Machine translation today

Mikel L. Forcada
Dublin City University

Machine translation (MT) is the translation, by means of a computer using suitable software, of a text written in the source language (SL) which produces another text in the target language (TL) which may be called its raw translation. This definition seems to imply that the resulting TL translation may be used as a professional product would, but machine translation and professional translation, even if closely related in purpose, are not interchangeable products (Sager 1994: 261).

Machine translation should be clearly distinguished from other translation technologies such as computer-aided translation*. In MT, translation is performed by the computer, with no human intervention in the process (although, as will be seen, there may be need for human intervention before or after it); in computer-aided translation, translation is performed by a professional with the aid of a range of translation tools* to help them.

1. Challenges and limitations of MT

Raw translations produced by an MT system are usually very different to those produced by translation professionals. This does not mean that MT is not useful in many areas; it only means that one has to be aware that it has some specific applications. What one needs is to identify the contexts in which one can use MT effectively and to know what can be expected of it.

Arnold (2003) classifies the obstacles faced by machine translation in four groups:

1. Form does not completely determine content. A text can have different interpretations. This refers, therefore, to the ambiguity of language. Sentences in a text can be ambiguous:
   - because one or more of its words have more than an interpretation (lexical ambiguity),
   - because the sentence has more than one possible syntactic structure (structural or syntactic ambiguity), or,
   - in some cases, because of both things at the same time.

Here are some examples:
   - If we read “I saw John walk by the bank”, it could be the case that we speak of a financial institution or of one of the edges of the river (lexical ambiguity of the word bank).
If we read “I saw the girl with the telescope”, it is not clear whether I used the telescope to see the girl or the girl had a telescope when I saw her (structural ambiguity: we do not know if the prepositional phrase “with the telescope” attaches to the noun phrase “the girl”, i.e., she had the telescope, or to the verb phrase “saw the girl”, where I have the telescope).

In most cases, ambiguity is a problem because the MT system has to choose the correct interpretation of a sentence in order to produce a suitable translation. Automatic resolution of ambiguity is far from being an easy task. While people can use context and their knowledge, their expectations, and their beliefs about the world to safely discard many interpretations (ideally, all but one), MT systems have to make these decisions using only programmable and computable processes which take a reasonable time and use a reasonable amount of memory to process the (often incomplete) information that may be extracted from the text surrounding the ambiguous element.

2. Content does not completely determine form. There exist many ways to express in a language a given content. Following Arnold’s (2003) example, only the last two of the following expressions genuinely express in English that we want to know the time of the day: “How late is it?” (German “Wie spät ist es?”), “Which hours are they?” (Portuguese “Que horas são?”), “How many o’clock is it?” (German “Wieviel Uhr ist es?”), “What time is it?”, “What’s the time?”.

3. Different languages use different structures to convey the same interpretation. Let’s consider the sentence “I like bicycles”. Its Catalan translation is “M’agraden les bicicletes” where one can see that the direct object in the English sentence (“bicycles”) has been turned into the subject of the sentence in Catalan (“les bicicletes”) which requires a 3rd person plural verb (“agraden”) and that the English subject pronoun (“I”) has turned into a proclitic object pronoun (“M’”). Although this is a relatively simple example, in general, the structures used by different languages can be so divergent that a simple, word for word translation would be unintelligible or even wrong.

4. The above three problems may be said to be the manifestation of intrinsic features of translation; there is also what Arnold (2003) calls the description problem. Current translation theories cannot formally express, either all the mechanisms underlying natural language translation or the mental processes involved. The intrinsic problems described above are tackled using methods that, in general, have to make radical simplifications (or even complete reformulations) of the professional translation process.

On the one hand, these problems are noticeably reduced when the languages involved in the translation are related: morphological, syntactic and semantic affinities simplify the design of these systems and allow one to easily obtain translations which are both easy to understand and correct. On the other hand, there are text types that can be
more easily translated, such as commercial letters, technical texts or economic texts, and others for which machine translation may be completely inapplicable such as advertising texts or poetry.

2. Applications of machine translation

Machine translation has basically found two main purposes, which are conventionally called assimilation and dissemination.

2.1 Assimilation

In assimilation, texts are machine-translated when one does not understand the SL and wants to have an approximate idea of the content of the text, its gist. One example might be browsing Internet pages through a machine translation system that translates them instantly to the chosen TL. In this application, translation errors are not too important if the system manages to convey the general sense of the text (the level of detail achieved depends on the machine translation system and the languages involved, but also on the user, who can indeed learn to use this new type of text quite profitably).

2.2 Dissemination

In dissemination, texts are machine-translated as an intermediate step in the production of a document in the TL that will be published (disseminated); raw MT results have to be post-edited (Allen 2003; Hutchins & Somers 1992: 152), that is, revised and corrected by a skilled professional. Indeed, replacing a human-translation-only environment by one that utilizes MT as a key component may bring about savings, and may be done in different stages, depending on the particular situation at hand:

1. Machine translation followed by post-editing: professional post-editors (ideally specially trained translators) edit the raw MT output into adequate text. This may apparently be advantageous if the joint cost of post-editing and MT is lower than the cost of human translation, but one should also consider additional costs arising from training and any required changes in the translation workflow.

