[*Machine Translation Today: Translating and the Computer 15*. Papers at a conference…18-19 November 1993 (London: Aslib)]

CONTROLLED ENGLISH WITH AND WITHOUT MACHINE TRANSLATION

Arthur Lee
Bull S.A.. 7. rue Ampère, 91343 MASSY Cedex, France

What is Controlled English? Its use and application.
The rules of Controlled English and their meaning.
Max - A Controlled English tool.
The Controlled English dictionaries.
Implementation of Controlled English within Bull.
Guidelines for implementing a Controlled Language system with or without Machine Translation.

INTRODUCTION

Although I am going to talk about Controlled English, in particular as used in Bull, most of my presentation applies to any language. In fact Controlled French is being developed at Bull along very similar principles. Controlled English was introduced in Bull by ILO (Internationalization and Localization of the Offer). This department also introduced Machine Translation to Bull, thus Controlled English and Machine Translation were, from the start, closely connected, although now the importance of Controlled English is seen to be much wider.

I want to answer three questions. What is Controlled English? Where should we use it? Why should we use it? I shall also explain the development of Controlled English within my own company and the rules and techniques which we have developed.

1. Controlled English is a set of grammar and style rules imposing simple structures on written English, plus a restricted vocabulary.

2. Controlled English is primarily for use in technical documentation. This covers all fields: electronic, computing, medical, scientific, etc. It is not designed for literature, poetry, love letters or conference papers for linguists.

3. There are several reasons why Controlled English should be used.

- In an international marketplace, many users of documentation have English as their second, or even third, language. Technical documentation should therefore use simple grammatical structures which are easy to understand.

- Controlled English provides a coherence of style and vocabulary throughout a manufacturer's documentation, thus avoiding jargon which may have different meanings in different documents or, inversely, different terms for the same concept.
  Recently, particularly in the field of computers, technology has become available to a wider public. The end

user is not necessarily a well educated, highly literate person and requires a simplified English in order to fully understand instructions given. Recently, I read an article in the French journal "Communication et langages" which discussed user instructions for domestic appliances, pointing out that many washing machines now have fifteen programs, but very few people use more than two because they do not understand the instructions. When they buy another machine they change manufacturer because they are not satisfied with the whiteness of their wash! The use of a Controlled Language therefore has a clear long-term economic impact on the manufacturer of any product which includes instructions on its use.

- Texts written in Controlled English are easier to translate by machine, owing to the lack of ambiguity, a simplified grammar and a known vocabulary.

- Easier machine translation results in lower publication and time-to-market costs.

- Standard terminology throughout a manufacturer's documentation makes the user's life easier and increases customer satisfaction.

- Documentation is easier to update, again reducing time and cost.

- Controlled English gives access to a wider global market. This leads to increased sales.

THE BULL CONTROLLED ENGLISH RULES

Bull Controlled English has ten rules. These are:

1. Make positive statements: avoid the passive voice; avoid the future tense.
2. Keep sentence length to a maximum of 25 words.
3. Use valid terminology; do not invent it. Use the Controlled English vocabulary.
4. One thought per sentence.
5. Use simple sentence structures.
6. Use parallel construction.
7. Avoid conditional tenses.
8. Avoid abbreviations and colloquialisms.
9. Use correct punctuation.
10. Use the tools available (Max. Grammar Checker, Spelling Checker).

What the Rules Mean

The first rule on passive and future tenses is about a basic element of style in technical writing. In general, the Technical Author should be living in a permanent, active, present. Rather than "an action is performed by the machine", we use "the machine performs an action". In using the future, we go into the realm of the unknown. The Author is not clairvoyant. When we say "the effect will be", this depends on the product remaining stable. For an electric toaster, which never has user upgrades, this is fine, but for aircraft, computers, military systems etc. we do not know if what is described now will always be the case. One can say that this is rather pedantic, as product upgrades imply documentation revision, but the positive present tense does clarify for later Authors what the product *actually* does.

I have often heard Technical Authors complain about the second rule. There is only one reason why sentences should be long, and that is when the Author does not understand the subject. It is extremely difficult to hide a lack of knowledge in short sentences and there is quite an art in writing long, complex sentences which give the same choice of interpretation that the Author has. There is also quite an art in translating such text so as to be equally vague in the target language.

In any field where a Controlled Language is used, it is vital that the terminology is correctly defined before asking Technical Authors to use this writing technique. Without this preliminary work, the third rule is meaningless. There must be a technical and general dictionary supplied to the Author. I have come across terms which I could not find in a standard office dictionary. How does a user with English as his second language going to understand?

The fourth rule on "one thought per sentence" and, as I say in my training courses, at least one thought, closely linked to the next rule. It leads to simple sentence structures and sentences which are easy for the user to follow.

The fifth rule on simple sentence structures allows for three basic structures: the statement, description or explanation, the step and action, and cause and effect. To take the example of getting a drink from a machine we can say. "Select the drink you want after putting the money in the machine. Take the cup when the machine indicates that the drink is ready." In Controlled English, we would say. "Put the money in the machine. Select your drink. When the machine indicates that the drink is ready, remove the cup." The Controlled English version follows a logical sequence of single steps.

The sixth rule on parallel construction is obvious to many foreigners, but not to native English speakers. In everyday English we quite often change tense in mid-sentence. This is not possible in French, for example. Sentences with mixed tenses are difficult for Machine Translation systems.

