

Machine Translation as a Model of Human Translation

PAUL BENNETT

*Centre for Computational Linguistics
UMIST, Manchester, UK*

Machine translation (MT) has generally been seen as a purely engineering enterprise, with virtually no attention paid to whether it provides any kind of psycholinguistic model of human translation (HT). This article investigates whether MT systems offer any kind of parallel to HT, by distinguishing three ways in which MT may simulate HT, namely in terms of input-output relations, knowledge, and processing. It is suggested that MT does not, in general, produce human-like output. And while the knowledge embodied in an MT system is broadly comparable to that of an expert human translator, the former is more compartmentalized, and specifically bilingual knowledge is sparser. Processing is hard to discuss without knowing more about how people translate. However, MT can play a useful role in prompting hypotheses about HT such as whether there is a human analogue of MT's complex transfer.

Introduction

Machine translation (MT) has generally been seen as a purely engineering enterprise, designed to produce translations of acceptable quality, rather than as an attempt to build a psychologically accurate model of the way in which human translators work. In this article, however, I wish to raise the issue of the psychological status of MT, to ask whether, and in which ways, current MT systems simulate human translation, and whether this is a useful design criterion. I am interested, therefore, in the extent to which MT research can be seen as part of Computational Psycholinguistics in addition to Applied Computational Linguistics (see Thompson 1983).

I shall begin by setting out three different respects in which MT might be said to simulate human translation, and then discuss each in turn. The tentative and preliminary nature of this enquiry should go without saying.

Kinds of modelling

I distinguish three ways in which a computational system might be described as a model of some aspect of human behaviour:

- (1) a. input-output relations
- b. knowledge
- c. processing (use of knowledge).

In the case of (1a), a system behaves in the same way as a human as far as its inputs and outputs are concerned, i.e. it relates a source and target text. An MT system which achieved this would produce human-standard translations, or at least would produce translations. This kind of mimicking of translators' behaviour may seem to be a trivial kind of modelling, though I shall suggest that it is not in fact trivial.

(1b) requires a system to represent, in some form, the knowledge of a human practitioner. For an MT system, this would mean at the very least that it embodied the kind of monolingual and bilingual knowledge, both grammatical and lexical, possessed by a translator. One might require further that this knowledge be structured in a way that paralleled that of the human, though clearly even this leaves a great deal of leeway.

The final kind of modelling, (1c), implies furthermore that the knowledge be used or exploited in a comparable way to that of a human. In our case, the process of translation would follow the same course in the person who translated and in the computational model of translation. I take it that a system that was a full-scale model of the human translator would produce human-quality translations, but I regard it for now as a moot point whether aiming at (1c) would be the best way of pursuing better translation quality in MT.

Having established in broad terms the kind of modelling I am concerned with, I shall now ask just how an MT system could be said to be a model in any of these three senses. Before that, however, I want to mention a reason why questions of psychological reality might well be seen as irrelevant to MT. Translation differs fundamentally from parsing in being a conscious, learned activity, not an automatic, unconscious reflex (as pointed out by Johnson 1983). Translation is not an **input system** in the terms of Fodor (1983), and is thus quite unlike monolingual language processing and vision, two principal areas of computational modelling of human behaviour, and there would be little reason to expect it to be amenable to comparable study. The processes of pre- and post-editing, which play an important role in most MT applications, also have little place in 'ordinary' language processing.

There are other reasons, too, why one might balk at such comparisons. For instance, Sager (1994:258) argues that a machine-translated text is simply not comparable to one produced by a human. In my view, however, such comparison is permissible provided that one bears the limitations of the procedure in mind and does not use it for the purposes of contrastive evaluation.

Modelling input-output relations

With regard to (1a), I claimed in Bennett (1993a) that there is no point in building an MT system which makes the same mistakes, or the same kinds

of mistakes, as human translators. If the aim is the best quality translation achievable, it makes no sense to build in to the system deliberate mistranslations. I would now like to qualify this view, however, by noting that it would be interesting if MT systems made human-like mistakes **without this having been specifically intended**. In a comparable way, the parser for English written by Marcus (1980) stumbles over garden path sentences such as *The horse raced past the barn fell*, just as people processing such examples do.

