

2007

Controlled language: The next big thing in translation?

Uwe Muegge, *Monterey Institute of International Studies*

Technology Spotlight

Controlled Language

by Uwe Muegge
muegge.cc

The Next Big Thing In Translation?

Many global organizations are beginning to see the productivity indicators for their translation and localization processes reach a plateau. That's an inevitable fact even for those organizations that use what's currently billed as the latest and greatest in translation technology, such as translation memory with automated workflow components or globalization management systems. Even with these tools in place, making content available in multiple languages remains a very expensive and time-consuming proposition. For those looking for ways to reduce the cost of translation to the point where almost all materials that should be translated, controlled language may be a viable option.

WHAT IS A CONTROLLED LANGUAGE?

A controlled language has two essential characteristics: The grammar of the controlled language is typically more restrictive than that of the general language, and the vocabulary of the controlled language typically contains only a fraction of the words that are permissible in the general language. This means that authors who write in a controlled language have fewer choices available when writing a text. For example, the sentence "Work must be spell-checked before publishing it" is a perfectly acceptable sentence in general English. Using the CLOUT™ controlled language rule set;

however, that sentence would have to be rewritten as "The authors must spell-check their documents before the authors publish their documents" to comply with rules regarding vocabulary, active voice, and avoidance of pronouns.

NOT ALL CONTROLLED LANGUAGES ARE CREATED EQUAL

The concept of controlled language is not exactly new: An early example of a controlled language is C. K. Ogden's *Basic English*, which was introduced in the 1930s. Since then, there have been dozens of controlled-language initiatives for English: e.g. *Avaya Controlled English* (ACE), GM's *Controlled Automotive Service Language* (CASL), White's *International Language for Servicing and Maintenance* (ILSAM), *Caterpillar Fundamental English* (CFE) and *Caterpillar Technical English* (CTE), IBM's *Easy English*, Kodak's *International Service Language*, *Nortel Standard English* (NSE), *Perkins Approved Clear English* (PACE), *Sun Controlled English*, and *Xerox Multilingual Customized English*. And then, of course, there is *ASD-STE100 Simplified Technical English*, aka *Simplified English*, the best-known and most widely used controlled version of the English language.

It may come as a surprise to some readers that while having a controlled language available certainly helps translation; many controlled languages have been



Translation and Localization Services

in all Languages • info@e-kern.com • www.e-kern.com

- Experienced and qualified translators and editors
- Document, software and website localization
- Use of specialized translation memory systems and localization tools
- Foreign-language DTP and printing services
- Over 30 years of experience in European, Asian and American markets

KERN Global Language Services in Europe (+49 69 75 60 73-0), Asia and America (+1 888 775 5376)

CONTROLLED LANGUAGE AND TRANSLATION

One of the biggest challenges facing organizations that wish to reduce the cost and time involved in the translation of their materials is the fact that even in environments that combine content management systems with translation memory technology, the percentage of un-translated segments per new document remains fairly high. While it is certainly possible to manage content on the sentence/segment level, the current best practice seems to be to chunk at the topic level. Which means that reuse occurs at a fairly high level of granularity. In other words: There is too much variability within these topics!

STAYING WITHIN THE TRANSLATION-MEMORY PARADIGM

The most effective way of optimizing a globalization environment that is based on translation memory technology is to normalize the source that feeds the translation memory tool. Normalizing the source means reducing variation between sentences. Writing in a controlled language reduces variation 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 function⁴. In a functional approach to controlled language authoring, there are specific rules for text functions such as heading, condition, process, or warning message. Here are two simple examples for functional controlled language rules:

Text function: Step (instruction)

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

Example: Click the button.

Text function: Result (instruction)

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

Example: The window "Expense Report" appears.

Implementing functional controlled language rules will enable authors to produce text where sentences with the same function have a very high degree of similarity. That not only makes sentence modules reusable within and across topics in a content management system, but also dramatically improves the match during translation.

MOVING UP TO THE MACHINE-TRANSLATION PARADIGM

Machine translation is receiving a lot of attention these days, and yet, by all accounts, the number of organizations that use machine translation as part of their globalization processes is very small. That's not really surprising as this technology is still not very well under-

"Machine translation is receiving a lot of attention these days, and yet, by all accounts, the number of organizations that use machine translation as part of their globalization processes is very small"

stood. Just ask any vendor of a machine translation tool or service what percentage of their own technical and marketing material was actually translated by machine - the answer might be surprising.

