

2010

Controlled language - Does my company need it?

Uwe Muegge

Controlled Language: Does My Company Need It?

By Uwe Muegge


(The following originally appeared in the April 2009 issue of *tcworld* and the July 2009 issue of *Language Tech News*, the newsletter of ATA's Language Technology Division, www.ata-divisions.org/LTD.)

Controlled languages use basic writing rules to simplify sentence structure. Here is how they work and how your company can benefit from introducing a controlled language.

What is Controlled Language?

A controlled language is a natural language, as opposed to an artificial or constructed language. Natural languages such as English or German are languages that are used for general communication. A controlled language differs from a general language in two significant ways:

1. The grammar rules of a controlled language are typically more restrictive than those of a general language.
2. The vocabulary of a controlled language typically contains only a fraction of the words that are permissible in a general language.

As a result, authors who use a controlled language have fewer choices available when writing a text. For example, the sentence "Check 

the spelling of a paper before publishing it” is a perfectly acceptable sentence in general English. Using CLOUT, a controlled language rule set I developed, the sample sentence would have to be rewritten as “You must check the spelling of your document before you publish that document” to comply with rules regarding vocabulary, active voice, use of articles, and avoidance of pronouns.

Why Do We Need Controlled Languages?

Facilitating language learning: When C.K. Ogden developed Basic English in 1930—probably the first controlled language—he had the explicit goal of dramatically reducing the five-plus years it is estimated to take to master Standard English.¹ Based on a vocabulary that contains 850 essential words (for comparison, the *Oxford English Dictionary* defines more than 600,000 words), Basic English is designed to be acquired in a few weeks only.²

Eliminating the need for frequent translation: One of the most widely used controlled languages today is ASD-STE100 Simplified Technical English, also known as Simplified English. Simplified English was originally developed by the European Association of Aerospace Manufacturers (AEEMA) in the 1980s.³ The main purpose of Simplified English was to create a variant of Standard English that aircraft engineers with only a limited command of English could understand, thereby eliminating the need to translate maintenance manuals into foreign languages.

Streamlining translation: Within the localization industry, many people familiar with the controlled language

It typically comes as a surprise that controlled languages can and have been used for purposes other than making the translation process more efficient.

concept associate it with automating the translation process. In fact, it typically comes as a surprise that controlled languages can and have been used for purposes other than making the translation process more efficient. By restricting both vocabulary and style, using a controlled language typically improves match rates in translation memory environments and translation quality in rules-based machine translation environments.

Enhancing comprehensibility: Helping authors avoid both semantic and syntactic ambiguity has been recognized as a goal worth pursuing in and by itself, especially in the domain of technical communication. Some organizations are deploying a controlled language for the sole purpose of improving the user experience of a product or service on the domestic market.

Common Features

A common characteristic of controlled languages is the fact that very little information about their rule sets and vocabularies is freely available. This is not really surprising considering that a controlled language can provide an organization with a distinct advantage over its competitors.

Apart from that, you will find very few similarities between controlled languages. Nortel Standard English,

for instance, has only a little over a dozen rules, while Caterpillar Technical English consists of more than 10 times as many. A recent comparative analysis of eight controlled English languages found that the number of shared features was exactly one: a preference for short sentences.⁴

Why Should My Organization Use a Controlled Language?

Here is why it can make sense for your company to use a controlled language:

Improved usability: Documents that are more readable and more comprehensible improve the usability of a product or service and reduce the number of support calls.

Objective metrics and author support: Tools-driven controlled language environments enable the automation of many editing tasks and provide objective quality metrics for the authoring process. Controlled-language environments also provide authors with powerful tools that give them objective and structured support in an environment that is typically rather subjective and unstructured.

Lower translation costs: As controlled-language texts are more uniform and standardized than uncontrolled ones, controlled-

Examples of Organizations That Have Created A Controlled Language

language source documents typically have higher match rates when processed in a translation memory system. Higher match rates mean lower translation cost and higher translation speed.

Some controlled languages have been specifically designed with machine translation in mind, (e.g., Caterpillar Technical English and CLOUT). Using a controlled language customized for a specific machine translation system will significantly improve the quality of machine-generated translation proposals and dramatically reduce the time and cost associated with a human translator or editor.

Impact on Translation

Even in environments that combine content management systems with translation memory technology, the percentage of untranslated segments can still remain fairly high in new projects. This can be a major challenge for organizations that wish to reduce the cost and time involved in the translation of their material. While it is certainly possible to manage content on the sentence/segment level, the current best practice seems to be to create chunks of content on a particular topic that can be reused in other documents (as opposed to creating new content for each new document). This means that reuse occurs at a fairly high level of granularity. In other words: there is too much variability within these topics for a high match rate.

