Interactive high-quality machine translation for monolinguals

D. Jones and J. Tsujii

Centre for Computational Linguistics,
UMIST,
Manchester,
UK.

ABSTRACT

This paper describes a machine translation system designed for use by monolinguals which guarantees high quality translations. A system has been developed for British Telecom which enables a user to compose texts in a foreign language with no knowledge of either the target language itself nor stylistic conventions pertinent to the text-type.

1. INTRODUCTION

This paper describes a machine translation system designed for use by monolinguals which guarantees high quality translations. A system has been developed for British Telecom which enables a user to compose texts specifically in a foreign language with no knowledge of either the target language itself nor stylistic conventions pertinent to the text-type.

Many conventional second generation machine translation systems are designed to work within a restricted domain but attempt to expand the coverage of the system as much as possible by increasing the size and coverage of the lexicon, grammar, and semantics. It is the thesis of this paper that this is an inappropriate methodology for certain contexts and that a different strategy should be adopted for situations where texts do not have to be static or 'given'. A 'given' source text is one where it is understood that neither the machine nor the human should modify the text in any way but should, instead, capture the intention of the writer as much as possible and transfer this information into a target language. Correctly interpreting the writer's intention is by no means straightforward so it makes sense to use machine translation in an environment where this information is freely available to the translator, i.e. during interaction with the computer.

2. REPRESENTATIONS

In order to achieve high quality machine translation output it is necessary to encode representative units of both source and target language equivalences in such a way that they are accessible given knowledge of user functional intent and surface expression thus:

[ω(expl,exp2) for [dc,con]]

where ω*(expl,exp2)* is a function from *expl* to *exp2* and *dc* and *con* are the corresponding functional description (see Dik, 1978) and context which hold for a particular equivalence relation. Equivalent expressions can be complete texts, paragraphs, sentences, or phrases. Appropriate expressions are identified by the user during one of two types of interaction with the system:

i) A menu-driven interface guides the user through a system of choice points eventually entering the contextual *(con)* description for a pair of equivalences. If the equivalences are of the level of text then further interaction will be required to satisfy any subcategorisation information in *dc* (see below).

ii) A user can freely enter expressions which are analysed and mapped onto the descriptions for any given functional equivalence. If there is a direct mapping then the target language equivalence is used. However, if there are any discrepancies the user is advised on any surface level modifications which may be required to bring the input and 'known' expressions up to parity. Some modifications may be made to *expl* although any changes will have repercussions in *exp2*. Such a modification may only occur if the *dc* specifies that *expl* and *exp2* are modifiable.

3. DEVELOPMENT

Work has been carried out at UMIST on equivalences at the level of complete text within the scope of a restricted text type. In order to build a system based on the representations described above, either a corpus of a given text type has to be examined in order to determine sublanguage regularities (Sager et al, 1987), or representative functional types can be used as a basis for composing language-pair specific text templates incorporating all the culture-specific stylistic detail essential for high quality translation (Saito and Tomita, 1986).

The current research involves the bringing together of these two approaches within one formalism whereby the same machine has the same type of data stored, but at different levels of scale. Functional representations can, therefore, be accessed and used for the analysis and synthesis of texts.

3.1. Pro-forma text composition

A basic system has been built where a database of 'pro-forma' business correspondence texts is used as data to allow monolinguals to compose letters of certain functional types e.g. *[complain], [offer], [enquiry],* etc. Such texts are templates with slots for phrases, names, dates, addresses which have to be entered by the user. Once a pro-forma text has been selected, question-answering interaction continues until all the slots are instantiated and the text is complete. The translation equivalence template (*exp2)* is used in order to produce the target language output. The pro-forma texts have been composed by bilingual experts in the field of business correspondence. They were composed on the basis of functional acceptability and sufficiency expressed by macro and micro functions where macro functions are understood to involve globally-orientated goals e.g. [offer] and micro functions the information-specific localization of macro functions e.g. what the writer is actually intends to write on a particular occasion. The monolingual user must be able to locate a text which will perform the macro function or functions appropriate to his/her need but it must also be sufficiently expressive of micro functions in order to accommodate specific and unpredictable input.

3.2. System Interface

At the beginning of an interactive session, the user knows what he/she wants to write but does not know how to express his/her intention in the current target language. On the other hand the system knows how to express a range of intentions in both the source and target languages but does not know the user's current intention. Obviously this state of affairs gives rise to the need for an interface where the system can guide the user to an appropriate pro-forma text i.e. a macro function equal to the communicative goal of the user. Each ω*(expl,exp2)* relation holds for a certain context This contextual information is important for accessing texts from high-level functional representations e.g.

