 31-38, 2013.

[11] A Jeffrey, *Post-editing* in *Computers and Translation: A translator's guide*, vol. 35, H. Somers (ed.), Amsterdam, John Benjamins Publishing Company, pp. 297 -318, 2003.

[12] L. G. Martínez, *Human Translation versus Machine Translation and Full Post-Editing of Raw Machine Translation Output*, Dublin City University, 2003.

[13] H. Somers, *Translation memory systems* in *Computers and Translation: A translator's guide*, vol. 35, H. Somers (ed.), Amsterdam, John Benjamins Publishing Company, pp. 31-48. 2003.

[14] D. Jurafsky and J. H. Martin, *Machine Translation and Encoder-Decoder Models* in *Speech and Language Processing* (3rd ed. draft), 2021.

[15] T. J. Sejnowski, *The Deep Learning Revolution*, Massachusetts, The MIT Press, 2018.

[16] J. M. Lopez, *Machine Translationness: A Concept for Machine Translation Evaluation and Detection*, 2015.

[17] D. Jurafskyand J. H. Martin, *Transfer Learning with Pre-trained Language Models and Contextual Embeddings* in *Speech and Language Processing* (3rd ed. draft), 2021.

[18] J. Munday, *Introducing Translation Studies. Theories and applications*, New York, Routledge. 2016.

[19] D. Ionescu, *Translation Strategies in the Approach to Scientific Texts* in *The Art and Craft of Translation*, L. Frățilă (ed.), Timișoara, Editura Universității de Vest, pp. 113-126, 2009.

[20] J. Van der Meer, *The Business Case for Machine Translation - The Breakthrough Is for Real*, The Lisa Newsletter, vol. 12, 2.6, 2003.