This leads us on to the more general issue of the extent to which the output of MT systems resembles expert human translations. Given the number of MT systems and human translators around, one might be reluctant to speak of 'typical' examples of either. However, I think it would be generally accepted that even good-quality MT systems do not on the whole produce human-like output, tending to produce literal or structure-bound translations (see Hutchins & Somers 1992:138; Bennett 1994). Systems specialized for particular domains and text-types perform better, of course. Where a translation produced by a machine is inadequate, it is often ill-formed in terms of the target language grammar, which human translations generally are not, so there is little case for psychological modelling here. To illustrate this last point, consider two examples of Russian-English translation by SYSTRAN. These are taken from Knowles (1979); in each case, I give the system's output followed by a decent human translation:

- (2) a. Yesterday we the entire hour rolled themselves on a boat.
 a'. Yesterday we went out boating for a whole hour.
 b. The final stage of landing is achieved with the aid of the fire system of high intensity.
 b'. The final stage of aircraft landings is accomplished with the help of high intensity lights.

(2a, b) are just not typical of human errors. It may be objected that SYSTRAN is not exactly state-of-the-art as far as MT research is concerned, but I have deliberately chosen a system which has been extensively used in real translation situations.

As a hardly original conclusion to this section, I would suggest that MT in its present situation cannot offer human-like translations, so from this point of view it does not constitute a model of the human translation process.

Modelling the translator's knowledge

I move on now to a consideration of (1b), noting first the observation of Johnson and Whitelock that "What we try to do when we build an MT

system is to incorporate all or part of the translator's expertise into a computer program" (1987:136). They postulate five kinds of knowledge on the part of a professional technical translator:

- (3) a. target language knowledge
- b. text-type knowledge
- c. source language knowledge
- d. subject area ('real world') knowledge
- e. contrastive knowledge.

Let us examine these in the light of standard MT architectures:

- (4) a. source text ANALYSIS \Rightarrow IR \Rightarrow TRANSFER \Rightarrow IR \Rightarrow SYNTHESIS \Rightarrow target text
- b. source text \Rightarrow ANALYSIS \Rightarrow IR \Rightarrow SYNTHESIS \Rightarrow target text

In each case, IR stands for 'intermediate representation'. (4a) shows the transfer scheme, with a bilingual component, the intermediate representations being language-specific; (4b) shows the interlingual arrangement, where the IR is an interlingua intended not to be language-specific (for more on these concepts, see Hutchins & Somers 1992:73-77). I shall concentrate on the two approaches seen in (4), as space does not permit consideration of recent work in MT such as shake-and-bake and co-description, or new paradigms such as example-based MT.

Let us for now consider (4) as specifying the components of the system, rather than how the components relate in the translation process, which I will deal with below. If we compare (3) and (4) at the broadest level, it is plain that the translator's source- and target-language knowledge is captured in the analysis and synthesis components of (4). Contrastive knowledge is reflected in the transfer module of (4a), whereas a strict interlingua system has no place for bilingual knowledge. Text-type and real-world knowledge can be included in MT systems, and many systems are, for instance, aimed at particular sublanguages. However, real-world knowledge is somewhat open-ended and so is problematic. An important additional point is that human knowledge is far more flexible and less compartmentalized than that of a machine (Sager 1994:247).

The more narrowly linguistic elements, viz. (3a, c and e), deserve further discussion. (3a) and (3c), the monolingual parts, can be divided in the first instance into grammatical and lexical knowledge, with the former again being divided into morphological, syntactic and semantic. Not surprisingly, these divisions are reflected in the organization of linguistic modules in MT systems, although the amount of semantic information

may vary. The question of contrastive knowledge - which, after all, is the one characteristic of translation - is less straightforward. Particularly in multilingual systems, one goal is to reduce the size of the transfer modules. In Eurotra (see Allegranza et al 1991), the intention was to limit transfer to simple lexical transfer (statements of the form *cup* \Rightarrow *tasse*). This would mean that, at best, there was no need to give statements of context in lexical transfer rules or to have rules which altered structure; to some extent I shall focus on Eurotra here, as it is the MT system I know most about. The intermediate representations were language-specific primarily as far as lexical items were concerned (for brief motivation, see Bennett 1994:14-16).

