[Proceedings of the *Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages*, Colgate University, Hamilton, New York, August 14-16, 1985]

Feasibility Study Of Personal/Interactive Machine Translation Systems†

Masaru Tomita

Computer Science Department,
Carnegie-Mellon University
Pittsburgh, PA 15213

ABSTRACT

Most existing practical machine translation systems are designed to translate documentation, such as technical papers and manuals. However, there is a growing need for translating not only large texts but also personal short texts such as letters and informal messages. The conventional machine translation systems, which are intended to translate large texts, are not very suitable for these kinds of small jobs. We need an interactive system which has a totally different design philosophy. This paper describes the design philosophy of personal/interactive machine translation system, and studies its feasibility. *Key Words:* Machine Translation, Personal Machine Translation, Interactive Machine Translation, Man-Machine Interaction.

1. Introduction

Most existing practical machine translation (MT) systems are designed to translate off-line large documents, such as technical papers and manuals [3]. However, there is a growing need for translating not only large texts but also personal short texts such as letters, telexes and informal messages. Because conventional MT systems, which are intended to translate large texts, are not very suitable for these kinds of small jobs, we need a different type of system, which we refer to as a Machine Interpretation (MI) system, in contrast with the conventional Machine Translation (MT) systems. The MI system, however, should not be just a miniature version or extended version of the conventional MT systems. We must approach the MI systems with a totally different design philosophy. In the next section, we contrast the design philosophy of an MI system with that of an MT system.

In this paper, we focus on general purpose systems, that are to handle texts of an arbitrary or wide domain. Special purpose systems, that are to handle texts of a very restricted domain (e.g. airplane reservation, business letters of a certain topic), would require different approaches. A number of approaches to those special purpose systems have been suggested [2, 7], but they are excluded from our discussion.

2. The Design Philosophy

It seems that all general purpose MT systems are not fully automatic, but require human assistance such as post-editing or pre-editing. To make an MT system practically useful, the cost of the human assistance must be significantly less than the cost of human translation of the entire text

† This research was sponsored by the Defence Advanced Research Projects Agency (DOD), ARPA Order No. 3597. monitored by the Air Force Avionics Laboratory Under Contract F33615-81-K-1539. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the US Government.

without a computer. This cost of human assistance is a crucial parameter in developing practically useful MT systems.

Next let us consider an MI system intended to translate personal and small texts. The primary goal of the MI system is not to reduce the cost of translating a document. Rather, the goal is to enable the user to translate a small text without a human translator or specialist, who may not be available to translate one or two paragraphs immediately on demand. Because the goals of an MI system are different from those of the MT systems, the conditions mentioned in the previous paragraph are not necessarily required for an MI system. In other words, the cost of human assistance is not as crucial in an MI system as the demand availability of the translator. Suppose a small text can be translated in 3 minutes by a human translator. If an MI system requires 6 minutes of non-specialist human assistance to do the job, the user probably does the job by himself using the MI system, rather than calling a human translator. It may require a 30 minute overhead to find the translator, send the document, track it and so on, which must be added to the 3 minute translating time. Also, one wants results in minutes, rather than days - as typically required for translation services who queue their documents before performing the 3 minute translation. Hence, one does not want to and may not be able to call a human translator each time one has such a small job. Thus, it is more acceptable for an MI system to count on human assistance than it is for a large scale MT system, because the MI system's jobs are always small.

On the other hand, in order to make an MI system practically workable, the system must satisfy a number of conditions that MT systems need not satisfy. First of all, in an MI system, the type of assistance required must be knowledge that all users possess, not expert knowledge requiring specialists. The system should assume that a user speaks only his own language, and does not know anything about foreign languages. Also, it should be assumed that the user is neither a computer engineer nor a linguist. Thus, the user must not be required to know the target foreign language, computer science or linguistics in general. This is because the user does not want to call any of these specialists each time he has a small job. On the contrary, an MT system can afford such specialists, as long as the total cost of human assistance is less than the cost of having human translators to do the same job without the MT system. Indeed, most existing practical MT systems require specially trained persons as post editors (a post editor is usually a full-fledged translator).