Controlled Authoring for Translation Memory Systems

An effective way of optimizing a globalization environment that is based on translation memory technology is to normalize the source that

Alcatel: Controlled English Grammar (COGRAM)
Avaya: Avaya Controlled English (ACE)
Caterpillar: Caterpillar Technical English (CTE),
Caterpillar Fundamental English (CFE)
Dassault Aerospace: Français Rationalisé
Ericsson: Ericsson English
General Motors (GM): Controlled Automotive
Service Language (CASL)
IBM: Easy English
Kodak: International Service Language
Nortel: Nortel Standard English (NSE)
Océ: Controlled English
Siemens: Siemens DokumentationsDeutsch
Scania: Scania Swedish.
Sun Microsystems: Sun Controlled English
Xerox: Xerox Multilingual Customized English



Some organizations are deploying a controlled language for the sole purpose of improving the user experience of a product or service on the domestic market.

feeds the translation memory tool. Normalizing the source means reducing variation between sentences. Writing in a controlled language reduces variability by limiting the

choices available to authors. This is especially true if the controlled language not only covers grammar, style, and vocabulary, but also text function.⁵ In a functional approach to

controlled language authoring, there are specific rules for text functions, such as instructions, results, and warning messages, that can be used for training a machine translation system. The rules are relevant to the structure of the document, the syntax, semantics, and the lexicon.

Here are two simple examples for functional controlled language rules:

Text function: Instructions

Pattern: Verb (infinitive) + article + object + punctuation mark.

Example: Click the button.

Text function: Results

Pattern: Article + object + verb (present tense) + punctuation mark.

Example: The window “Expense Report” appears.

Implementing functional controlled language rules will enable authors to write text in which sentences with the same function have a very high degree of similarity. This not only makes sentence modules reusable within and across topics in a content management system, it also dramatically improves match rates in a translation memory system.

Controlled Authoring for Rule-based Machine Translation Systems

While uniformity is the decisive factor in improving efficiency in a traditional translation memory environment, reducing ambiguity in the source text makes machine translation more productive. The problem that rule-based machine translation systems like Systran struggle with is the fact that in uncontrolled source texts, the (grammatical) relationship between the words in a sentence is not always clear. To enable rule-based machine translation systems to generate better transla-

A controlled language can provide an organization with a distinct advantage over its competitors.

tions, the controlled language needs to have rules like the following that help the machine translation system successfully identify the part of speech of each word in a sentence.

Write sentences that have articles before nouns, where possible.

- Do not write: Click button to launch program.
- Write: Click the button to launch the program.

Write sentences that repeat the noun instead of writing a pronoun.

- Do not write: The button expands into a window when you click it.
- Write: The button expands into a window when you click the button.

With rules in place that mitigate the weaknesses of rule-based machine translation systems, the quality of their output is bound to improve dramatically.

Do I Have to Develop My Own Controlled Language?

Not at all! Today, many organizations that wish to reap the benefits of controlled-language authoring opt for a software-driven solution that comes with a built-in set of grammar and style rules. Systems like acrolinx IQ Suite, IAI CLAT, or Tedopres HyperSTE have enabled thousands of organizations to improve the quality and productivity of their authoring and translation processes. In a soft-

ware-driven authoring environment, organizations do not have to maintain the staff of highly trained linguistic experts needed to develop and deploy a proprietary controlled language. Instead, the organization simply selects the rules that are most suitable for a given content type from a set of preexisting writing rules. Typically, these checking tools support the definition of multiple sets of rules for multiple types of content (e.g., stricter rules for user documentation than for knowledge based articles).

From a technology standpoint, it is relatively easy to implement the rules part of a controlled language; the terminology part is typically more labor intensive. It is certainly true that many controlled language software solutions include a module for collecting terminology, but still the task of creating a corporate dictionary, which is what this job amounts to, might be a daunting one. Not only will all synonyms among the possibly thousands of terms in use within the organization have to be identified, but these synonyms will also have to be categorized into preferred and deprecated (do not use) terms. While creating a corporate dictionary may be a challenge, once it is available, that dictionary may also be the feature most valued by the users of the controlled language system.