One consequence of this is that it is difficult to exploit language-pair specific translation strategies above the lexical level. So one cannot directly incorporate rules of thumb such as 'SL structure X is regularly translated as TL structure Y' (e.g. a French active clause with *on* as subject rendered as an English agentless passive), since (a) transfer modules act on intermediate representations which neutralize many surface aspects of structure, and (b) they ought not to contain such structure-changing rules anyway. This conclusion would only be avoided in the case of major structural differences which were not neutralized in the IR, such as the rendition of *to like* as *plaire* or *piacere* (with switched-round arguments), or the translation into Romance languages of motion expressions such as *swam across the river* (cf. (6c) below).

Let's take one more specific example of human contrastive knowledge:

In Romance language medical texts, you normally assume that a SL noun plus adjective group, e.g. *radioactivité plasmaticque*, is going to be switched round to a noun-plus-noun compound, 'plasma radioactivity', provided the SL adjective is formed from a noun of substance. (Newmark 1988:213)

Adjectives such as *plasmaticque* are commonly known as relational adjectives; the correspondence noted by Newmark applies not just to medical language. In Eurotra research on compounds (see Bennett 1993b, Carulla 1994), such cross-linguistic differences were solved by representing relational adjectives as their underlying nouns in the intermediate level, e.g. *plasma* for *plasmaticque*. This permits simple lexical transfer although the problem of when to produce a relational adjective in synthesis remains. My point here, however, is just that the piece of knowledge set out by Newmark is not captured as such, rather it is spread over several modules (French analysis says such adjectives are represented as nouns, while English synthesis says modifying nouns in compounds are generally realized simply as nouns); French-English transfer does not deal

with this relation.

This conclusion holds *a fortiori* for interlingual systems which have no transfer components at all. I take it, then, that one big difference between the knowledge of an MT system and of a human translator is that the contrastive knowledge of the latter is far richer and not purely lexical. Expert translators presumably acquire masses of instances of the kind of specific knowledge illustrated in the passage from Newmark. Of course, one might argue from this that MT systems should be bilingual only, or that only bilingual systems merit comparison with human translation.

One level of linguistic knowledge omitted in the discussion above was that of text or discourse structure. If present, this could be spread over (3a, c, e) in an MT system, but generally MT systems pay little attention to textual coherence or cohesion. I proposed (Bennett 1994) that discourse considerations are best confined to synthesis, as they are TL-dependent. The fact remains, however, that MT systems are less successful in representing this kind of knowledge - understandably so, given the subjective nature of many of the concepts used in linguistic research on these topics.

To conclude this section, it seems reasonable to say that MT systems are partial models of the human translator's knowledge, in that they represent knowledge of TL and SL grammar and lexicon, plus some text-type and real-world knowledge. But, while they necessarily contain contrastive information, this is, in the ideal case, far sparser than with human translators.

Modelling the translation process

Our final aspect of modelling is where a system translates in the same way a human does, going through the same processes and not just using the same knowledge. An immediate problem with discussing this is that relatively little seems to be known with any degree of confidence about the way people translate. The fact that translation is a conscious, deliberate process (see above) seems to make it less amenable to the psycholinguistic techniques used to study human sentence processing and production. For instance, I am not aware of any evidence concerning the time taken to translate different kinds of structure, e.g. as to whether examples where a phrase (italicized in (5)) is a long way from its canonical position (shown by a bullet) take a relatively long time to translate (see further below):

- (5) a. *What* do you think John is intending to do • next?
 b. *Mary* appears to be believed to have been arrested •

For instance, one might wonder whether (5a) is easier/quicker to trans-

late into a language with wh-fronting than one without, such as Japanese, and how difficult is translation of (5b) into a language like French that does not allow such repeated raisings. Such questions might be compared to psycholinguists' (former) investigations into the Derivational Theory of Complexity, although this has now lost its appeal.