A second condition for an MI system to be practically workable is that the system response must be reasonably quick since an MI system runs on an interactive, real-time environment. The user does not want to wait for minutes to be asked each questions as he translates a small text. By contrast, since an MT system usually runs as a batch system, the response time is not such an important factor; only the time for human assistance matters.

The final condition for an MI system is that the system must be reasonably inexpensive so that every user can afford to run it in his home or office. MT systems are usually very expensive and installed only at major institutions. The user cannot bring a text to an institution each time he has a small job. Personal computers are therefore ideal for running MI systems.

To summarize, an MI system has the following tolerance. It can count on human assistance more than an MT system. However, it has 3 major constraints. It must not require its user to be bilingual or a specialist of any kinds, must respond in real time, and must be affordable for every user.

3. The Design Decisions

Assuming that the user knows only his own language, we can think of two translation directions; (1) foreign language to user's language, and (2) user's language to foreign language. Direction 1 is not suitable for MI systems, because it is difficult for the user to input foreign sentences without knowledge of the foreign language. We should therefore focus on direction 2; translating sentences in the user's language into a foreign language.

It is fairly easy to show that the interactive method is the most suitable human assistance method for an MI system that translates the user's language into a foreign language. The pre- or post-editing methods of human assistance are inappropriate in the following ways: To pre-edit an input text, the user must know the exact grammar and vocabulary the system can handle, and we have assumed that the user is neither a computer scientist nor linguist. To post-edit an output text, the user must know the foreign language.

In the interactive method, the user inputs a sentence in the user's language. The system may then ask the user several questions, and on receiving answers, the system finally outputs a foreign sentence. The final outputs does not require post-editing because ambiguity and other problems are solved by interaction with the user. The questions must be asked in the user's language, and must not require any knowledge of the foreign language, computer science or linguistics to answer them. A major concern of such an interactive system is the number of questions the system would ask the user. The toughest problem is perhaps syntactic ambiguity, which grows exponentially as sentence length grows [4]. A sentence can easily have over a thousand possible parses. Although an MI system can count on human assistance more than MT systems, the system would not be workable if it asks, say, 30 questions per sentence. In the following sections, however, the experiment shows that syntactic ambiguity can be resolved by asking at most a couple of questions per sentence, if the system has a little semantics.

4. The Experimental Interactive System

This section describes an experimental MI system [8] which we have built at Kyoto University by extending Nishida and Doshita's English-Japanese machine translation system† [6]. Our system asks its user questions to resolve input sentence ambiguity which cannot be resolved by the system itself. That is to say, when ambiguities are encountered, our system first tries to resolve them using relatively simple semantics, and only ambiguities which could not be resolved by the system are resolved by asking the user interactively.

Potential users of our system are English speakers, and the system is to be used to translate English into Japanese. That is, the user inputs an English sentence, the system asks the user questions in English whenever needed, the user answers those questions and finally the system produces a (probably not fluent but) correct Japanese sentence. In doing this we assumed the following conditions:

- The user does not know the target language (Japanese) at all.
- The user is neither a computer specialist nor a linguist.
- Human assistance is achieved only by an interaction, and no pre- or post-editing is required.

As discussed in the previous section, this kind of system is particularly useful for translating personal and small texts*. The experimental system has been implemented in Lisp on HITAC M-240H at Kyoto University. To implement sentence disambiguation by asking, we adopt the general algorithm described in Tomita [9] with some modifications.

We classify sentence ambiguity into 4 categories, as follows:

Type A: Ambiguity by multi-part-of-speech words

Type B: Ambiguity in conjunction scoping

Type C: Ambiguity in modification (such as PP attachment)

Type D: Ambiguity by general and specific pattern

To resolve each ambiguity, the system asks a question in a different manner.

† This work was done with Toyoaki Nishida and Shuji Doshita at Kyoto University, Japan. The author's role in the project was to modify the parser and to implement user interaction. All part of the English-Japanese machine translation system, except the user interaction module, was built only by Nishida and Doshita.

* Kay [5] argued that this kind of interactive machine translation systems could be practical for translating large documents as well, in the following two situations:

1. When the document is to be translated into several languages.
2. When the document is so technical that even professional translators can barely translate it.

