

Machine Translation without a source text

Harold L. SOMERS, Jun-ichi TSUJII and Danny JONES

*Centre for Computational Linguistics
UMIST, PO Box 88
Manchester M60 1QD, England*

Abstract

This paper concerns an approach to Machine Translation which differs from the typical 'standard' approaches crucially in that it does not rely on the prior existence of a source text as a basis of the translation. Our approach can be characterised as an 'intelligent secretary with knowledge of the foreign language', which helps monolingual users to formulate the desired target-language text in the context of a (keyboard) dialogue translation systems.

Keywords: Machine translation; natural language interface; dialogue

Introduction

Machine Translation (MT) or natural language translation in general is a typical example of the 'under-constrained' problems which we often encounter in the field of artificial intelligence¹. That is to say, the same 'messages' can and should be translated differently depending on the surrounding contexts (where and when they are used), and on the speakers' intention (what they really want to express) etc. It is all too often the case that this information, which is necessary for the selection of the appropriate overall target text structure, is not made explicit in source texts prepared for translation. The author of the source text naturally follows the 'rules' of the source language in preparation of source texts and assumes that the factors which will affect the selection of target expressions are self-evident.

MT systems developed so far or being developed have been trying to compensate this genuine property of language translation by extending the units of translation from sentences to texts (e.g. Rothkegel 1986, Weber 1987) or

by introducing 'understanding' based on 'domain specific knowledge' (as in the 'sublanguage' approach - cf. Kosaka *et al.* 1988, Lehrberger & Bourbeau 1988). This course of research would be inevitable if we were to confine ourselves to translation of prepared texts which already exist before translation. In such cases, we have to recover from text itself or by using extra 'knowledge', such implicit information which is necessary for formulating target expressions.

However, we can imagine a quite different course of research for developing a different type of MT system, i.e. an 'expert' system which can play the role of an 'intelligent secretary with knowledge of the foreign language'. Such a system does not require the user (the writer) to prepare full source texts in advance. It starts from rough sketches of what the writer wants to say and gathers the information necessary for formulating target texts by asking the writer questions, because the writer is the person who really intends to communicate and has a clear idea about what s/he wants to say. We can get much richer information through such interactions than in the usual written text translation by professional translators. Through interaction, we can get information concerned with, for example, the user's intention which is not explicitly expressed in the 'text' to translate but which is nonetheless necessary for producing quality target texts.

This sort of system is different from the widely promoted 'Translator's Workbench' idea (e.g. Kay 1980, Melby 1982), the main aims of which are to help *translators* to translate texts. In this scenario, both the system and the user have knowledge about both source and target language, and it is sometimes difficult to see where the most appropriate division of labour should occur: indeed, there is sometimes a conflict between what the system offers the translator-user, and what the user already

¹ The authors would like to acknowledge the contribution to this work of the other members of the project team: Bill Black, Jeremy Carroll, Anna Gianetti, Makoto Hirai, Natsuko Holden, John Phillips and Kenji Yoshimura.

knows, or between the extent to which the system or the user should take the initiative, which might differ from occasion to occasion. On the other hand, in the proposed expert system scenario, the partition of knowledge is clear: the system knows mainly about translation, the writer knows only about the desired communicative content of the message. There is no conflict between what the system assumes to be the extent of the writer's (the user's) knowledge, nor in the writer's expectations. In this respect we are following the line taken by Johnson & Whitelock (1987), and the work here at UMIST on the ENiran project (Whitelock *et al.* 1986, Wood & Chandler 1988) developing an MT system for a monolingual user.

MT systems so far have been developed based on the implicit assumption that source texts contain all (or almost all) the information necessary for translation. We take as a starting point that this assumption is not necessarily true, especially when we consider pairs of unrelated languages where cultural as well as linguistic differences contribute to this problem.

Notice that the concept of 'source text' in the above is quite different from that in the normal context of MT. That is, we do not have a source text to translate as such, but instead, the user has his/her communicative goals and the translation system can help to formulate the most appropriate target linguistic forms by gathering information necessary to accomplish these goals through 'clarification dialogues'.

