October 15, 2013

Implementing a controlled language is now cheaper and easier than ever

Uwe Muegge

Implementing a controlled language is now cheaper and easier than ever!

The controlled language (CL) model is not new: Caterpillar Fundamental English was rolled-out in 1972, and many other enterprise- and a few industry-level controlled languages followed since then. What is new is the fact that now complete commercial rule sets and powerful CL tools are available for free, making controlled language authoring attractive for even the smallest organizations.

By Uwe Muegge

Why use a controlled language?

Improving the readability of technical texts
The first controlled language, C.K. Ogden's Basic English, consisted of 850 words and was published in 1930 with the goal of reducing the time needed to learn English. Modern controlled languages typically have two components:
- Grammar rules that are more restrictive than the rules of the general language;
- A vocabulary that contains only a fraction of the words of the general language.

These types of restrictions help authors avoid writing long, convoluted sentences and using uncommon words, all of which are arch enemies of easy readability and comprehension.

Supporting the translation process
When the European Aerospace Manufacturers Association (AECMA) introduced Simplified Technical English, the primary goal was not to simplify translation, but to completely eliminate it. How? By providing authors of aircraft maintenance manuals, which AECMA members were primarily concerned about, with a form of English that was specifically designed for an audience of non-native speakers of English. In many markets the documentation for a wide variety of products and services is subject to strict translation requirements. Using a controlled language for the creation of technical documentation typically supports the translation process in a number of ways:
- The use of a restricted vocabulary typically reduces the number of terminology-related queries from translators, which speeds up the translation process.
- The use of a restrictive rule set typically improves the uniformity of source texts across documents, which helps leverage existing translations, thereby reducing translation cost.
- The use of a restrictive style guide typically reduces the number of words in source texts, which reduces translation cost and speeds up the translation process.

One of the most promising applications of controlled language is using Do-it-yourself statistical machine translation service (DIY SMT) for translating controlled language content. Once a statistical machine translation system has been customized with a controlled language corpus, this type of MT system should provide superior translation quality. The main benefit of the new DIY machine translation services is that these new services make customization very easy and inexpensive. Once an organization has approx. 10,000 sentences of translated controlled language content, the transition to using statistical machine translation is an easy one with a DIY SMT service.

Supporting the writing process
'Speaking with one voice' is a common challenge many organizations have that work with teams of technical writers. 'Speaking with one voice' refers to the ability of every member in a team of writers to use the same style and vocabulary in the texts a writer creates. Providing technical writers with a shared set of writing rules and a shared vocabulary helps teams of writers create texts that are more uniform and require less editing. In fact, with a comprehensive set of writing rules in place, it is possible to automate certain aspects of the editing process using a controlled language checker.

Controlled language checkers, also known as style checkers, evaluate the compliance of texts with the grammar rules and the vocabulary rules of a controlled language. Below is a list of commercial controlled language checkers:

acrolinx

Content Optimization Software

Your Content,
Only Better

Acrolinx features:

❯ **Spell and grammar checking**

❯ **Style and terminology checking**

❯ **Multilingual terminology management**

❯ **Reusable text modules**

❯ **User-friendly operation**

Benefits for your organization:

❯ **Reduced translation costs**

❯ **Accurate and consistent content**

❯ **Search Engine Optimization**

❯ **Automated reporting**

contact@acrolinx.com

- Acrolinx by Acrolinx
  (www.acrolinx.com)
- CLAT by IAI Saarbrücken
  (owl.li/nvU2f)
- HyperSTE by Tedopres
  (www.simplifiedenglish.net)
- MAXit by Smart Communications
  (www.smartny.com/maxit.htm)
- SEC by Talisen Technologies
  (owl.li/nvVjO)

Most of the products listed above are very mature, server-based enterprise solutions that were designed to support the large, geographically distributed teams of technical writers.

Are there any no-cost/low-cost controlled language rule sets and tools?

Traditionally, controlled language environments were primarily used in large and very large organizations. However, one of the most exciting developments in the controlled language field is the recent release of several free CL resources. With powerful rule sets and a user-configurable vocabulary checker available at no cost, controlled language authoring (in English) is now an option for businesses of any size.

ASD-STE 100

In January of 2013, the AeroSpace and Defence Industries Association of Europe (ASD), made Issue 6, which is the latest version of ASD-STE100, available to the public for free. ASD is the successor to AECMA, and ASD-STE100 is the updated version of Simplified Technical English. The PDF document that ASD distributes is 368 pages long and contains the complete set of writing rules as well as the entire dictionary of approved words. To receive a free copy of ASD-STE100, submit a request on the following web site: www.asd-ste100.org/request.html

Term Checker

The STE Term Checker is a special version of LanguageTool, an Open Source grammar, style and vocabulary checker, that has been optimized for ASD-STE100. The free version of the STE Term Checker implements the entire

(approved and non-approved) vocabulary of ASD-STE100 Issue 3, plus a number of vocabulary-related rules. This tool identifies not-approved words in a text, part-of-speech challenges and several other issues related to ASD-STE100 vocabulary.