4.1. Ambiguity by multi-part-of-speech words [Type A]

An example sentence with this kind of ambiguity is:

Time flies like an arrow.

Because the words, "time", "flies" and "like" are multi-part-of-speech words, several parse trees are generated. In some parse trees, the word "time" is a verb, and in other parse trees, it is a noun. The system asks questions as follows.

(Time) flies like an arrow
The word TIME is
1: verb
2: noun
NUMBER? 2

Time (flies) like an arrow
The word FLIES is
1: verb
2: noun
NUMBER? 1

The first line of each question is a copy of the input sentence, and the word being talked about is highlighted to get attention. Note that the system does not ask a question about the word "like", because the part of speech of the word can be uniquely determined after asking the two questions.

4.2. Ambiguity in Conjunction Scoping [Type B]

This kind of ambiguity occurs almost always when a sentence includes a conjunction such as "and". The example sentence is

I visited cities in Japan (and) Hawaii.

In this case, the system asks a question such as the following.

I visited cities in Japan (and) Hawaii
1: (cities) and (Hawaii)
2: (Japan) and (Hawaii)
NUMBER?

The system first reprints the input sentence, highlighting the word being talked about. Another example is shown below.

noisy boys (and) girls
1: (noisy boys) and (girls)
2: (boys) and (girls)
NUMBER?

4.3. Ambiguity in Modification [Type C]

This includes the problem of prepositional phrase attachment as in the sentence

I saw a man with a telescope.

The question to resolve this is:

I saw a man (with a telescope)

1: (a man) is (with a telescope)
2: The action (saw a man) takes place (with-a telescope)
NUMBER?

Another example with this kind of ambiguity is N-N compound attachment and the question to resolve it is:

(large) file equipment
1: (file equipment) is large
2: (file) is large
NUMBER?

4.4. Ambiguity by General and Specific Rules [Type D]

Suppose that the system has specific rules such as

VP --> 'compare' + NP_1 + 'with' + NP_2

in addition to the general rule

VP-> VP + PP.

Then 'with B' in the verb phrase

compare A with B

can modify 'compare' in two ways: "compare two objects A and B" using the specific rule, or "compare A by means of B" using the general rule. Since Japanese translations for these two interpretations are different, the system must resolve the ambiguity. The question the system asks to disambiguate the verb phrase above is

......compare A (with B) .....
1: (with B) is used idiomatically for the verb (compare)
2: (with B) is used not idiomatically
NUMBER?

5. Empirical Results

In this section, we describe the result of our experiment on how much interaction is required by our system to translate technical texts. We first define the number of questions. It is clearly not fair to count both a 2-alternative question and a 5-alternative question as 1 question. Thus, we define the number of questions of a n-alternative question to be n/2; that is, a 2-alternative question to be 1, a 3-alternative question to be 1.5, and so on. We think that this definition is a reasonable approximation†. As a sample text to be translated by the system, we take four abstracts of papers in the area of computer science. The whole text consists of 40 sentences and 841 words, and the average length of one sentence is 21 words, not including periods. The system's grammar and dictionary are well adopted to the text so that the system can handle all of the 40 sentences. One might suspect that this is unfair in the sense that the system would not perform equally well with sentences other than those 40 sentences. The purpose of the experiment is, however, not to evaluate the performance of a particular translation system. Our purpose is, rather, to estimate the number of questions of an interactive machine translation system in general, assuming that the system has a fairly comprehensive

† People, particularly information scientists, tend to think that $\log_2(n)$ is a better approximation. But this is not the case, because answering one 256-alternative question clearly requires much more effort than answering four 4-alternative questions.

grammar, dictionary and semantic information.

The result of the experiment with this text is as follows.

The number of questions
Total: 98
Per sentence: 2.45
Per word: 0.12

The result shows that, on average, the system asks roughly two to three questions per sentence. The relationship between the length of a sentence and the number of questions for the sentence is shown in the following.

L	S	N	A
0-9	2	1.0	0.0
10-14	10	2.0	0.7
15-19	5	14.0	2.6
20-24	12	16.8	3.4
24-	11	19.1	3.5

L: Length of sentence
S: The number of sample sentences
N: The number of parse trees generated per sentence
A: the number of questions per sentence

Table 1: Sentence Length and Number of Questions

As we can see in the table, the longer a sentence, the more questions are needed to disambiguate it.