2. Pre-editing: When several TLs are involved, it may be advantageous to pre-edit (Hutchins & Somers 1992: 151) the source text: each well-chosen edit to the source text may avoid several edits in more than one language (note that, in general, post-editing cannot be avoided altogether). However, pre-editors have to be trained to anticipate MT problems and this adds extra costs.

3. Controlled languages: Finally, repetitive pre-editing may be avoided by defining a controlled language (Nyberg et al. 2003; Arnold et al. 1993: 147; Hutchins & Somers
1992: 151; Lockwood 2000) to be used by authors, a variant of the human SL with lexical and syntactic restrictions designed to avoid problems in MT. Designing or adapting a controlled language to the task at hand is costly, and tools have to be provided for effective authoring. Therefore, it will only be profitable if heavy repetitive pre-editing would otherwise occur.

3. Approaches to machine translation

3.1 Two main approaches

At present there are two main approaches to machine translation. Until the nineties the dominant approach was rule-based or knowledge-based machine translation: teams comprising computer and translation experts programmed the morphological analysers, the parsers, etc. in the MT engine and compiled dictionaries and grammatical rules to transform sentence structures, etc. in formats that could be processed by that MT engine. However, since the beginning of the nineties we have witnessed a growth of what can be called corpus-based machine translation: MT programs “learn to translate” from enormous corpora of bilingual texts where millions of sentences in one language have been aligned with their counterparts in the other language (these corpora are not unlike huge translation memories: see Computer-aided translation). There is also room for hybrid approaches (a very active field of current research) such as statistical MT systems (see 4.3.2) incorporating some type of linguistic knowledge (such as morphological dictionaries: Koehn 2009: 314) or statistical MT techniques for domain-targeted “automatic post-editing” of rule-based MT output (Simard et al. 2007).

Rule-based systems take longer to build (it is necessary to encode explicitly the linguistic information that the system will use) whereas corpus-based systems can be constructed more quickly but only provided that a large volume of sentence-aligned bilingual text is already available. Therefore, corpus-based systems are difficult to apply, for instance, to minority or less-resourced languages without extensive bilingual corpora.

3.2 Rule-based machine translation

Among rule-based MT systems, the most usual are transfer systems. An ideal transfer system (Hutchins & Somers 1992: 75) has three well-defined stages:

1. **Analysis** produces, from the sentence in SL, an abstract intermediate representation, in which linguistic classifications and groupings are established to allow for the application of general rules of translation. For example, in English – Spanish translation, if the analysis indicates to us that the English segment “a comfortable cushion” consists of a determiner, an adjective and a noun, it is possible to later apply a general rule that reorders this sequence into a determiner – noun – adjective sequence.
2. *Transfer* converts the intermediate representation delivered by analysis into a new intermediate representation for the TL, looking words up in the bilingual dictionary (*lexical transfer*) and for instance, applying rules such as the one just mentioned (*structural transfer*).

3. *Generation* produces a concrete TL sentence from this abstract intermediate representation.

The fact that only the transfer stage is bilingual induces some *modularity* in the system: analysis for language pair $A - B$ may be used, say, for language pair $A - C$, and generation for language pair $A - B$ may be used, say, for language pair $D - B$.

The intermediate representations can be more or less complex: analysis may indeed be deeper, leading to a complete syntactic parsing of the sentence, or even to a subsequent semantic representation of it.

Indeed, analysis may be so deep that a language-neutral intermediate representation needing no transfer is obtained: only analysis towards it and generation from it are necessary. Such systems are called *interlingua* systems (for instance, the Kant system: Mitamura et al. 1993). They have the advantage that no bilingual knowledge is needed to add a new language to an existing system, and that just two modules (analysis and generation of the new language) have to be added, but designing a general-purpose interlingua is tantamount to designing a complete model of the real world (and even hypothetical worlds), which restricts interlingua systems to limited-domain translation tasks.

### 3.3 Corpus-based machine translation

Corpus-based machine translation, also called *data-driven* machine translation, may be divided into two main paradigms: *example-based machine translation* and *statistical machine translation*. A corpus of sentence-aligned bilingual parallel text is a prerequisite for both approaches.

#### 3.3.1 Example-based machine translation

Example-based machine translation (EBMT, Carl & Way 2003) was first formulated by Nagao (1984) as “translation by analogy” and is generally described as consisting of three distinct phases:

- **Matching**: the new sentence to be translated is segmented and the segments are matched against identical or similar segments in the SL side of the bilingual examples in the corpus.
- **Alignment**: the corresponding fragments in the TL side of the matched bilingual examples are determined, to build “translation units”.
- **Recombination**: the TL sides of these “translation units” are combined into a translation for the new sentence.
These three phases are usually seen as being parallel (Somers 2003b) to the analysis, transfer, and generation phases in the transfer approach of rule-based machine translation (see Section 4.2).

