



## MT+TM+QA: The Future is Ours

Alan K. Melby  
Brigham Young University at Provo

### Introduction

What will be the future of MT (machine translation)? How will it affect human translators? During the 1950s and early 1960s, we heard that MT would soon replace human translation, but it did not. The ALPAC report (1966) put a damper on research in machine translation world-wide for a number of years, but in the early 1980s, people were again suggesting that machines would soon replace human translators. The 1990s were relatively calm, with modest claims by the promoters of MT systems. Then, more recently, in the early years of the 21st century, developers of statistical machine-translation systems are enthusiastically announcing, yet again, that the quality of raw MT output will soon meet or exceed the quality of human translation. Is this just another false alarm, or is something fundamentally different this time? In this article, I will make claims about the future of MT systems, the future of translation memory (TM) systems, and the role of quality assurance (QA) in the future of human translators.

### 1. Machine Translation (MT)

*My first prediction is that traditional, hand-coded, ruled-based machine-translation systems will receive less attention in the next few years.*

Such systems consist of three phases of processing: analysis of the source text, transfer (to accommodate differences between the source and target languages), and generation of the target text from an intermediate representation. They require enormous amounts of human time to develop the rules, and the quality of the raw output is low unless the system has been tailored to a very narrow domain and the source text conforms to this domain.

These systems are built on the following assumption about the nature of language: that meaning can be computed from the bottom up (that is, starting from isolated, individual words, and combining them into larger and larger units). At first glance, this assumption seems obviously true. How else could we figure out the meaning of a sentence other than by combining the meanings of individual words? However, in fact, humans do not deal with words in isolation when analyzing a text. Context is continuously being taken into account, even if we are not consciously aware of it. In most rule-based machine-translation systems, context is only brought in during word-sense disambiguation. Once the sense of a word is identified, it is assumed that the word-sense can thereafter be

treated without further reference to context.

In terms of linguistic theory, rule-based systems are typically associated with some branch of generative grammar, if there is even a full syntactic analysis at all. There is an extensive literature describing rule-based systems and the linguistic models on which they are based. See, for example, the references in Hutchins (1986). Of course, the study of syntax did not originate in the 1960s when the generative approach began to dominate the linguistic scene. As Chomsky himself points out, his approach to syntax is not entirely original and shares much with Cartesian philosophy. (Chomsky 1966).

In contrast with rule-based systems, which can now be called classic MT systems, there is a substantially different approach called statistical machine translation (SMT). While rule-based MT systems can be viewed as being based on grammars and dictionaries, which have been around for thousands of years, statistical machine translation systems, on the other hand, are based on bilingual corpora. After initial experiments in the 1990s (Brown et al 1990) and then a period of little activity in SMT, there has recently been a flurry of activity in this area. In statistical machine translation, the starting point is an extensive collection of pairs of documents. Each pair, often called a bitext, is a source text and a target text. The target text is normally a human translation of the source text. Each bitext is segmented, usually at the sentence or paragraph level, and corresponding segments in the source text and target texts are linked. Then the source text is fully indexed for rapid retrieval of segments containing a particular word or phrase, along with the corresponding segments of target text, which presumably contain the translation of the word or phrase in question. In addition, an extensive statistical analysis of the corpus of bitexts results in a table of correspondences between source language words or phrases and target language words or phrases. In a sense, this table can be thought of as a bilingual dictionary that has been automatically derived from the bitext corpus. However, this does not mean that a statistical machine-translation system is equivalent to a rule-based system. Perhaps the most significant difference, other than the obvious difference of whether the bilingual dictionary is created manually or automatically, is that the machine translation in an SMT system is not just a one-to-one mapping of source-language words to target-language words. An SMT system is not just a glorified word-for-word dictionary lookup and substitution procedure. Instead, context is taken into account by matching chunks of source text with chunks of target text whenever possible. This matching is not done by applying a linguistic model of language but rather by using statistical methods that have proven very effective in automatic speech recognition. In a purist approach to SMT, there is a degree of disregard for the classic linguistic levels (morphology, syntax, and semantics). It would seem logical that morphological processing will eventually be needed in SMT, especially for highly inflected languages, in order to map between base forms of words instead of treating each inflected form separately. For example, the word "shoe" should probably correspond to the same base form in another language regardless of whether that word is inflected one way as the subject of a sentence and another way as the direct object. Also, some differences in word order that involve long distances dependences, such as the placement of the verb at the end of dependent clauses, will best be expressed with some sort of syntactic representation.