Nevertheless: Machine translation works, and it has been working in production environments for many years. In fact, this author has implemented a machine translation environment at a major global player that produces translations that don't require any human post-editing⁵. Currently, this system is only capable of translating product descriptions in a highly controlled language, e.g. "Plate 245536-BA right-angle blue 15 mm 1 ea". While the product database certainly constitutes only a very small percentage of the translatable content available in a global organization, the ability to automatically generate product descriptions in multiple languages and push those translations out to all systems that need them, is certainly highly desirable.

The big question really is: Can today's machine translation systems handle more complex challenges such as technical documentation? And the answer is: Yes! However, most machine translation scenarios will involve some degree of human post-editing. And controlled language can play a major role in reducing the amount of human intervention to a minimum.

Unlike in a traditional translation memory environment, where uniformity is the decisive factor in improving efficiency, the big factor for making machine translation systems more productive is reducing ambiguity in the source text. The problem that rules-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 rules-based machine translation systems to produce better translations, the controlled

developed with other goals in mind than supporting translation, let alone machine translation. Both *Basic English* and *Simplified English* are geared towards facilitating language learning. In other words, their goal is in fact to *avoid translation altogether* by making source texts available in a variant of the English language that users can learn in a few weeks time - as compared to the 5+ years of learning it typically takes to master Standard English.

A further indicator of the different goals of these controlled languages is the fact that they do not have lot in common in terms of their rules base. *Nortel Standard English*, for instance, has only a little over a dozen rules, while *Caterpillar Technical English* consists of more than ten times as many. And a recent comparative analysis of eight controlled English languages found that the number of shared features was exactly one, i.e. a preference for short sentences.²

WHAT ARE THE BENEFITS OF USING A CONTROLLED LANGUAGE?

Enabling authors to produce text that is easier to read, comprehend, and retain, as well as more consistent in terms of vocabulary and style, has many advantages for the organization that provide a controlled language authoring environment. Here are a few of the most important reasons for introducing a controlled language:

- Documents that are more readable and more comprehensible improve the usability of a product or service and reduce the number of support incidents.
- Controlled-language environments provide authors with powerful tools that give objective and structured support in a typically rather subjective and unstructured environment.
- Tools-driven controlled language environments enable the automation of many editing tasks and provide objective quality metrics for the authoring process.
- The more restrictive the controlled language, the more uniform and standardized the resulting source document, the higher the match rate in a translation memory system, and the lower the translation cost in a conventional translation environment.
- A controlled language that was designed for machine translation will significantly improve the quality of machine-generated translation proposals and dramatically reduce the time and cost associated with human translators editing those proposals for producing translations for previously un-translated material.



IF CONTROLLED LANGUAGES ARE SO GREAT, WHY ISN'T EVERYONE USING ONE?

Even though controlled language has been established as a practice in an industrial context more than 30 years ago³, there are very few organizations that have embraced a comprehensive controlled language philosophy - at least, there are not many that talk about it. This means that anyone new to the field may have a hard time finding reliable, vendor-independent information on what solutions are available and what the costs and benefits of deploying those solutions are. The fact that many controlled-language tools have been designed with corporate customers - and their deep pockets - in mind, hasn't really helped spreading this authoring approach beyond a very small circle of companies. While there are a few controlled-language tools available in the \$1000-\$5000 price range, e.g. the *MaxIt Checker* from Smart, many more tools reside in the \$50,000-\$100,000 range, e.g. *acrocheck* from acrolinx or *CLAT* (Controlled Language Authoring Tools) from IAI. With the high-end tools receiving much more publicity than the lower-priced ones, it may be difficult for a smaller organization to make a convincing ROI case.

Finally, deploying a controlled language solution means implementing an environment, in which authors have much less creative freedom. Some authors have pushed back at the roll-out of content management systems as these systems force authors to create content in chunks and reuse those chunks instead of creating new ones at the author's discretion. Therefore it is fair to expect that the introduction of a system that forces authors to make prescribed choices concerning the grammar and style of every sentence and every specialized term every author writes requires a major education and training effort.

language needs to have rules like the following that helps the machine translation system to 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

rules-based machine translation systems, the quality of the output produced by these machine translation systems is bound to improve dramatically. In a recent study I conducted as part of the advanced computer-assisted translation course I teach at the Monterey Institute of International Studies, the productivity of students jumped approx. 50% when post editing a machine-translated text that was written in a controlled language compared to post editing a simple machine-translated uncontrolled text of the same length and level of difficulty. And, by the way, these results were achieved using exclusively no-cost translation software and services.