An important property of EBMT is that, when the new sentence is identical to a sentence in the corpus of examples, its translation is recovered and used, as it would in a translation memory system (see Computer-aided translation*); this is not the case with statistical machine translation (Section 4.3.2). EBMT systems differ with respect to the level of pre-processing of the corpus of bilingual examples before the translation process per se, which may involve the use of linguistic resources such as parsers or bilingual dictionaries. Each one of the three phases faces specific challenges which are still an open subject of research: the nature of segments and the segmentation process itself is crucial for successful matching, alignment is usually far from being trivial, especially when languages are not closely related, and successful recombination may be hampered by issues at segment boundaries (lack of agreement, repeated words, etc.).

3.3.2 Statistical machine translation

Statistical machine translation (SMT, Koehn 2009) developed independently of EBMT, as a result of the seminal paper of Brown et al. (1988), is currently the dominant paradigm in MT research and has a growing share of the MT market. In SMT, one says that a SL and a TL sentence are a translation of each other with a certain probability. Indeed, in principle, any TL sentence can be the translation of the SL sentence at hand, and the task becomes reduced to finding the TL sentence for which this probability is the highest possible. The approach assumes that a reasonable estimate of such a probability may be computed using a probability model inferred from the bilingual corpus. Additionally, the search is only approximate since the system cannot explore all possible translations of the sentence at hand.

The probability model is usually made of several components such as a translation model consisting itself of lexical probabilities (probabilities that a certain TL word and a certain SL word are mutual translations) and alignment probabilities (describing processes such as word reordering), a TL probability model describing how “natural” (how likely) a TL sentence is (independently of the SL sentence), as well as various other probability models or “features” of the SL and TL sentences. Note that no linguistic information is used in “pure” SMT: the two models are inferred or learned by using complex statistical estimation techniques on (usually very large) sentence-aligned bilingual corpora (the “training corpus”). Also, state-of-the-art SMT systems use segments longer than words, called phrases (Koehn 2009: 127) even if they are not syntactic units in the linguistic sense.

4. Evaluation of machine translation

MT depends strongly on the quality of the raw machine-translated text. However, defining MT quality in general terms has proven to be very difficult, and indeed, the adequacy of raw output may vary from one purpose to another purpose. For instance, a raw machine-translated text may be almost perfectly understandable to a native speaker of the TL, but may still need heavy post-editing to make it fit for publishing. And, conversely, MT errors that make a substantial part of the raw text unintelligible for that native speaker may be very easy to spot and correct by a skilled post-editor.

4.1 Manual evaluation

Traditional manual evaluation measures use human judges to score sentences according to two intendedly independent criteria: on the one hand, the intelligibility or fluency of the translation, independently of the original text, and, on the other hand, its fidelity or adequacy, that is, how much of the meaning of the original sentence is conveyed by the raw translation. Multiple judges are ideally used and their scores are averaged out to get a more stable indicator.

4.2 Automatic evaluation

Manual evaluation is clearly very expensive and cannot be performed repeatedly, for instance, when comparing the output of different MT systems or to adjust parameters during MT system development. As a result, a number of automatic evaluation measures have been proposed, which try to measure how close each raw machine-translated sentence is to one or more reference human translations. These measures may be very efficiently computed whenever they are needed, and can indeed be repeatedly used during the development of SMT systems. The most commonly used measure or “metric” is called BLEU (“Bi-Lingual Evaluation Understudy”, Koehn 2009: 226), and simply measures how many segments of one, two, three or four words the raw translation has in common with the references.

4.3 A critique

One problem that has been observed is that automatic evaluation measures do not always correlate with human evaluation results (Koehn 2009: 231). However, another important problem with both manual and automatic measures is that they are very indirect and tend to evaluate the quality of machine translation by comparing it with professional translation, even if they are not directly interchangeable in any real-world application (Sager 1994: 161). Indeed, they measure quality in ways that may not directly relate with real-life applications of machine translation. Consider MT output that is going to be post-edited for publication: subjective assessments of adequacy may not clearly correlate with post-editing effort, and fluency may not correlate at all (non-fluent translations with apparent, easy-to-correct errors may be preferred sometimes by post-editors). And objective assessments of how close the raw output is to one or more references may not give an idea of how much effort is needed to correct it to produce another adequate translation not in the reference set.
5. Conclusion

In view of the discussed challenges and limitations, it is quite clear that MT will never take the place of professional translators. On the contrary, in certain situations that should be clearly evaluated, one can expect a good MT system to free translators from the most mechanical part of the translation task, so that their productivity increases to match the increasing demand, but one should never expect the MT system – however good – to understand the text, to always solve ambiguities properly and to produce texts conforming to the TL norms or fit for the intended purpose of the translation. Research and development will keep on producing improved systems that will be available as a component to be integrated, where appropriate, in new translation workflows, to successfully address the growing demand for translation.

References


