Johanna Schels

605 35th Street West, Hastings, Minnesota 55033 USA | Phone: 1+ 612-615-1730
johanna.lorah@comcast.net | johanna.schels@gmail.com | Skype: johanna.schels2

Master of Science in Technical Communication

Academic Literature Review

Controlled Authoring: Writing for Translation and Content Re-use

for

Department of Communication, Writing and the Arts

Metropolitan State University, Saint Paul, Minnesota

Contents

Controlled Authoring: Writing for Translation and Content Re-use

Introduction

In the age of globalization, every organization involved in marketing products and/or services to an international audience requires the translation and localization of diverse marketing collateral, from product manuals, to brochures, to website content. Knowledge exchange between researchers and scientists who speak different languages relies on the accurate and timely translation of their work. Real-time communication via new media platforms among people in different countries requires instantaneous translation. The size of the volume of content to be translated is inconceivable. Human translators and machine translators (also called automated translators) work to bridge the linguistic barriers of multilingual communication by translating from source languages into other target languages.

Background

Professional translation, localization, and post-translation editing services are expensive. In the translation services industry, professional translators are paid per word to be translated from the source language into the target language. The source document is analyzed for word count with a translation software tool such as SDL Trados. Word counts are comprised of *no matches*, *fuzzy matches*, and *exact matches*. The translator charges the highest amount for *no matches*, followed by *fuzzy matches*; there is no charge for *exact matches*, because the translation already exists in the translation memory (TM) of the programmed translation software tool. Machine translation can be a faster and cheaper option. However, it requires programming of the machine, diligent quality control, and post-editing of machine-translated content by a linguist. Compared to human translation, machine translation offers high-speed and high-volume output, but costs for programming, developing terminology, and post-editing of machine-translated content can easily offset any savings.

There is a direct correlation between translation costs and word count of a document: When the word count is reduced, costs are reduced. Technical writers can help their organizations reduce costs for the translation and localization of documents. This is accomplished by using controlled authoring.

This literature review discusses how the technical writer can prepare and write content for translation and localization to save costs and add efficiency to the translation and localization processes, without jeopardizing accuracy and quality of the output. As a result, international customers and clients benefit from content that is understandable and consistent in the use of terminology, grammar, syntax, semantics, style, and tone. This creates trust between the customer and the business.

Definition of Controlled Authoring: Writing for Translation *is* Writing for Content Re-use

In his article, *Controlled Authoring–Writing for Re-Use*, Helbich (2006) suggests that savings are "waiting to be realized when and where the content is created" (p. 2). According to Helbich (2006), controlled authoring reduces translation and localization costs; controlled authoring

(within the context of Helbich's article) is "writing for re-use or writing for translation" (p. 2). Helbich (2006) explains, "It's a process that integrates writing with localization so that the text can be written for re-use and at the same time written for efficient translation" (p. 2).

Helbich (2006) outlined four authoring controls writers can use. Following the definition of each authoring control, I provide a table with examples, which are taken from Helbich's article and from TURCK's RM-89 product manual, a document I revised for translation.

Text Modularization

Text modularization is the practice of separating content that changes from content that remains the same. Static content is translated once and re-used, which saves costs.

Helbich's example

Combined text	**Modularized text**
Clipping Baggers	Clipping Baggers
Your lawn tractor has a cutting deck size of 36 inches. DO NOT install an aftermarket bagger on this lawn tractor. The bagger can become clogged and damage your lawn tractor because the discharge chute is too narrow. Aftermarket baggers installed on a lawn tractor without the proper discharge chute can cause damage and possible injuries.	Your lawn tractor has a cutting deck size of 36 inches. DO NOT install an aftermarket bagger on this lawn tractor.
	The discharge chute is too narrow and may become clogged.
	The installation of an aftermarket bagger when the discharge chute is too narrow may result in permanent damage to your lawn tractor. It may also cause you or others to be injured.
If you must use a bagger on a tractor with a cutting deck size of 36 inches, be sure to use only an approved bagger from your dealer and MAKE SURE that you install the appropriate chute adapter kit. See your dealer for factory approved baggers and adapter kits.	Use ONLY a bagger that is factory approved and BE SURE to install the appropriate chute adapter kit. See your dealer for factory-approved baggers and adapter kits.