It could be argued that this generation of a target text on the basis of something other than a source text is not 'real translation'. Such an argument might derive from an overly traditional view of translation where a translator gets some text (say, in the post) and sits at a desk with a bilingual dictionary and translates 'blind' i.e. with no actual knowledge of the writer's intentions, goals, etc. There is a sense in which second generation MT systems simply reflect this scenario of a translator. Of course, the best translations are done by a translator who can ask the original author "What did you mean when you said...?"; by the same token we believe we can build a better translation system if we can elicit such information from the originator of the 'text' at the time of 'writing'.

General background to the research

This research is undertaken in the context of the more general activities of the Japanese ATR

research programme into automatic interpretation between English and Japanese of telephone conversations. As such it is oriented towards translation of dialogues. One approach to dialogue translation has been the 'phrasebook' approach of Steer & Stentiford (1989). In this speech translation prototype system, set phrases are stored, as in a holidaymaker's phrasebook; they are retrieved by the fairly crude, though effective, technique of recognising keywords in a particular order in the input speech signal. The main disadvantage of this system is its inflexibility: if the phrase you want is not in the phrasebook, you cannot say anything.

In the research programme to be reported here, we are not concerned with speech processing per se, and we assume the context of an on-line keyboard conversation function such as *talk* in UNIXTM (cf. Miike *et al.* 1988). It has been found that keyboard conversations have the same fundamental features as telephone conversations, notwithstanding the obvious differences between written and spoken language (Arita *et al.* 1987, Iida 1987). Furthermore, we restrict ourselves to goal-oriented dialogues, i.e. dialogues where one participant is seeking information from the other: our experimental domain is dialogues for a conference registration and hotel reservation system.

When such conversations are subjected to the additional distortion of being transmitted via a traditional MT system, several further problems accrue, as the *talk* experiment mentioned above showed, notably when mistranslation occurs. The problem of human-machine interaction in the specific area of clarification dialogues for MT must be studied. The need to incorporate different types of clarification dialogue has general implications for the question of system architectures for interactive MT systems. This aspect is discussed in detail below.

In the above scenario, the system tries to gather information necessary for formulating target texts through interactions. This means the system formulates target texts by adding information to 'source texts' (in the conventional sense). We can extend this idea further. In the extreme case, we can imagine a system which has stereotypical target texts in certain restricted domains (e.g. business correspondences in specific areas), retrieves appropriate texts through dialogues with users and reformulates them to fulfill the specific

requirements expressed by users. In this scenario, the MT system becomes a kind of multilingual text generation system and adds a lot of information not contained in the 'source text' at all. This idea has been investigated here at UMIST in the context of a research programme for British Telecom (Jones & Tsujii 1990), and has significantly influenced the research reported here (a similar idea for 'automated text composition' in Japanese has been suggested by Saito & Tomita 1986).

Dialogue MT

It is important to emphasize that there is a basic difference between Dialogue Machine Translation (DMT)² systems on the one hand and conventional MT systems on the other, namely the difference of user types. In DMT, users are dialogue participants who actually have their respective communicative goals and who really know what they want to say. On the other hand, the users of conventional MT are typically translators who, though they have enough knowledge about both languages, lack 'complete understanding' of texts to be translated.

This difference in user-types leads to different characterizations of interactions between MT systems and their users. We have to take into account what this difference implies in designing actual DMT systems. The main implications can be summarized as follows.

In DMT, the system can ask in theory any questions to elicit the information necessary for translation which is not explicitly expressed in the 'source text'. This is impossible in conventional MT, because the users do not have 'complete understanding' of the context in which the texts are prepared, and the users (who are translators) simply could not answer such questions. (It is often the case that even human translators would like to consult the authors of the original texts in order to produce a good translation.) In order to exploit this advantage in DMT however, we have to overcome several related difficulties.

² Our concept of DMT should be distinguished from 'Dialogue-based MT' as proposed by Boitet (1989), in which dialogue is used to clarify the author's intentions in the context of a personal MT system. This is also the case in our DMT, with the crucial difference that the object of translation in our case is also part of a dialogue, i.e. the user's dialogue with a third party. Clearly however, there are significant areas of overlap between our project and Boitet's.

