[From: IFTT'89, April 26-28, Oiso Prince Hotel, Japan]

Controlled Language

Hiroshi Uchida
Fujitsu Laboratories Ltd.

In Japan, more than 10 machine translation systems are developed and marketed now. Several of these systems are actually used commercially. However, present machine translation systems can not produce perfect results. The main reason is that machine translation systems use very little information about input sentences, compared with the information humans use when they understand or translates these sentences. For example, most machine translation systems translate sentence by sentence without using contextual information, and the knowledge bases used in semantic analysis are very poor.

However, this does not mean that we cannot use present machine translation systems for practical use. Present machine translation systems have enough capability for practical use, when we use them skilfully. And to use controlled language is one of such skills.

If there is a ellipsis of the subject or the object, postpositions, the ending of words in a sentence, a computer which translates sentence by sentence cannot understand the sentence. Therefore, we must add information such as contextual information to the computer from the human side, and also restrict the usage of language for the computer to understand sentences and get the good results from the computer. If we make such a language control, both human side and computer side will be happy. This is the main objective of language control.

Several controlled languages for English have been proposed. Basic English, proposed by C.K.Ogden in 1936, has a restricted vocabulary of 12 verbs and about 800 other words. But it was designed to use combinations of verbs to represent various meanings. Caterpillar Fundamental English was proposed in 1972, but this controlled language was also for human readers, not for computers.

In Japan, several language control methods have been proposed, for example Japanese control language by Yosida of Kyuushuu University in 1979, and a machine adaptable language by Nagao of Kyoto University in 1983. Both controlled languages were for computers to understand sentences. Also, language control methods for special machine translation systems were proposed by Fujitsu for ATLAS, Hitachi for HICATS and so forth.

1.Controlled language for machine translation

We need two types of language control method to use machine translation systems.

1) Weakly controlled language to get better results from machine translation
2) Strongly controlled language to get perfect results from machine translation

Weakly controlled language is still a natural language, and humans can read sentences of this controlled language naturally. So this controlled language is helpful for human translators as well as machine translation systems. This kind of control language is used when translators use machine translation systems as a translation tool.

Strongly controlled language is an artificial language to some extent. This controlled language is used by a person who does not know the target language and wants to get nearly perfect results from a machine translation system without the help of human translators. Sentences of this controlled language are not natural language, but humans can read it easily.

Both controlled languages should restrict the syntax of language rather than vocabulary. The reason is that a machine translation system can handle a large vocabulary, so when the syntactic structure of a sentence is clear, the system can interpret the meanings of words using its knowledge base. This is the main characteristic of controlled languages for machine translation systems, while controlled languages for human readers mainly restricts vocabulary.

2. Weakly controlled language

The main purpose of weakly controlled language is to make the syntactic structure of the sentence clear. There are no semantic restrictions on word usage to lesson restrictions on a natural language.

The main restrictions on syntactic structure are as follows.

1) Write a simple sentence principally,
but there are many exceptions. Some complex sentences which have clear syntactic structure is permitted.
2) Modify-modified relations must occur between nearest component of the sentence.
3) Coordinate clauses must be at the same level
4) etc.

This kind of language control method is rather the guidelines for writer. This kind of controlled language is developed by JEIDA.

3. Strongly controlled language

In addition to the restrictions of the weakly controlled language, the strongly controlled language has the following restrictions and conventions.

1) Usage of [ ] to clarify the syntactic structure.
[ ] means some sentence component.
ex. I [saw a bird] with telescope.
I saw [a bird with telescope].
2) denotation of part of speech
3) replacement of the word
substitute a less ambiguous word which has the same meanings

These kinds of controlled language are proposed by Fujitsu, Hitachi, etc..

4. Small Japanese

The Small Japanese is a controlled Japanese proposed by the Machine Translation System Committee of JEIDA. This controlled language is a weakly controlled language and is designed as guideline for original writers of documents rather than rules for pre-editors of existing documents. The basic concept of controlled Japanese are the following.

1) Restrictions are mainly for clarifying syntactic structure of sentences.
2) No restriction for the semantic usage of words.
3) Prohibition of omission of information which cannot be extracted from sentences not using contextual information or background knowledge.

General rules of Small Japanese are as follows.

1) Write simple sentences.
2) Phrases which modify verbs, adjectives, and adjectival verbs must be positioned near the thing modified.
3) Place "rentai" modifying phrases nearest the thing modified.
4) Attach the postpositions 'to',' ka','ya','toka' in between and at the end of coordinated sentence elements(noun phrases).
5) Do not omit predicate endings or case postpositions.
6) Write Chinese words in kanji.
7) Write numbers in arabic numerals.
8) Enclose examples in parentheses.
9) Avoid literary expressions and verbosity.
10) Language of quotations, thought, and commands should be enclosed in 「 」.
11) Enclose formulas, variables, and strings in 「 」.
12) When necessary, divide loanwords by ".".

5. Small English

Small English is a controlled English also proposed by the Machine Translation System Committee of JEIDA. This controlled language is also a weakly controlled language and is designed as a guideline for writers rather than rules for pre-editors of existing documents. The basic concepts are identical to Small Japanese.

General rules of Small English are as follows.

1) Compose sentences consisting of no more than two clauses.
2) Do not omit syntactically necessary elements such as the verb, subject, or object, the 'to' in a to-infinitive clause, the subordinator 'that' relative pronouns which do not include the antecedent noun.
3) Choose a word order which makes the syntactic structure clear. Do not use inversion or insertion. Place prepositional phrases and adverbials which can move to the front of the sentence, at the front.
4) Do not use participial clauses other than adverb + participle.
5) Do not use independent, non-finite clauses which cannot be used idiomatically.

4. Conclusion

Small Japanese and Small English were evaluated with the following method: existing documents were pre-edited into Small Japanese and Small English, and translated with machine translation systems. Then the efficiency of post-editing with and without the pre-editing was compared. The result showed that post-editing for pre-edited documents was considerably faster than for non pre-edited documents, but the time saving was less than the time required for pre-editing. This shows that documents should be written from the start in Small Japanese and Small English.

The ability of computers to understand natural language is far lower than that of humans. When we talk to small children, we use simple sentence and simple words. The situation is the same when we talk to computer. That is why we need controlled language for machine translation systems.