Our sample text, by the way, is a quotation from actual publications, and therefore was written without considering the fact that the text is to be input to a machine. In contrast with this sample text, the input of an MI system, in general, tends to be composed by the user himself. Since the user knows that sentences he writes will be input to the system, it is expected that he does not input unnecessarily complex sentences. It is therefore interesting to do the same experiment with another sample text which is provided by a person who takes account of the fact that the text will be input to a machine.

We first gave the sample text used for the previous experiment to a native English speaker, and asked him to read and understand the text. We then asked him to rewrite the text, giving two comments as follows.

1. Your text will be input to a machine translation system, and very complex sentences may result in mis-translation.
2. Nevertheless, you do not have to feel restricted or to make special efforts.

As the result, our sample text was rewritten to 72 sentences and 878 words. The average length of a sentence in this new text is 12.2, which is almost 40% less than the original text. The result of our experiment with this new sample text is shown in the following.

The number of questions
Total: 64.5
Per sentence: 0.9
Per words: 0.07

As we can see, the system asks significantly fewer questions with the new sample text than with the original text in terms not only of the amount of interaction per sentence but also the total number of questions, despite the fact that both texts have the same semantic content.

Those data themselves are not definitive, because the result depends heavily on the grammar of the system. We can, however, conclude at least that an interactive machine translation system will not ask the user an unreasonable number of questions per sentence, if the system has a little semantic knowledge.

6. The Potential System

The experimental system we have described in the last sections has been built to try to show the feasibility of the interactive approach to machine translation, and it is far away from being called personal/interactive machine translation system. In this section, we discuss a potential personal/interactive system with a design philosophy described in section 2, that is, (1) It can count on human assistance more than MT systems, (2) It must be easy to use, (3) It must respond reasonably quickly and (4) It must run on low cost computers.

6.1. Amount of Human Assistance

People tend to think that interactive sentence disambiguation is not even feasible, because the number of parses out of a highly ambiguous sentence sometimes exceeds a thousand. However, a little semantics can reduce sentence ambiguity dramatically, as we have seen in the previous sections. Furthermore, not many people have realized that, even if a sentence has 1000 different parses, it can be disambiguated by asking as few as five 4-alternative questions. The total human effort required to answer questions may or may not exceed the total effort for a professional translator to translate the text without the system. However, in our new philosophy, this is a secondary concern, as long as short and personal texts are concerned.

6.2. User-friendliness

We do not think that the way the experimental system asks questions is particularly user-friendly. Nevertheless, taking the simplicity of the implementation into account, we think it is sufficiently acceptable at this stage. It does not require any knowledge of target language (Japanese) or computer science.

Our main concern at this point is, rather, handling ill-formed input sentences, a frequent problem not implemented in the experimental system. Input sentences for MT systems are provided usually from published texts, and we can assume that every sentence is well-formed (i.e. mistake-free). On the other hand, input sentences for MI systems are composed dynamically often at the terminal, and the user often inputs erratic sentences. The errors include misspelling, incorrect segmentation, missing constituents, spurious constituents and out of order constituents. Most of them are rather minor, careless mistakes. It is not acceptable for MI systems to simply reject a sentence each time the user makes such a minor mistake. The system must be able to handle ill-formed sentences robustly. Fortunately, a fairly good amount of work has been done to handle ill-formed input sentences, and summarized recently by Carbonell and Hayes [1].

6.3. Parsing Time Efficiency

The experimental system has not taken account of parsing time efficiency. It first produces all possible parses out of an ambiguous sentence and asks the user questions to choose the intended one. Theoretically speaking, this takes exponential time because the number of parses grows exponentially as the sentence length grows. In practice, the system takes a couple of minutes to parse a highly ambiguous sentence with hundreds of possible parses even on the large computer, although most of sentences can be parsed in seconds. On smaller computers, the problem would be much more serious. We therefore need a very efficient parsing algorithm that produces all possible parses in polynomial time. Note that such algorithm requires an efficient representation of all possible parses, so that the size of produced parses does not grow exponentially. Recently, such an efficient ail-paths parsing algorithm with an efficient packed-shared representation of alternate parse trees was introduced by Tomita [10, 11].