WATCH CONTROLLED LANGUAGE AND MACHINE TRANSLATION WORK

For even more compelling evidence that controlled language and machine translation make for a winning team, visit www.muegge.cc, a site that was designed from the ground up for enabling machine translation and all text was written in CLOUT, the Controlled Language Optimized for Machine Translation. On the home page, click on any of the language combinations into English, i.e. German > English or French > English and watch how Google's free machine translation system turns a complete website into a fully navigable, highly comprehensible virtual English version in real time. And that's just a glimpse of how controlled language authoring and machine translation can transform globalization processes.

(ENDNOTES)

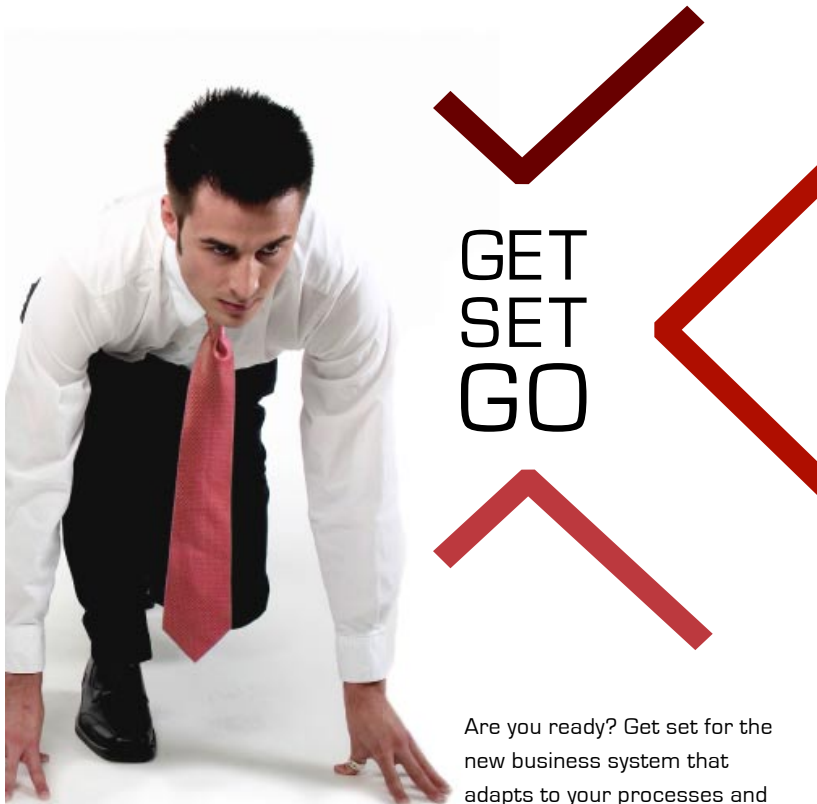
¹ The CLOUT™ rule set was developed by Uwe Muegge specifically for the purpose of helping authors write source text for subsequent machine translation. CLOUT stands for Controlled Language Optimized for Uniform Translation.

² O'Brien, S. (2003). "Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets", EAMT/CLAW 2003, Dublin, Dublin City University

³ Caterpillar started using Caterpillar Fundamental English in the early 1970s.

⁴ One example of a well-developed authoring rule set based on a functional approach is Funktionsdesign® [functional design], developed by professors Robert Schäflein-Armbruster and Jürgen Muthig.

⁵ Muegge, Uwe (2006): "Fully automatic high quality machine translation of restricted text: A case study", in "Translating and the computer 28. Proceedings of the twenty-eighth international conference on translating and the computer, 16-17 November 2006, London", London: Aslib.



Are you ready? Get set for the new business system that adapts to your processes and manages them – all the way from quote to invoice. Whether you localise, interpret, translate, publish, teach or consult, or do a bit of everything - LTC WORX can handle it.

Visit us today on:
www.ltc-worx.com



No More Compromises.