The seventh rule on conditional tenses is particularly important for clarity and safety. The word 'should' implies a choice. Imagine for a moment the effect of using 'should' throughout an aircraft maintenance manual! Similarly, the word 'may' should be avoided. I recently saw "This may lead to unpredictable results." So even the unpredictability was unpredictable. The Author should use 'must', 'can' and the present tense, for example "This leads to unpredictable results."

Rule eight, on abbreviations and colloquialisms, concerns clarity of the text. Abbreviations often have several meanings. Of course it is not always possible to give the expansion of abbreviations throughout the text and maintain readability. Every document should have an appendix listing all the abbreviations used and their meaning. Colloquialisms such as 'O.K' or 'power up' are vague. The Author should always be precise in values, conditions and actions.

Rule nine is self evident. Incorrect punctuation can change the sense of a sentence to the reader and will often confuse a Machine Translation system.

Rule ten is important for Machine Translation. The ideal is to use a Controlled English tool, ensuring that the vocabulary used is correct. If not. many grammar checkers will pick up the structural aspects of Controlled English. Spelling mistakes can lead to mistranslation by Machine Translation systems.

MAX - A CONTROLLED ENGLISH TOOL

At Bull, we use a tool for implementing Controlled English where the text is to be translated. This tool is called Max, and was produced by Smart Communications of New York. It comes in two forms: batch and interactive.

What does Max do?

Max applies the first nine rules of Bull Controlled English. It checks grammar, sentence structure, punctuation and vocabulary. Max uses three dictionaries to control vocabulary.

The synonym dictionary. This is the reverse of a normal synonym dictionary. Whereas normally we go from one word to many words, the object of the Max synonym dictionary' is to pick up a wide range of words and phrases with the same meaning and to propose a small number of words and phrases to replace them. The primary entries do not appear in the other dictionaries, but all the proposed alternatives appear in one of the other dictionaries. The basic synonym dictionary was provided by Smart Communications and refined by me for Bull's use.

The general dictionary. This dictionary contains everyday words of the English language. The general dictionary was supplied by Smart Communications and modified by me for Bull's use.

The technical dictionary. Max is supplied with a range of technical dictionaries for various domains. However our technical dictionary used by Bull was generated entirely by ILO.

IMPLEMENTATION OF BULL CONTROLLED ENGLISH

The first stage

Controlled English was first used in Bull in ILO. Max was used in batch mode for pre-editing text before translation. The resulting text was never intended to be read by human beings, the Controlled English produced being adapted specifically for the translation engine.

The second stage

The next stage was to implement Controlled English with the Technical Authors. By this time ILO had developed a multi-lingual terminology database. The foundation of this database was the English terminology, supplied with the codes necessary for Max. The translation of the database was then performed for our initial target languages (Dutch. French, German, Italian and Spanish). The database was used to produce the technical dictionary for Max and the dictionaries for the translation engine. Next we put the general dictionary into the database and created the entries for the target languages. If we change our translation system, we will be able to generate the technical and general dictionaries for the new system from our existing database. In this way, all technical terms and general vocabulary used by the Authors are known by the translation system.

Technical Authors and Machine Translation. The Authors use the interactive version of Max. They write a paragraph, then "Max" it. They are not obliged to follow the rules, but should rarely deviate from them. The text files arrive at ILO with the associated error files. Here, in the pre-editing stage, we can see if unknown terms have been used. In a fast moving industry new terminology is being created daily, and by spotting new terms immediately they are used, we can update our database and translation dictionaries, and reissue the Max technical dictionary. Obviously, not all terms used which are not found in the dictionaries are valid new terms. These will be modified at the pre-editing stage. If a particular Author is consistently failing to apply the rules and use the correct vocabulary, we can follow the problem up with the Documentation Department.

Max provides a temporary dictionary. This can be used by the Author for terms which are not defined in Max, but which he uses frequently. We can use this temporary dictionary as input to update our terminology database.

The third stage

The final stage, which is currently being implemented, is Controlled English throughout the company. For the second stage (Technical Authors) a manual was written explaining the rules of Controlled English and also providing guidelines on style. This manual, along with paper versions of the Max technical and general dictionaries, is available for everyone writing internal or user documentation, whether or not it is to be translated. The eventual aim is to provide Max to everyone. The advantage in using Controlled English for internal documentation is that many originators and readers of internal documentation are not native English speakers. We are also looking into an English language editor designed specifically for French speakers.

Guidelines for Implementing a Controlled Language System with or without Machine Translation.

These guidelines are based on the experience of Bull and will help you to implement a Controlled Language system which is easy to use and can evolve with your activities.

Basic requirements

The first step is decide on your rules. Controlled Languages systems follow more or less the rules I have described, with language specific variations and differences according to the method of working. For example, if the Author reads his translations, include the review of translated text with the original English, in order to gain experience of which forms are easier to translate.

The second step is to create your technical and general dictionaries. You can have one single technical dictionary or individual dictionaries for different domains. Controlled Language systems *must* have dictionaries with the controlled vocabulary.

Machine Translation

Make sure that your translation system knows the Controlled Language vocabulary. Our experience of building a terminology database starting with the source language (in our case, English) has proved invaluable in maintaining coherence at every stage in document production, from writing through pre-edition, translation, post-edition and revision. Coherence of the Controlled Language and translation dictionaries saves time and money.

Conclusion

I hope that in this brief overview of Controlled Language, with specific reference to English, I have managed to demonstrate the advantages of Controlled Language systems, not only from an economic point of view, but also as a tool for easier communication in our global village.