Text Standardization

Text standardization excludes information and data specific to a product so that text can be re-used for a different model with identical features, eliminating the need for retranslating the same text.

Helbich's example

Product-specific text	Standardized text
The Commando APC-4260 has a ground clearance of 12.5 inches and an axle to ground clearance of 10 inches while maintaining a low silhouette and a low center of gravity.	This vehicle has a ground clearance of 12.5 inches and an axle to ground clearance of 10 inches while maintaining a low center of gravity.

Text Simplification

Text simplification reduces words, eliminates redundancy and ambiguity, and increases readability. As a result, translation costs are reduced.

Example from the RM-89 revision

Complex text	Simplified text
All RM-89 resolver based encoders have two status LED's to help you determine the state of the device. The LED's are always located on the back cover of the RM-89. 30 words	All RM-89 resolver-based encoders have two status LEDs that indicate the status of the device. The status LEDs are on the back of the RM-89. 25 words

Text Reduction

Text reduction eliminates irrelevant words, phrases, and sentences, also resulting in reduced word counts and thus reduced costs (pp. 3-4).

Example from the RM-89 revision

Wordy text	Reduced text
The position data can also be preset which allows you to align the position data with your machine position without having to physically rotate the shaft. 26 words	Preset the position data to align it with the machine. Do not rotate the shaft. 15 words

From my perspective as a technical writer, text simplification and text reduction can be combined because they produce the same result: the reduction of wordiness and of redundancy. From my experience as a technical translator, excessive simplification and text reduction may lead to an incorrect translation, especially when this practice results in erroneous grammar,

syntax, and semantics. As a result, the translator must conduct additional research or contact the author, which results in decreased productivity. In the event of a linguistic challenge, where the equivalent word or phrase of the source text does not exist in the target language, the translator must work around the obstacle by providing an explanation. Complex product features or processes might also require an explanation, especially when they are reduced to noun phrases with ambiguous relationships.

Text modularization will increase translation accuracy. It eliminates the use of incorrect terminology and reduces ambiguity because the text module has been edited at least once. Text modularization also increases translation productivity, especially in regard to technically difficult text that takes longer to translate (Helbich, 2006).

In addition to the four authoring controls, Helbich (2006) suggests the use of an "authoring memory (AM) system and a translation memory (TM) system" (p. 3).

In the past, I used both AM software and TM software. AM software allows quick access to content. It also allows effective version control of documents. It eliminates the tedious process of searching databases and shared file servers for re-usable content. TM software is similar. It stores already translated text. During the translation process, the translation software tool compares the text that is being translated to stored translations in its TM, and then it provides the translator with an exact match, a fuzzy match, or a no-match.

To streamline the translation process and manage the progress of multiple projects, professional translators use software tools such as SDL Trados to combine project management and the functions of AM and TM. Technical writers and translators have the unique opportunity to work together by utilizing some new technologies that accommodate both authoring and translation functions (Helbich, 2006).

In his article, *Optimizing the Source Using Translation Memory*, Campo (2009) describes a pilot project that utilized existing TM to help author and revise content in a way that reduced word counts by increasing the number of matches. Campo (2009) writes, "What writers need is a TM tool that runs side-by-side with an authoring application and can semi-automatically offer suggestions on how to better match new text to the existing TM" (p. 7). Like Helbich, Campo (2009) emphasizes the need to integrate the translation and localization processes into the authoring process and indicates that translation costs are saved with a TM, "but at a cost of labor and time" (p. 7). From his pilot project, Campo (2009) concludes that a TM increases the consistency and quality of documents and that "savings were achieved by both re-use of existing text and aggressive word-count reduction" (p. 7).