The future of statistical machine-translation systems is probably a hybrid approach in which morphology and syntax are somehow taken into account. This may involve using some explicitly rule-based components, such as a

morphological analyzer, or it may involve alternative approaches to morphology, such as the use of analogical models of language (AML). In AML, many exemplars, such as specific inflected forms each paired with the appropriate base form, along with features and a distance metric, are the input to the system, rather than hand-coded rules. These hybrid systems will probably still be focused on bilingual corpora rather than traditional rules, and thus we will call them "data-driven" systems as opposed to rule-based systems.

*My **second** prediction is that whenever sufficient quantities of high-quality bilingual corpora are available for the domain being treated, data-driven machine-translation systems will soon outperform classic rule-based systems in quality of output, though probably not in speed.*

As computing power becomes even less expensive, the speed difference between rule-based and data-driven systems will, of course, become a less important factor. However, processor speed cannot make up for a lack of a sufficiently large and suitable bitext corpus.

## **2. Translation Memory (TM)**

Having made some testable predictions concerning the future of MT, let us turn to the second part of the title of this article: TM (Translation Memory). Traditional TM is sentence-level and language independent. An unordered list of translation units, each consisting of a source-language segment and a target-language segment, is indexed. Then a source text to be translated is segmented and compared with the TM database. Exact matches and "fuzzy" matches (that is, source segments that partially match against the source-language segment in a translation unit), are displayed for a human translator to accept as is, edit, or reject. Source segments that do not result in either an exact match or a fuzzy match above a certain threshold of similarity, do not result in any target text being displayed.

There is the basic difference between an MT system and a TM system. An MT system attempts to produce a complete target text that can be used in its raw form or after post-editing by a human translator. A TM system, on the other hand, generally does not produce a complete translation but instead makes suggestions to a human translator who is responsible for producing a suitable target-language text. If a sufficient number of retrieved translation units are used by the human translator with little or no editing, a TM system may result in a much faster translation than a translation "from scratch" in which every segment of source text is translated by a human. Of course, if only a small percentage of the segments in a source text result in the retrieval of a translation unit, the use of a TM system may not significantly increase translation speed.

What is the direction of development of TM systems? While traditional TM systems are highly effective when translating a slight revision of a previously translated document (for example, the documentation for a new version of a product that involves only minor changes or a revised version of documentation that was translated before a product was finalized), they are not very effective in other contexts. For situations where the percentage of "hits" (source segments for which a usefully similar target segment is retrieved) is rather low, other TM tools are needed. One additional tool is a subsegment-level lookup feature that

searches for portions of a segment, sometimes called a "chunk", and displays all those translation units that contain that chunk of text. The translator examines those translation units and decides whether they contain useful information.

One challenge of subsegment-level lookup is that there can be an overwhelming number of hits to look through. Another is knowing which chunks are going to be found in the database. Looking up a chunk and retrieving no translation units is a waste of time. One approach to dealing with these challenges is to automatically look up subsegment-level chunks and display for the translator those chunks that were found, ranking the target language units for each chunk according to likely relevance, for example, according to the number of words surrounding the chunk that are found in both the source segment and the target segment. For inflected languages, the lookup of chunks will be more effective if language-specific morphological processing is performed on the bilingual corpus to allow for matches when the source-language chunk exists in the translation memory but in a different inflected form. Another trend in TM systems is toward the retention of the integrity of the source and target texts as bitexts, rather than as unordered sets of translation units in isolation. Bitexts allow the translator to explore as much context as desired surrounding both the source chunk and the target chunk.

There is no need to state a prediction that TM systems are moving toward automated subsegment-level lookup of chunks. This feature is already available in several commercial systems.

*My **third** prediction is that TM systems with automated subsegment-level lookup will begin to offer morphological analysis for some languages and that these systems will begin to exploit the existence of a bitext-oriented translation memory by providing features that cannot be provided when the translation memory consists of unordered translation units in isolation.*

### **3. Convergence and Quality Assurance**

It is probably obvious from the title of this article that there will be a prediction involving the convergence of MT and TM. It is not a huge step from (1) a TM system that automatically looks up and ranks chunks of text to (2) an MT system that puts those chunks together into a target-language sentence.