Examples of Controlled Language

For more information on controlled languages and to see a demonstration of

a simple controlled language designed for machine translation, visit www.muegge.cc. The entire site was written in CLOUT, the Controlled Language Optimized for Machine Translation. On the home page, click on any of the language combinations into English (e.g., German-English or French-English) and watch how a free machine translation system turns a complete website into a fully navigable, highly comprehensible virtual English site in real time. Click on the link Controlled Language/Rules for

Machine to see 10 sample CLOUT writing rules that have a high impact on the comprehensibility and (machine) translatability of instructional text in English.

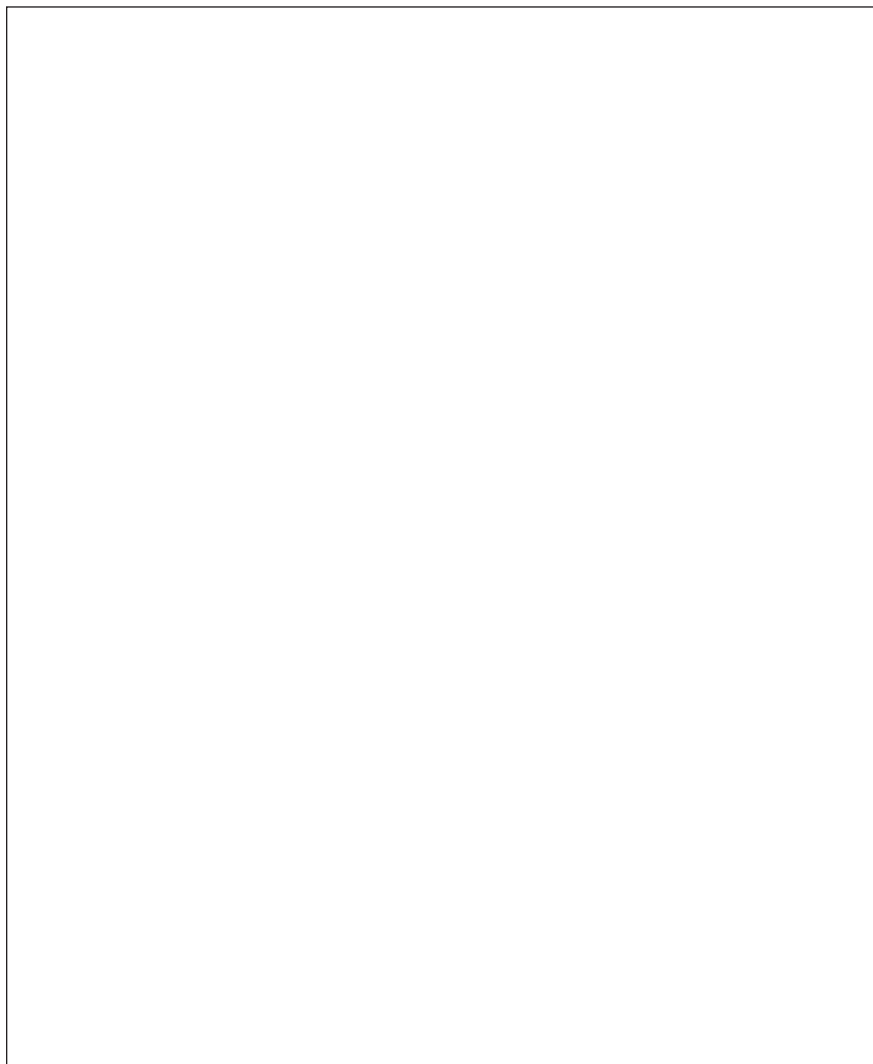
References

1. Ogden, Charles Kay. *Basic English: A General Introduction with Rules and Grammar* (London: Treber, 1930).
2. Ogden's Basic English Word List. Ogden's Basic English (Basic

English Institute, 1996), [Online] <http://ogden.basic-english.org/words.html> (Accessed February 3, 2009).

3. ASD-STE100 - Simplified Technical English - International Specification for the Preparation of Maintenance Documentation in a Controlled Language. ESSAS Electronic Supporting System for ASD Standardization (AeroSpace and Defence Industries Association of Europe, 2005), www.asd-stan.org/sales/asdocs.asp (Accessed February 3, 2009).
4. O'Brien, Sharon. 2003. "Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets" (Machine Translation Archive, 2003), www.mt-archive.info/CLT-2003-Obrien.pdf (Accessed February 3, 2009).
5. One example of a well-developed authoring rules set based on a functional approach is Funktionsdesign® (functional design), developed by professors Robert Schäflein-Armbruster and Jürgen Muthig.

ata



STAR	
www.star-group.net	WebTerm
Global and collaborative terminology management	