Again, I have been a technical translator for many years. I will not accept a document with a high number of pre-translated (re-used) text segments because I still need to translate what remains, and this requires reading the entire document. This means more time that I cannot expense. From my experience, most pre-translations require editing, especially when they are done by a monolingual writer. However, I will accept machine-translated text that requires quality-testing and post-translation editing. Controlled language authoring is necessary to produce content for machine translation.

Definition of Controlled Language Authoring: Controlled English or Global English

In his article, *Controlled Authoring to Improve Localization*, Ō Broin (2009) explains controlled authoring is "the process of applying a set of predefined style, grammar, punctuation rules and approved terminology to content (documentation or software) during its development" (p. 12). Helbich (2006) suggests that controlled language authoring, Controlled English or Global English, "refer to a set of language and authoring rules in which terminology, grammar and syntax are tightly constrained, and for the purpose of preparing authored text for machine translation" (p. 2).

As some examples of Controlled English or Global English developed and used for many years by well-known, global corporations, Ō Broin (2009) refers to "Caterpillar Technical English, Nortel Standard English, the Plain English Campaign, GM's Controlled Automotive Service Language, and Global English" (p. 12).

Small businesses do not have the resources to develop their own controlled language. When budgets are tight, technical writers on international documentation teams can take the lead and develop glossaries and style guides, although on a much smaller scale. Ō Broin (2009) lists some examples of controlled language rules, such as "standardized spelling, length of sentence, number of clauses, use of active versus passive, simplifying tenses, rules for noun phrases, modifiers, syntactic cues, past particles, gerunds, avoidance of Latin phrases, slang and so on" (p. 12). However, Ō Broin offers no explanation of his examples.

In her article, *Planning and Writing For Translation*, Sichel (2009) advises, "Simple, straightforward text is easiest to translate. Say what you mean as concisely as possible, and keep sentences short and limited to a single idea" (p. 3). This approach works well to produce high-quality human translations, but it is not enough for machine translation. An automated machine translator must be programmed, requiring tight constraints on language. In her paper, *Controlling Controlled English–An Analysis of Several Language Rule Sets*, O'Brien (2003) conducted an in-depth analysis of eight existing Controlled English rule sets for types and commonality of rules and concludes there is no core rule set for Controlled English, "although there is some commonality of rules across some rule sets" (p. 1). O'Brien (2003) reasons that "a core set of Controlled English rules would be useful for any individual or organization" (p. 1). Every technical writer and professional translator will agree with O'Brien, especially when teams must cooperate to complete multilingual communication projects, while preserving consistency, tone, and style.

O'Brien (2003) proposes to develop Controlled English rules that cover four main categories:
Lexical rules—vocabulary, spelling, abbreviations, acronyms, etc.
Syntactic rules—subject-verb agreement, nouns, adjectives, adverbs, articles, pronouns, etc.
Text structure rules—layout, sentence length, information load
Pragmatic rules—textual devices, information specificity, text type structure, etc. (pp. 2-4)

To demonstrate O'Brien's four categories, I took some examples from *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*, authored by John R. Kohl, who is a renowned researcher and contributor to the refinement of Global English guidelines. I also used examples of my revision of a technical product manual for the RM-89 DeviceNet™ Encoder manufactured by TURCK Inc. In the tables below, the correct examples are in the right column.

Lexical rules

Eliminate unnecessary abbreviations (Kohl's list of examples, p. 220).

ca.	about, approximately
i.a.	that is
a.k.a., AKA	also known as
n.a., N.A., n/a, N/A	not applicable, not available, none

Eliminate non-standard spelling (Kohl's list of examples, p. 215).

hi	high
lite	light
low	low
thru	through

Eliminate incorrect technical terms (Kohl's list of examples, p. 215).

box graph	box chart
message area	status bar
shortcut menu	pop-up menu
preset style	predefined style

Syntactic rules

Make each sentence syntactically and semantically complete (RM-89 example, p. 5).