First, in DMT there are several different types of dialogues, any of which may start up or be resolved at any given time: these dialogues include

- (a) user-user object-level dialogues
- (b) user-user meta-level dialogues (e.g. in which one participant in the dialogue asks the other participant questions to clarify the meaning or intentions of his/her statements)
- (c) user-system dialogues typically initiated by the system, concerning the progress of the object-level dialogue, disambiguating ambiguous object-level dialogue, i.e. what the user wants to *say next*.
- (d) user-system meta-level dialogues typically initiated by the user, concerning clarification of the object-level dialogue, i.e. what was *just said*.

One of the foreseeable difficulties in DMT is how to distinguish these different modes of dialogue, that is, how systems can distinguish, first of all, utterances of types (a) and (b) to be translated and transmitted, from utterances of type (d) which should not be translated. In particular, dialogues of types (b) and (d) may be difficult in some cases, because the user posing questions of clarification cannot generally recognize whether the difficulties of understanding come from 'errors' in translation or from the other participants' utterances themselves. For examples of this effect, see Miike *et al.* (1988).

Dialogues of type (c) are found in some form in most conventional interactive MT systems; note that with monolingual users such dialogues are quite different from those found in the 'Translator's Workbench' type of system, since it is particularly difficult to phrase interactions concerning problems of transfer when the user is not expected to know anything about the target language, and when current frameworks do not allow us to specify the relationships among possible translations defined by different structural correspondence rules. On the other hand, regarding problems with analysis, a particularly useful result of the research on ENtran was to see to what extent potential ambiguities could be recognised on the basis of structures computed by more or less traditional parsing techniques (i.e. charts). For dialogues of type (c) we are guided by the work of Jones & Tsujii, mentioned above.

The British Telecom work concerns a system for generating business letters in French,

German and Spanish on the basis of an essentially menu-driven interface (in English). The system has a set of pretranslated fragment pairs some of which have slots for variable elements to be inserted (e.g. the name of a company, or a product) which may or may not be translated in a conventional manner. The system-user dialogue aims at selecting the appropriate target-language expression (TLE) fragment corresponding to some source-language expression (SLE) and compiling the TLEs in the appropriate sequence so as to generate the required output. Notice that, since the fragments have been pretranslated (presumably by a competent translator), the result is of a guaranteed high quality.

This idea is developed in the following ways. First, we assume that the interface menu is replaced by a much more complex 'model dialogue' (see below). In the sense that the pretranslated fragment pairs are associated with particular points in the model dialogue, they can be said to be not just pairs of SLEs and TLEs but in fact triples, since they are identified by a description of the dialogue context (DC) which conditions the equivalence of the two expressions, by specifying the point in the model dialogue at which they are identified, thus: <SLE_i,TLE_i,DC_i>. It is possible for a given SLE, there may be several TLEs depending on the particular DC, thus:

<SLE_i,TLE_i,DC_i>
 <SLE_i,TLE_j,DC_j>
 <SLE_i,TLE_k,DC_k>

For example, the English response *OK* in a dialogue may correspond to Japanese *wakarimashita* when something is being explained, *ii desu yo* when asserting agreement, or *ijoo desu* when it indicates completion of the discussion and a change of topic.

The task of the DMT system can now be divided between first locating the appropriate set of triples involving a given SLE, and then locating the appropriate TLE for that SLE according to the DC.

If we assume that the SLEs are not just 'canned texts', but actually types of text templates of varying linguistic complexity (i.e. from set phrases through to syntactic patterns – see below), it can be seen that the first part of the above task can be achieved by traditional techniques of parsing or by some other matching procedure. The set of different DCs for a given SLE can be used to trigger a clarification dialogue so as to determine the

appropriate TLE.

In this scenario the user has taken the initiative in the dialogue, by 'typing in' what s/he wants to say, and having the system find the appropriate triple.

Two other scenarios are also possible. In one, the system retains the initiative, and rather like in the menu-driven system, selects (or seeks via a meta-dialogue) the next appropriate DC, and then offers a range of appropriate SLEs for selection. In this sense the <SLE,TLE> pair for a given value of DC can be regarded as a 'conditioned equivalence pair'.