6.4. Economy.

On small computers, we cannot afford, at least at present, fancy operations utilizing, for example, real world knowledge, or a huge dictionary containing a million words. Our effort should therefore focus on how to solve natural language problems by asking without very fancy operations or huge data bases. For instance, when the user inputs a word which is not in the dictionary, the system can ask the user to use an alternative word. Thus, the size of the dictionary should be perhaps around 10,000 words, which is small enough to run on small computers and large enough to cover most of words so that the system does not ask the user very often due to unknown words. This kind of tradeoffs can be found everywhere in designing an MI system. To resolve referential ambiguity and word-sense ambiguity, as well as syntactic ambiguity, the system must have knowledge small enough to run on small computers but large enough to solve most of the problems by the system itself so that it does not ask the user questions very often to solve the problems.

7. Conclusion

As we have seen, the design philosophy of personal machine translation system must be totally different from that of conventional machine translation systems. (1) It can count on human assistance more than MT systems. However, (2) it must be user-friendly, (3) it must respond in real time, and (4) it must run on low cost computers.

As mentioned in subsection 6.1, people tend to think that interactive sentence disambiguation is not even feasible, because the number of parses out of a highly ambiguous sentence sometimes exceeds a thousand. However, the experiment has shown that the number of questions can be reduced remarkably if a system has little semantic knowledge. Furthermore, not many people have realized that, even if a sentence has a thousand different parses, it can be disambiguated by asking as few as five 4-alternative questions.

Finally, it has been suggested that we should focus more on man-machine interaction and parsing efficiency, which have not necessarily been essential for conventional systems.

ACKNOWLEDGEMENTS. I would like to thank Toyoaki Nishida, Shuji Doshita, Jaime Carbonell, Herb Simon, Jun-ichi Tsujii and Makoto Nagao for discussions that led to many useful ideas, and Cynthia Hibbard for helping to produce this document.

8. References.

[1] Carbonell, J. G. and Hayes, P. J.
Recovery Strategies for Parsing Extragrammatical Language.
Technical Report CMU-CS-84-107, Computer Science Department, Carnegie-Mellon University, Feb, 1984.

[2] Carbonell, J. G., Cullingford, R. E. and Gershman, A. V.
Steps toward Knowledge-Based Machine Translation.
IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-3(4), July, 1981.

[3] Carbonell, J. G. and Tomita, M.
New Approaches to Machine Translation.
In Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages. Colgate University, August, 1985.

[4] Church, K. and Paul, R.
Coping with Syntactic Ambiguity or How to Put the Block in the Box on the Table.
Technical Report MIT/LCS/TM-216, Lab. for Computer Science, Massachusetts

Institute of Technology, April, 1982.

[5] Kay, M.
Machine Translation.
American Journal of Computational Linguistics vol. 8(no.2):pp.74-78, April-June, 1982.

[6] Nishida, T. and Doshita, S.
Application of Montague Grammar to English-Japanese Machine Translation.
In Proceedings of Conference on Applied Natural Language Processing, pages 156-165. 1983.

[7] Saito, H. and Tomita, M.
On Automatic Composition of Stereotypic Documents in Foreign Languages.
Technical Report, Computer Science Department, Carnegie-Mellon University, 1985.

[8] Tomita, M., Nishida, T. and Doshita, S.
User Front-End for disambiguation in Interactive Machine Translation System.
In IPSJ Symposium on Natural Language Processing (in Japanese). Information Processing Society of Japan, 1984.

[9] Tomita, M.
Disambiguating Grammatically Ambiguous Sentences by Asking.
In Proceedings of 10th International Conference on Computational Linguistics (COLING84). 1984.

[10] Tomita, M.
An Efficient Context-free Parsing Algorithm for Natural Languages.
In 9th International Joint Conference on Artificial Intelligence (IJCAI85). August, 1985.

[11] Tomita, M.
An Efficient Context-free Parsing Algorithm for Natural Languages and Its Applications.
PhD thesis, Computer Science Department, Carnegie-Mellon University, May, 1985.