Default of CW increasing when looking at the shaft.	When you look at the shaft, you will see the default CW increase.

Text structure rules

Revise noun phrases through text reduction (RM-89 example, p. 4).

The position data can also be preset which allows you to align the position data with your machine position.	Preset the position data to align it with the machine.

Pragmatic rules

Create lists for instructions (RM-89 example, p. 3).

Parameters allow you to set the count direction, format of the velocity data, number of counts per turn, and the rollover count, which is the number of counts before the position returns to zero.	Use the parameters to complete the following tasks: • Set the count direction; • Format the velocity data; • Format the number of counts per turn; and • Format the rollover count, which is the number of counts before the position returns to zero.

Use hyphenation to identify how nouns relate in noun phrases (RM-89 example, p. 5).

Figure 1.1 shows a flange mount unit with end connectors.	Figure 1.1 shows a flange-mount unit with end-connectors.

No doubt, machine translation requires diligent programming that relies on dictionaries and termbases that have been compiled over time. In their article, *Linguistic tasks on translation corpora for developing resources for manual and machine translation*, Dash and Basu (2010) write, "Within the area of machine translation research, the central point of the debate has been the question about the level of complexity involved in the task of translation corpora analysis" (p. 3). As Dash and Basu (2010) explain, translation corpora analysis involves linking "segments of source texts with those of their translations in the target texts" (p. 3).

Machine Translation Versus Human Translation

The literature I reviewed shows there are proponents and opponents of machine translation. In his article, *Dispelling the myths of machine translation*, Muegge (2008) writes, "using any of the free online translation environments gives only a glimpse of the true power of a full-fledged professional machine translation system" (p. 22). Muegge (2008) refers to a case study at Symantec, which showed that "productivity of human translators can double when unknown sentences are pre-translated in a machine translation system" (p. 23). According to Muegge, "a considerable percentage of machine-generated translations turn out to be perfect" (p. 23). In her article, *Questions and Answers about Machine Translation*, Gerber (2011) admonishes translators to edit machine translations because of its "inability to handle language nuances and culture" (p. 25). Gerber (2011) writes, "It is never appropriate to submit unreviewed, unrevised MT output to a client as a finished translation" (p. 22). It is obvious that Muegge and Gerber have opposing views. However, many clients use machine translation and ask translators to post-edit for a reduced price.

Machine Translation as a Viable Option

Both human translation (HT) and machine translation (MT) are linguistic processes that involve the written translation of one language into one or multiple other languages. In his article, *How to implement machine translation*, Muegge (2002) suggests, "there are approximately 400 commercial machine translation systems available on the market" (p. 3). All MT systems are based on three different methods and work best when combined with controlled language authoring and translation memory (TM) technology.

Types of Machine Translation (MT)

Direct MT

Direct MT is also known as dictionary-driven machine translation. Direct MT translates a word or phrase from the source language into the target language with no regard to grammar, syntax, semantics, style, and tone (Muegge, 2008).

This type of MT is regarded as obsolete because of its limitations. However, Muegge (2008) believes that Direct MT still has its place; it is an ideal solution for technical documents such as product descriptions with numeric data strings. In his article, *Wie mit proaktivem Terminologiemanagement Übersetzungen auf Knopfdruck möglich werden*[1], Muegge (2008) describes how Medtronic tested its translation workflows for cost reduction, translation quality, and fast turnaround. As a result, Medtronic chose Direct MT for the translation of product descriptions and specifications for thousands of products.

In his article, *Dispelling the myths of machine translation*, Muegge (2009) suggests that Direct MT is suitable when vocabulary and syntax are standardized. For example, "weather reports, financial profiles, and many e-commerce applications" (p. 119).

Rules-based MT

Rules-based MT is also known as transfer machine translation. The translation process has three steps:

1. **Analysis**: The tagged segment (phrase or sentence) of the source text is parsed to create the syntactic structure (blueprint) of the tagged segment.
2. **Transfer**: The syntactic structure of the tagged segment of the source text is transferred into the corresponding structure of the target text.
3. **Generation**: The corresponding structure of the target text is populated with the translated words from the tagged segment of the source text.