Finally, in a mixed-initiative scenario, the user and the system collaborate in the following way: first, a communicative goal is established, and with it a sequence of DCs corresponding to the 'dialogue plan'. The user then makes a proposal for the next utterance in the dialogue, and the system searches its database for the nearest apparently appropriate <SLE,TLE,DC> given the user's input (corresponding to the SLE) and the DC as given by the dialogue plan. If an exact match is found, the TLE is generated and the object-level dialogue continues. However, if an exact match is not found, the system gets the user to modify the SLE until it more closely matches the SLE selected by the system.

Model dialogue

The important issue in the above is that the equivalence relation of the two expressions is not guaranteed by the expressions themselves but by the DCs which are given rather independently of the informational content of the two expressions in the triples. In a context such as business correspondence, it might be the case that much less information is necessary to identify the relevant triple than that conveyed by the actual linguistic expressions and that, because each individual language usually has its own conventions which letters must follow, the actual informational contents of the two expressions might be different. The same is true of certain types of dialogues. For example, there are conventional phrases used in Japanese phone calls (Nagasaki 1971) which, if translated literally, would probably mystify the non-Japanese dialogue partner:

*Sorry to disturb you when you are busy /
 eating / about to go to bed / still asleep
 (depending on time of day)*

Sorry to have had to disturb you

Sorry for having talked too much
Excuse me for bothering you
Thank you for going out of your way to
answer the phone
I assume it is inconvenient for you now, but...
I am sorry for phoning you without warning
I wasn't expecting to phone you, but...

One important research question is what exactly the DC should look like. Our current assumption is that DC will actually refer to a point in a 'model dialogue', probably a flexible network of script-like structures indicating possible dialogues that the system can translate, perhaps along the lines of work by Wachtel (1986) and Reilly (1989). We have not yet finalised our ideas in this area, but we are considering in particular how to model suitably flexible dialogue structures within the domain in question, the problem of interactions between the model dialogues and the meta-dialogues, as well as the mechanisms which enable the system to navigate its way through the model dialogue network in response to the user's input.

'Canned text' and extensions

It was stated above that the nature of the SLE and TLE pairs should be varied. In particular, because of the need for flexibility as compared to the British Telecom work described in Jones & Tsujii (1990), we assume that the system will permit some degree of conventional compositional translation. So SLEs and TLEs are not always texts, or 'paratexts' (i.e. texts with slots for proper names or simply translated noun phrases, etc.) but, in some cases, structural descriptions of a more conventional kind. This clearly implies that within the system there is a need for analysis (and generation) of the kind found in conventional MT systems. In particular, where appropriate texts or paratexts are not found for a given input, and the dialogue management part of the system is satisfied that 'free input' is an available option at this point in the model dialogue, then the system becomes more like a conventional MT system, though with the special characteristics of an MT system which interacts with a monolingual user.

For the most part, however, it is assumed that there is a stereotyped set of functions involved in performing a global communicative function in a restricted domain. We can assign surface representations to these functions which restrict the form of expression to a certain extent in

order capture functional regularities in communication and to guarantee high quality translations. When the system encounters unexpected input, it has a choice of trying to steer the user towards input which is more within its expectations, or to abandon temporarily its assurance of high-quality translation and operate in a more traditional manner.

It may be asked why we need the model dialogues, the canned text and paratexts, and conditioned equivalence pairs: would it not be better simply to have a long pre-composition phase where the writer interacts with an expert system which asks lots of questions about intentions and goals and then uses this knowledge (if required) in a conventional parse-and-disambiguate system. Of course this would be another way of addressing the problem of under-specified texts, but it is not clear what type of questions could be asked unless a specific domain of composition was pin-pointed. This brings us back to domain knowledge, which in this case is expressed as knowledge about what the user can ask next, which we capture in the model dialogues.

Conclusion

It is nowadays accepted that we cannot expect to have fully automatic high-quality MT. We have to develop systems which allow flexible and effective human interventions. Our idea is to explore diversified approaches to interactive MT and in particular we seek to develop an interactive system for monolingual users. Furthermore, it seems that several interesting new approaches become apparent once we escape from the basic assumption of the existence of a concrete source text, and explore the idea of 'MT without source texts'.