Because of the problems of experimental study of translating, investigators have tended to rely on think-aloud protocols whereby translators report on their mental processes while translating (e.g. Lörscher 1991). For the sake of concreteness, though, I shall here take the model offered by Bell (1991, ch. 2), which is based more on research in psycholinguistics, and less on work specifically relating to translation.

It would follow from (4a) that MT systems perform translation in three sequential stages, analysis - transfer - synthesis. A naive initial view would be that human translators do the same thing, though the stages will overlap with each other rather than being sequential (Lörscher 1991:17). Bell, however, offers an analysis-synthesis picture more like (4b). The analysis procedure he sees as common to both translator and monolingual reader, though this is itself a controversial assumption, and one that needs testing. Analysis consists of syntactic, semantic and pragmatic analysis, though Bell stresses that these processes are integrated, not simply consecutive. The output is a semantic representation which contains information about clause structure, propositional content, discourse structure, register and illocutionary force. This representation is now handed over to the synthesis component, whereby pragmatic, semantic and syntactic synthesizers produce the TL text; synthesis is discussed at much shorter length, though. His sample semantic representation (p. 67) does contain some words of English, so I assume that he is not claiming that these representations are entirely in some universal semantic metalanguage. In any case, I find it surprising that his model reduces a translator's expertise to knowledge of SL and TL, with no apparent place for contrastive knowledge. Leaving this aside, though, how can MT systems be said to measure up to this model?

One large difference is that the pragmatics modules of Bell's model have little if any place in most MT systems (see above). Another is that processing in NLP systems in general, not just in MT systems, tends to be much more sequential (e.g. first syntax, then semantics). This is particularly so in the case of the stratificational approach of Eurotra where analysis involves assigning a surface syntactic analysis, then a relationally-based one, then the more semantic-oriented IR, with synthesis going through the reverse steps. In lacking a transfer module, Bell's model is clearly close to that of the interlingua architecture, (4b). One particular question prompted by these remarks is the nature of synthesis in MT, in particular how it relates to synthesis in other areas of NLP, especially

what kind of representation is seen as the input to synthesis. I would also argue that the idea of translating compositionally, via the combination of ever-larger translation units (Bennett 1994) may be common to both MT and human translation. We should however note the claim (Somers 1993) that, for human translators, structure-bound translation is the last resort, rather than the first choice.

Rather than continue with such reasoning, however, I wish to pursue a slightly different line. I discussed above the idea of transfer and simple vs. complex transfer. It seems to me that a worthwhile question to ask is whether human translation has any analogue of the process of complex transfer. In other words, do cases where ST and TT differ greatly in terms of morphosyntactic structure take longer to translate than those where ST and TT are parallel in terms of surface structure, or those where they differ on the surface but are similar at the IR level (cf. the example of *radioactivité plasmatique*)! I would suggest that these are the kinds of question that MT opens up for research into the translation process. And given that different MT systems will offer differing classifications of individual cases, they may help to show which architecture or system is closest to modelling the human translator.

Let's make this point more precise. ST-TT relations may be divided for these purposes into three types, bearing in mind that the notion of 'same/different IR' is system-dependent:

- (6) a. Same syntactic structure in both:
Charles works in London. / Charles travaille à Londres.
- b. Different syntactic structure, same IR:
I believe Charles to be rich. / Je crois que Charles est riche.
- c. Different at both syntactic and IR levels:
Charles ran across the street / Charles a traversé la rue en courant.

One might hypothesize that (6a-c) represent increasing difficulty for translation, and correspondingly take longer to translate. I do not know if this is correct, and the experimental problem would be to distinguish difficulty of translation from difficulty of analysis, but this ought in principle to be possible. One could add further cases, e.g. those where the TT contains more information than the ST (e.g. about number or aspect). Some other examples of complex mappings worth investigating from this point of view are listed in (7):

- (7) a. Simple predicate to support verb:
Charles murdered Marie. / Charles a commis un meurtre contre Marie.

- b. Verbal category expressed by inflection or via an auxiliary:
Marie will leave soon. / Marie partira bientôt.

Some of the restructurings examined in Tsutsumi (1990) are also relevant here. MT-oriented linguistics, therefore, can offer a classification of constructions which may be tested for their processing difficulty.