Rules-based MT is widely used because the machine (system) can be programmed. For example, customized termbases, dictionaries, and settings for grammar, syntactic, and semantic rules.

[1] Translation of article title: How translations become possible with proactive terminology management and by pushing a button.

Muegge (2009) points out that writers must adjust their writing styles and incorporate the rules, because "the rules base of any system is by necessity limited" (p. 118). SYSTRAN, Babel Fish, and @ prompt Expert are well-known rules-based MT systems (Muegge, 2009).

Statistical Machine Translation (SMT)

SMT has two components:

a. **Translation Model**: Aligns existing source text and already translated target text (stored in databases, repositories, and cloud-based termbases, dictionaries, and TMs) to create a model. Generates a translation proposal based on this model, focusing on identical or similar word sequences.

b. **Language Model**: Selects a translation proposal that is closest to an existing example in the target language.

SMT systems are trained. Muegge (2009) writes, "once an SMT system has been trained on customer-specific data, this is the MT technology that typically produces the highest translation quality" (p. 118). However, an astronomically high volume of source text and translated target text is needed to train the machine. The benefit of this significant investment in resources is an in-house capacity to render a high volume of quality translations in a short period of time, and when products need to get on the market fast (Muegge, 2009).

Muegge (2009) writes that Language Weaver is "the leading vendor of statistical machine translation systems" (p. 118). Two large corporations that use SMT to translate online content are Microsoft and Google.

Implementation of Machine Translation

Most professional translators I know do not invest in MT to increase their productivity, also because of concerns with cost. I do not use it because from my experience, and depending on the type of content to be translated, it can be more time consuming to edit machine-translated content than to render the translation myself (translation always includes editing).

In his article, *How to implement machine translation*, Muegge (2002) states, "Machine translation technology promises fast and cheap translations. However, this technology is not easily implemented" (p. 5). For machine translation to work, companies must complete a time-consuming list of tasks:

- Create comprehensive dictionaries that are product- or field-specific.
- Manage terminology.
- Program rules for grammar, syntax, and semantics (possible with high-end MT systems).
- Write content in a controlled language such as Global English or create style guides.
- Edit the content before translation.
- Edit the content after translation.
- Integrate communication technology (from content management system, to authoring tools, to machine translation, to translation memory) to avoid duplication of work (Muegge, 2002).

Costs and time can be saved when companies recognize what MT involves, and when they are ready to invest in technical writers who are skilled in writing in a controlled language environment. Without controlled language such as Global English, MT is not viable.

Controlled Language Optimized for Uniform Translation (CLOUT)

In his article, *Implementing a controlled language is now cheaper and easier than ever*, Muegge (2013), the developer of CLOUT, provides this definition, "CLOUT is a minimalist controlled language that is effective at improving the comprehensibility and translatability of technical documents" (p. 12).

CLOUT has ten simple writing rules. In the table below, I compared these rules to John R. Kohl's first ten Global English guidelines for human translation. The comparison shows that the Global English guidelines provide clearer instructions than the more simplistic CLOUT rules.

CLOUT Rules	**John R. Kohl's Global English Guidelines**
1. Write sentences that express only one idea.	1. Limit the length of sentences.
2. Write the same sentence if you want to express the same content.	2. Don't use a telegraphic style.
3. Write sentences that are grammatically complete.	3. Consider defining, explaining, or revising noun phrases.
4. Write sentences that have a simple grammatical structure.	4. Make each sentence syntactically and semantically complete.
5. Write sentences in the active form.	5. Don't add verb suffixes or prefixes to nouns, acronyms, initialisms, or conjunctions.
6. Write sentences that repeat the noun instead of using a pronoun.	6. Revise dangling –ing phrases.
7. Write sentences that use articles to identify nouns.	7. Use nouns as nouns, verbs as verbs, and so on.
8. Write sentences that use words from a general dictionary.	8. Use standard verb complements.
9. Write sentences that only use words with the correct spelling.	9. Avoid ambiguous verb constructions.