References

- H. ARITA, K. KOGURE, I. NOGAI, H. MAEDA & H. IIDA (1987) 'Media ni izon suru kaiwa no yōshiki: denwakaiwa to kiibōdo no kaiwai no hikaku (Media-dependent conversation manners: comparison of telephone and keyboard conversations)'. *Jōhō Shori Gakkai* 87.34 (Jōhō Shori Gakkai Kenkyū Hōkoku, Shizen Gengo Shori 61-NLP-5, 1987.5.22).

- C. BOITET (1989) 'Speech synthesis and Dialogue Based Machine Translation'. ATR Symposium on Basic Research for Telephone Interpretation, Kyoto, December 1989. Preprints, 6-5-1-9.
- H. IIDA (1987) 'Distinctive features of conversations and inter-keyboard interpretation'. Workshop on Natural Language Dialogue Interpretation, Advanced Telecommunications Research Institute (ATR), Osaka, November 1987.
- R.L. JOHNSON & 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.
- D. JONES & J. TSUJII (1990) 'High quality machine-driven text translation'. Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, Austin, Texas, June 1990.
- M. KAY (1980) The proper place of men and machines in language translation. Research Report CSL-80-11. Xerox Palo Alto Research Center, Palo Alto, California, October 1980.
- M. KOSAKA, V. TELLER & R. GRISHMAN (1988) 'A sublanguage approach to Japanese-English machine translation'. In D. Maxwell, K. Schubert & T. Witkam (eds) *New directions in machine translation*, Dordrecht: Foris, 109-120.
- J. LEHRBERGER & L. BOURBEAU (1988) *Machine translation: Linguistic characteristics of MT systems and general methodology of evaluation*. Amsterdam: John Benjamins.
- A.K. MELBY (1982) 'Multi-level translation aids in a distributed system'. In J. Horecký (ed) *COLING 82: Proceedings of the Ninth International Conference on Computational Linguistics*, Amsterdam: North-Holland, 215-220.
- S. MIIKE, K. HASEBE, H. SOMERS & S. AMANO (1988) 'Experiences with an on-line translating dialogue system'. 26th Annual Meeting of the Association for Computational Linguistics. Buffalo, NY, June 1988. Proceedings, 155-162.
- K. NAGASAKI (1971) (*Hito ni warawarenai Kotoba dzukai to hanashi kata*. Tōkyō: Bunwa Shobo.
- R. REILLY (1989) 'Communication failure in dialogue: implications for natural language understanding'. In J. Peckham (ed) *Recent developments and applications of Natural Language Processing*, London: Kogan Page, 244-261.
- A. ROTHKEGEL (1986) 'Textverstehen und Transfer in der maschinellen Übersetzung'. In I. Bátori & H.J. Weber (Hgg) *Neue Ansätze in Maschinellem Übersetzung: Wissensrepräsentation und Textbezug*, Tübingen: Max Niemeyer Verlag, 197-227.
- H. SAITO & M. TOMITA (1986) 'On automatic composition of stereotypic documents in foreign languages'. Presented at 1st International Conference on Applications of Artificial Intelligence to Engineering Problems, Southampton, April 1986. Research Report CMU-CS-86-107, Department of Computer Science, Carnegie-Mellon University.
- M.G. STEER & F.W.M. STENTIFORD (1989) 'Speech language translation'. In J. Peckham (ed) *Recent developments and applications of Natural Language Processing*, London: Kogan Page, 129-140.
- T. WACHTEL (1986) 'Pragmatic sensitivity in NL interfaces and the structure of conversation'. 11th International Conference on Computational Linguistics, Proceedings of Coling '86, Bonn, 35-41.
- H.J. WEBER (1987) Converging approaches in Machine Translation: domain knowledge and discours [sic] knowledge. Linguistic Agency University of Duisburg Series B, No.164.
- P.J. WHITELOCK, M.M. WOOD, B.J. CHANDLER, N. HOLDEN & H.J. HORSFALL (1986) 'Strategies for interactive machine translation: the experience and implications of the UMIST Japanese project'. 11th International Conference on Computational Linguistics, Proceedings of Coling '86, Bonn, 329-334.
- M.M. WOOD & B.J. CHANDLER (1988) 'Machine translation for monolinguals'. In D. Vargha (ed) *COLING Budapest: Proceedings of the 12th International Conference on Computational Linguistics*, Budapest: John von Neumann Society for Computing Sciences, 760-763.