The developers of the STE Term Checker (and of LanguageTool) provide information on how to customize this application with user-specific words. If properly configured, the STE Term Checker recognizes not only user-specific vocabulary but also inflected word forms. More information on the STE Term Checker and a download link are available on the following web site:
www.simplified-english.co.uk/index.html

Controlled Language Optimized for Uniform Translation (CLOUT)

CLOUT is a set of ten writing rules that this author found to have a high impact on the translation quality of several low-cost machine translation systems across several language pairs. CLOUT is a minimalist controlled language that is effective at improving the comprehensibility and translatability of technical documents. The CLOUT rules are intentionally simple so that even writers who have little or no prior writing training can quickly learn and use these rules. Here are a few examples of CLOUT rules:
- Write sentences that are shorter than 25 words.
- Write sentences that express only one idea.
- Write the same sentence if you want to express the same content.

In addition to the ten writing rules, CLOUT also includes 'Do' and 'Do not' examples for each rule. Here is an example of a CLOUT rule in its entirety:

Controlled language rule 5:
Write sentences that have a simple grammatical structure.
Write:
Show that you can organize your thoughts by using a simple sentence structure in your texts.
Do not write:
You, in your texts, to show that you can organize your thoughts, should use a simple sentence structure.

CLOUT is available as a one-page PDF document from this website: owl.li/nJscx

Figure 1: Source text in the top window, STE Term Checker results in the bottom window
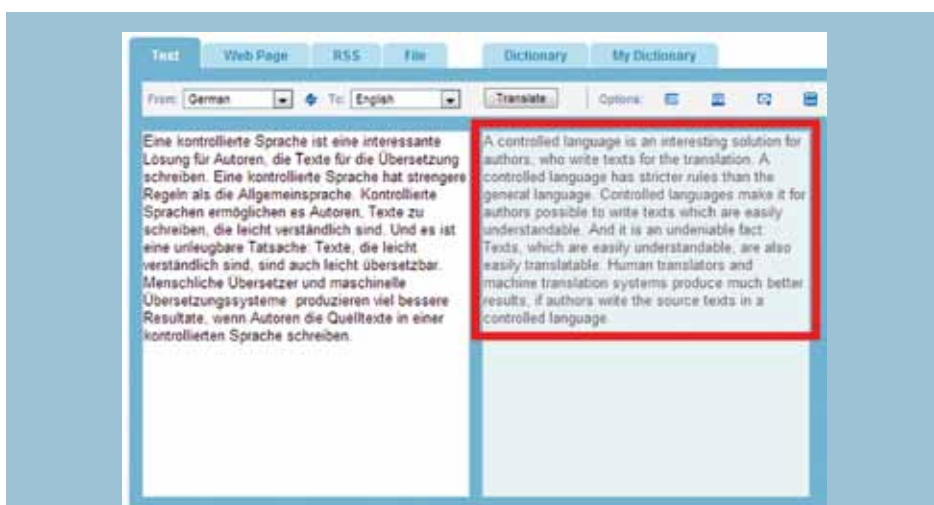

Figure 2: Example of a highly comprehensible machine translation from a German controlled language source text into English generated by a free (rules-based) machine translation service (SYSTRAnet). No human post-editing was applied.

Summary

For many years, large organizations have been using controlled language authoring as a strategy for improving the readability and translatability of technical documents. Now that several powerful controlled language resources, including the ASD-STE100 rule set, are available for free, getting started with controlled language writing no longer requires a major investment.

While getting started with controlled language writing is now easier and cheaper than ever, it should be noted that changing the way technical authors write is still a major undertaking. However, the many benefits controlled language writing offers should convince many more managers of technical writers to give controlled language writing at least a try.

Further reading and resources

· MUEGGE, U. Controlled Language Optimized for Uniform Translation (CLOUT). 2002. Available from World Wide Web: http://owl.li/orRQ3

· MUEGGE, U. Do-it-yourself MT: Taking (statistical) machine translation to the next level. tcworld. 2013, vol. 8, no. 3, pp. 20-23. Available from World Wide Web: http://owl.li/orRRq

· MUEGGE, U. Möglichkeiten für das Realisieren einer einfachen Kontrollierten Sprache. Lebende Sprachen. 2002, vol. 47, no. 3, pp. 110-114. Available from World Wide Web: http://owl.li/orRSW

· MUEGGE, U. TechScribe STE Term Checker: Uwe Muegge reviews a free vocabulary checking tool for ASD-STE100. Communicator. 2013, vol. 24, no. 2, pp. 28-29. Available from World Wide Web: http://owl.li/orRV0 ISSN 0953-3699.

Uwe Muegge has more than 15 years of experience in the translation and localization industry, having worked in leadership functions on both the vendor and buyer side. Uwe has been with CSOFT International, a provider of language services based in Beijing, since 2008, and he currently serves as Senior Translation Tools Strategist for North America.

uwe.muegge@csoftintl.com
www.csoftintl.com