Without far more being known about how people translate, it is hard to discuss whether MT systems actually model the way human translators work. I therefore see this research programme as an instance of "simulation of poorly understood systems" (Simon 1981): this can still be useful, especially if we are willing to abstract from the detail of some set of phenomena, which makes modelling easier.

Concluding remarks

So does MT provide a model of human translation? I hope it will be clear that no simple yes or no answer is appropriate. MT systems do not model human translators in the sense of producing human-standard translations or making human-type errors. They may contain the same kind of information/knowledge possessed by human experts, but they often divide this up differently. It is hard to say whether MT systems translate the same way people do, but one useful question raised by our study is whether there is a human equivalent of complex transfer, and whether what takes a person longer to translate also takes a machine longer. At the very least, MT provides for the formulation of questions that the empirical study of human translation should address.

Acknowledgements

I am very grateful to Juan Sager and Harry Somers for comments on an earlier version.

References

- Allegranza, V., P. Bennett, J. Durand, F. Van Eynde, L. Humphreys, P. Schmidt and E. Steiner (1991) 'Linguistics for Machine Translation: the Eurotra Linguistic Specifications', in C. Copeland, J. Durand, S. Krauwer and B. Maegaard (eds) *The Eurotra Linguistic Specifications*, Luxembourg: Office for Official Publications of the European Communities, 15-123.
- Bell, R. (1991) *Translation and Translating*, London: Longman.
- Bennett, P. (1993a) 'A Comparative Evaluation of the Eurotra Preference Mechanism', in P. Bennett and P. Paggio (eds) *Preference in Eurotra*, Luxembourg: Office for Official Publications of the European Communities, 91-103.

- Bennett, P. (1993b) 'A Multilingual Translation-oriented Typology of Compound Nouns', *Traitement Automatique des Langues* 34(2):43-58.
- Bennett, P. (1994) 'The Translation Unit in Human and Machine', *Babel* 40:12-20.
- Carulla, M. (1994) 'Relational Adjectives in the Translation from Germanic Nominal Compounds into Romance Languages', in *Proceedings of the Workshop on Compound Nouns: Multilingual Aspects of Nominal Composition*, Geneva: ISSCO. 103-107.
- Fodor, J. (1983) *The Modularity of Mind*, Cambridge, MA.: MIT Press.
- Hutchins, W. J. and H. L. Somers (1992) *An Introduction to Machine Translation*, London: Academic Press.
- Johnson, R. (1983) 'Parsing - an MT Perspective', in K. Sparck Jones and Y. Wilks (eds) *Automatic Natural Language Parsing*, Chichester: Ellis Horwood, 32-38.
- Johnson, R. and P. Whitelock (1987) 'Machine Translation as an Expert Task', in S. Nirenburg (ed) *Machine Translation: Theoretical and Methodological Issues*, Cambridge: Cambridge University Press, 136-144.
- Knowles, F. (1979) 'Error Analysis of SYSTRAN Output', in B. Snell (ed) *Translating and the Computer*, Amsterdam: North-Holland, 109-133.
- Lörscher, W. (1991) *Translation Performance, Translation Process, and Translation Strategies*, Tübingen: Gunter Narr.
- Marcus, M. (1980) *A Theory of Syntactic Recognition for Natural Language*, Cambridge, MA.: MIT Press.
- Newmark, P. (1988) *A Textbook of Translation*, London: Prentice-Hall.
- Sager, J. C. (1994) *Language Engineering and Translation*, Amsterdam/Philadelphia: John Benjamins.
- Simon, H. (1981) *The Sciences of the Artificial*, Cambridge, MA.: MIT Press, 2nd edition.
- Somers, H. L. (1993) 'Current Research in Machine Translation', *Machine Translation* 7: 231-246.
- Thompson, H. (1983) 'Natural Language Processing: a Critical Analysis of the Structure of the Field, with Some Implications for Parsing', in K. Sparck Jones and Y. Wilks (eds) *Automatic Natural Language Parsing*, Chichester: Ellis Horwood, 22-31.
- Tsutsumi, T. (1990) 'Wide-range Restructuring of Intermediate Representations in Machine Translation', *Computational Linguistics* 16:71-78.