[From: *Encyclopedia of literary translation into English*, edited by Olive Classe
(London: Fitzroy Dearborn Publishers, 2000), pp. 884-885]

---

# Machine translation

W.John Hutchins
(University of East Anglia)

The translation of texts from one natural language into another by the use of computers--known, since research began in the mid 1940s, as machine translation (MT)--requires the precise specification of bilingual equivalences and of the processes of conveying the meanings of expressions (words and sentences) from a source language (SL) text into a target language (TL) text. (For general introductions see Arnold et al. (1994) or Hutchins & Somers (1992).)

The main linguistic problems arise from the inherent ambiguities of words and sentence structures out of context, such as:

- homonymy and polysemy: *cry* as 'weep' or 'shout', *bank* as 'edge of river' or to 'financial institution', *light* as noun, verb or adjective with two or more possible meanings ('not heavy' or 'not dark').
- bilingual lexical differences: *river* can be *rivière* or *fleuve* in French, either *Fluss* or *Strom* in German; *wear* has multiple Japanese equivalents according to the object-type: coat or jacket (*haoru*), shoes or trousers (*haku*), hat (*kaburu*), ring or gloves (*hameru*), belt or tie or scarf (*shimeru*), etc.
- differences of structure: e.g. German *Das Mädchen spielt gern Tennis* vs. English *The girl likes to play tennis*; Japanese *Jishin de kenbutsu ga kowareta* is literally: 'Earthquake-by buildings collapsed', i.e. The earthquake destroyed the buildings.

Most of these differences can be handled successfully in context by the specification of grammatical categories (noun, verb, adjective, etc.), lexical collocations, 'fixed' compounds, semantic features ('human', 'animate' etc.), and case relationships (agent, instrument, etc.) together with the analysis and identification of syntactic structures and semantic relations. However, certain linguistic phenomena resist straightforward rule-based treatment, such as:

- the selection of articles (in English, French, and German) when translating from languages such as Russian and Japanese with few indicators of definiteness and indefiniteness (*Zhenshchina vyshla iz domu* 'The woman came out of the house' vs. *Iz domu vyshla zhenshchina* 'A woman came out of the house')
- the identification of pronouns and their antecedents. For example: *The soldiers shot the women. They were buried next day*. We know that the pronoun *they* does not refer to *soldiers* and must refer to *women* because we know that 'shooting' implies 'killing' and 'injury' or 'death' and that 'death' is followed (normally) by 'burial'; this identification is crucial if the TL pronoun must indicate male or female (e.g. French *elles* and *ils*.)
- differences of style, e.g. when one language prefers nominalisation (as in English technical documents: *The possibility of rectification of the fault by the insertion of a wedge is discussed*) while other languages prefer verbs (*We discuss whether it is possible to rectify the fault by inserting a wedge*).

Such examples demonstrate that the disambiguation of SL sentences and the correct selection of TL equivalents are often impossible without knowledge of (non-linguistic) features and properties of the actual objects and events described. A number of knowledge-based MT systems have been developed, mainly restricted to specific subject domains, which go beyond traditional linguistics approaches. Furthermore, there have also been other recent developments in statistical (probabilistic) methods and the

use of corpora of example translations which aim to overcome the limitations of rule-based systems (Hutchins 1995a). For example, idiomatic translation of *have an effect on* is assisted by reference to a text corpus containing aligned equivalents such as: *have a direct effect on <-> ont une influence directe à; had a direct effect on <-> ont eu une répercussion directe sur; has had a marked effect on <-> a largement influencé; had a positive effect on <-> s'est avérée positive dans*; etc.

Nevertheless, MT is unlikely to ever achieve human-quality idiomaticity. The rule-based and probabilistic nature of MT limits it, in the main, to 'literal' translations adhering relatively closely to the lexical and structural features of the source language. It works best with relatively unambiguous, terminologically 'normalised' and controlled language; indeed, Melby (1995) argues that computer-based analysis (as in MT) is inherently unable to deal with colloquial free and dynamic language.

Output from MT systems is rarely good enough for immediate publication--it has to be revised (or 'post-edited') by human translators, often at considerable cost. However, there are circumstances where poor-quality unedited output is acceptable, e.g. where rapid access is needed to vital information or where publication is not envisaged.