*My **fourth** prediction is that there will be integrated systems using the same bitext corpus that combine TM and MT under the control of the translator.*

At a general level, this is actually a very old prediction that dates back more than twenty years. What is new is to specify that these integrated systems will involve a convergence of TM and MT using the same bitext corpus.

What are the challenges for such integrated systems? The major challenge is quality. It could become easy for a translator to accept low-quality target text sentences in the interest of efficiency. It would go beyond the scope of this article to discuss quality extensively. Instead, we will introduce two questions: what is quality and how important is it?

I suggest that we use the definition of quality found in the ASTM International

standard F2575 on quality assurance in translation (ASTM 2006): degree of conformance to an agreed-upon set of specifications. Immediately, we can conclude that by this definition quality is important, since it is defined in terms of what all parties have agreed to be important and have formalized in written specifications. This is far from an absolute definition of quality. Instead, it is a flexible definition relative to a particular translation project. Theoretically, this definition is consistent with Functionalism in translation studies.

Despite this flexibility, let us examine three specifications that are commonly used: coherence, consistency, and accuracy. A translation that has no textual coherence is very difficult to read. A translation that does not use key terms consistently is likely to be confusing. And a translation that is factually inaccurate or departs from standard terms, even if the non-standard terms are used consistently, is often unacceptable. A perhaps astounding fact is that neither TM nor MT guarantees these properties in a translation. The larger the bitext corpus, the more variety will be found in translation units retrieved for a chunk of text. The hit ranked highest by some mechanical procedure that does not involve understanding real-world context as well as surrounding text may not be the best hit. Consistency is best managed by terminology management. Most translator tools already include terminology management, so it does not make sense to predict that it will be available. It is available. The question is whether we will use it effectively. Effective use of terminology management and other tools involves constant awareness of all aspects of the context of a translation. There is no indication that MT and TM will achieve this in the foreseeable future, but humans are particularly good at it.

## Conclusion

*My **fifth** and final prediction is a bit scary: in the future, the only kind of non-literary translator who will be in demand is one who can craft coherent target texts that, when appropriate, override the blind suggestions of the computer.*

This is actually good news for translators: they will be more human, rather than less. They will be involved in the entire quality assurance (QA) process of creating specifications appropriate for the audience and purpose of a particular translation and making sure they are adhered to at every step of the project. If that sounds like a project manager, so be it. The future is ours: do we want to be viewed more as file clerks afraid of being replaced by document management systems or as confident professionals who are gradually being freed from the drudgery of detailed, mechanical text manipulation so that they can focus on the bigger picture of quality assurance in information management?

The methodology for testing whether my predictions are accurate is simple: wait a few years and look at machine translation systems, translation memory systems, and the profile of well-paid human translators. I suspect that most of my predictions will come to pass within five to ten years. Let's get together again at that point and see.

## References

Hutchins W J 1986. Machine translation: past, present, future. Ellis Horwood, Chichester, UK. (Halstead Press, New York)

Chomsky, Noam 1966. Cartesian linguistics: a chapter in the history of rationalist thought. Harper & Row (New York).

P. F. Brown, J. Cocke, S. Della Pietra, V. Della Pietra, F. Jelinek, R. Mercer, & P. Roossin, 1990. "A Statistical Approach to Machine Translation," Computational Linguistics 16(2)

Skousen, Royal. Analogical Modeling of Language. Kluwer (Dordrecht, The Netherlands)

Melby, Alan K. 1982. "Multi-level Translation Aids in a Distributed System" A paper presented at the Ninth International Conference on Computational Linguistics (COLING82), Prague, Czechoslovakia, July, 1982, and published in North Holland Linguistic Series #47--COLING82, ed. by Jan Horecky, North-Holland Publishing Company pp. 215-220.

ASTM. 2006. ASTM F 2575-06, Standard Guide for Quality Assurance in Translation. Philadelphia, USA: ASTM International

Nord, Christiane 1997. Translating as a Purposeful Activity: Functionalist Approaches Explained. St Jerome Publishing

Desembre 2006