ωl(expl,exp2) {dcl.conl} -> ω2(expl,exp2) {dc2,con2}

By utilising this type of information the system can deduce at any given time the inclusion of functions and, therefore, expressions which would be appropriate for those contexts. For instance, given the above representation, the system can determine that where *dcl* is:

$$COMPLAIN\ (X\ 1)_{agent}\ (X\ 2)_{goal}\ (X\ 3)_{recipient}$$

and *dc2* is:

$$not\_RECEIVE(X\ l)_{agent}(X2)_{goal}(X3)_{source}$$

an associated context for both is [REPLY , WRONG_ORDER ; FAULTY_GOODS]. 'REPLY' establishes that in this context there must be an existing relationship between writer and addressee and (,) that *expl* and *exp2* are semantically attuned to states of affairs where a wrong order has been received or (;) faulty goods have been delivered.

Using this contextual information, a user can be guided through a network of choice points by the system. At any point the system can determine the likely intention of the user until a pro-forma text or expression of a different level of representation, e.g. sentence, phrase, etc., is offered as a candidate.

Viewing the pro-forma text as a large template means smaller templates can be embedded into it using contextual information as a grammar of textual compositionality. As all expressions have contextual markers, any other expression can be selected for inclusion as long as it is marked for the appropriate context, *dc* information is then used in order to check for surface level expectations and requirements as a particular surface realisation of an expression may be inappropriate for a given context. Thus novel texts can be built up by embedding template expressions.

Alternatively, 'free input' slots may be provided where users can enter any phrasal or sentential expression they may wish. The system will attempt to accommodate this into the template if possible given the high degree of semantic and syntactic expectations which pertain in certain environments. If an equivalent target language expression can be generated for the *exp2* template then the input is accepted, otherwise a conventional transfer grammar can be used to achieve a local translation. Further work is required in this area.

The role of the user and the status of the source text need not be regarded in the way described above if an environment is chosen where the source text is not given. In other words, if what the system has to translate is not fixed then the system has the opportunity to dynamically modify the input in such a way that users are satisfied that the source text is sufficient for their purposes and that the machine can produce translations of a high quality. Thus, the role of the user is transformed from one of manual fail-safe override to a dynamic interactive role with the system.

The human user, in such an environment, will have an overall goal which will be to compose some source language text. The goal of the system will be to translate that text to the best of its ability and to do so will involve rejecting any constructions the user generates which it cannot definitely translate given its knowledge of the source language, the domain of usage and its associated current expectations. The system will only accept source text input once there has been an 'agreement' between the user and itself that any modifications required to the input are acceptable.

3.3. Environments

Suitable environments for such a methodology are themselves naturally restricted in the sense that such a machine translation system requires an environment where it would not be unnatural to compose a source text in a restricted enough domain to allow for useful encoding of representative linguistic knowledge.

Business correspondence is a good example of a suitable environment as texts are composed on a 'throw away' basis. This means that such texts are composed quickly for a particular function which is relatively easy to discern. Once composed, delivered, received and read their function has been performed. A technical reference manual, on the other hand, is different. It is written over a much longer period of time with the aim of being read by an indefinite number of people over an indefinite period. One does not write something the size of a reference manual in five minutes and, therefore, the compo-

sition environment will be quite different.

It remains to be seen to what extent such techniques can be used for both the composition and translation of other restricted text types of a lengthier variety such as a manual, both during the composition stage, as well as for the translation of texts once they have been written. One possible application would be that of a translator-summarizer. As the system holds expressions defined in terms of possible deep surface realizations and contexts, if the machine could map these representations onto 'given' texts this would be comparable to a closest match interpreter. Again, though, this is an area for further research.

4. SUMMARY

A different type of machine translation is envisaged where a system is not restricted to attempting to translate material when the information required to do so is quite possibly underspecified. In certain environments it is possible for machine and user to reach a compromise on surface level expressiveness for high quality translation along a functional axis. Such a machine translation system has the ability to demonstrate what it can translate rather than what it cannot.

A system has been built whereby a monolingual user can compose and translate restricted text types with guaranteed high quality even at the level of sublanguage-specific stylistic conventions. It is noted that in certain environments of text composition it is quite natural to elicit information from the user as to functional intent as well as to the appropriateness of choosing expressions which the machine suggests as functionally but not syntactically equivalent to one which the user initially intended. Under such circumstances the user knows that there is a pay-off between absolute freedom of expression and high quality translation. On the one hand the user is limited in the detail of what users can say but on the other they know they are producing high quality translations. In the simplest type of system the degree of restriction on the user is the greatest although more sophisticated analysis and synthesis techniques can be used in association with *dc* and *con* knowledge in order to allow for guided text composition/translation and dynamic augmentation of equivalence templates with either additional templates or 'free' surface-level input.

References

Dik, S.C., *Functional Grammar.* North-Holland, 1978.

R Saito and M. Tomita, *On Automatic Composition of Stereotypic Documents in Foreign Languages.* Research Report CMU-CS-86-107, Computer Science Department, Carnegie-Mellon University, December 1986.

Sager, N., *Medical language processing: computer management of narrative data.* Addison-Wesley, 1986.