The quality of MT output can be substantially improved if dictionaries and grammars can be restricted to specific subject fields (sublanguages), and/or if the vocabulary and style of input texts can be controlled to ensure terminology consistency and one-one translation equivalents. In these circumstances post-editing can be reduced or eliminated, but the constraints and costs of SL control make this approach feasible only with large-volume translation throughput. However, control of terminology and simplification of style and sentence structure is desirable (even essential) in many areas of technical documentation whether intended for translation or not.

In the early years of MT research it was assumed that the goal of MT was the development of fully automatic systems capable of translating texts on any subject at a quality equivalent to that of any human translator; there were even some researchers who thought MT systems could translate literary works. This 'perfectionist' ideal of high-quality general-purpose MT has been abandoned since the mid 1960s (for the history of MT see Hutchins 1986, 1995b). MT development concentrates on systems which are cost-effective in specific environments:

- systems for multinational companies producing large-volume technical documentation in many languages within relatively restricted subject ranges. In these organisations terminology can be controlled and style can be standardised to decrease ambiguity and complexity and to improve MT output. Currently, many millions of pages are being translated in this way every year (Brace et al. 1995).
- systems for the globalisation and localisation of products (such as software) and their supporting documentation
- systems to produce rough 'less than perfect' low-quality translations for basic information needs. Organisations of all sizes want to keep abreast of commercial and technical documents reporting developments in other countries (with the option of later full translation); individuals need to understand the gists of documents, academic papers, correspondence (e.g. electronic mail), newsletters, etc. in languages they know poorly or not at all. Many cheap translation software packages are now available.
- systems to produce good-quality output from highly constrained SL texts. Business communication often follows regular patterns, and MT systems are being developed to enable messages to be interactively composed in defined frameworks and then translated into an unknown language with the assurance of accurate and fluent output.
- systems for translating spoken language in restricted domains. Under investigation are systems for telephone hotel booking and conference organisation, for interaction with databases (e.g. flight information) and for business negotiations.

These are all applications which lie outside the traditional range of human translation. Although sometimes translators are involved as post-editors of MT output, it is now more common for specially trained linguists to work in conjunction with terminologists and technical authors in multinational companies and MT services. It is clear that translators do not want to be the 'slaves' of MT systems, correcting errors which no professional would commit.

Since the mid 1960s there have been many developments in computer-based translation tools for the translation profession (some as by-products of MT research): automatic dictionaries, facilities for the management of terminology and creation of glossaries, multilingual word processing facilities, on-line access to external information databases, electronic transmission of documents, optical character recognition devices, concordance making facilities, and storage of and access to bilingual corpora of previous translations. These facilities are now being integrated in translator's workstations, and are contributing markedly increased productivity for specialist translators in science, technology, engineering, medicine, etc. wherever terminological consistency is desirable and where access to previous translations is invaluable.

In nearly all respects, the translation of literary, philosophical, biographical and cultural texts does not meet the criteria for the successful application of MT systems: vocabulary is not (and cannot be) controlled or narrowly defined; one-one translation equivalence is rare, is undesirable and is usually avoided; there are no large volumes of repetitive documents; and high quality output is demanded and expected. Literary translation avoids the 'literal' tendency found in most MT; but it is above all the virtual impossibility of accounting for cultural context in the target language which makes MT unfit for literary translation.

Although MT as such is inappropriate for literary translators, it could well be that they could make good use of translator's workstations, whenever there is need for access to specialised dictionaries (e.g. 16th century French when translating Rabelais), access to concordances, or access to other translations of the same author (e.g. Balzac or Dickens).

In general, MT is no threat to the human translator (and least of all, to the literary translator) since it concentrates on areas of translation demand which have not been, and cannot be, met traditionally by the translation profession.

**Further reading**


Arnold, D. et al. (1994): *Machine translation: an introductory guide.* Manchester/Oxford: NCC Blackwell.

Brace, C. et al. (1995): MT users and usage: Europe and the Americas. *MT News International* 12, pp. 14-19.

Hutchins, W.J. (1986): *Machine translation: past, present, future.* Chichester: Ellis Horwood.

Hutchins, W.J. (1995a): A new era in machine translation research. *Aslib Proceedings* 47 (10), pp.211-219

Hutchins, W.J. (1995b): Machine translation: a brief history. In: Koerner, E.F.K. and Asher, R.E. eds. *Concise history of the language sciences* (Oxford: Pergamon), pp.431-445.

Hutchins, W.J. and Somers, H.L. (1992): *An introduction to machine translation.* London: Academic Press.

Melby, A.K. (1995): *The possibility of language: a discussion of the nature of language, with implications for human and machine translation.* Amsterdam/Philadelphia: John Benjamins.