From my perspective as a translator, I agree that short sentences that follow rules for grammar and syntax are easier to translate than incomplete sentence structures with missing syntactical cues such as capitalization and punctuation. I want to emphasize one Global English guideline because of its importance to my work as a translator: The relationship of nouns in noun phrases must be made clear through hyphenation, especially in technical content. It is very time consuming to separate a string of technical terms to determine their relational meanings. This is important because an incorrect translation in an instruction manual can lead to personal injury and property damage.

Conclusion

The second part of this academic literature review is an application of Global English, especially a set of guidelines suggested by John R. Kohl. I revised the first ten pages of an instruction manual for a new product, the RM-89 DeviceNet™ Encoder manufactured by TURCK Inc., a global leader in industrial automation.

Based on a word count analysis with the translation software tool SDL Trados, the revised content of the instruction manual had 320 words more than the source text. However, the revised content is more precise, and thus easier to translate, eliminating costly errors. The revised content, although technical in nature, is not suitable for machine translation, because it is too complex.

Controlled language such as Global English is suitable for writing content that is easier to understand by native and non-native speakers. Global English focuses on standard terminology, correct grammar, complete syntax, unambiguous semantics, and short text segmentation, also by using punctuation for emphasis and clarity.

Global English is also suitable for translation, localization, and content re-use if the main consideration for its use is improved communication to support the audience. If the consideration is limited to saving costs and time, I recommend a cost-benefit analysis for each international communication project. Global English can be implemented by small and large organizations, but project-specific implementation should be analyzed. Here, it is helpful to understand the translation and localization process. Machine translation is suitable for large volumes of technical content if the machine is programmed. Again, a cost-benefit analysis can help determine whether the business or the organization can spend the money for the required resources.

References

Campo, Joseph. (2009). Optimizing the Source Using Translation Memory. *MultiLingual, Writing for Translation Getting Started: Guide*, 5-7 Retrieved from http://www.multilingual.com/gsg

Dash, Niladri S., & Basu, Pronomita. (2010, June). Linguistic tasks on translation corpora for developing resources for manual and machine translation. *SKASE Journal of Theoretical* Linguistics, 7(2), 1-18 Retrieved from http://www.skase.sk/Volumes/JTL16/pdf_doc/01.pdf

Gerber, Laurie. (2011, November/December). Questions and Answers about Machine Translation. *The ATA Chronicle*, XL (11), 21-27

Hearne, Mary, & Way, Andy. (2011, May). Statistical Machine Translation: A Guide for Linguists and Translators. *Language & Linguistics Compass*, 5(5), 205-226 Retrieved from http://web.ebscohost.com.ezproxy.metrostate.edu/ehost

Helbich, Carl. (2006). Controlled Authoring – Writing for Re-Use. *MultiLingual , Writing for Translation Getting Started: Guide*, 17(7), 2-4 Retrieved from http://www.multilingual.com/gsg

Kohl, John R. (2008, July). *The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market*. Cary, NC: SAS Institute Inc.

Muegge, Uwe. (2008, August). Dispelling the myths of machine translation. *tcworld*, 3(4), 22-25 Retrieved from http://works.bepress.com/uwe_muegge/5

Muegge, Uwe. (2002). How to implement machine translation. *The Selected Works of Uwe Muegge*. Retrieved from http//works.bepress.com/uwe_muegge/47

Muegge, Uwe. (2013, October). Implementing a controlled language is now cheaper and easier than ever. *tcworld*, 8(4), 11-13 Retrieved from http://works.bepress.com/uwe_muegge/91

Muegge, Uwe. (2008, January). Wie mit proaktivem Terminologiemanagement Übersetzungen auf Knopfdruck möglich werden. *The Selected Works of Uwe Muegge.* Retrieved from http://works.bepress.com/uwe_muegge/18

O'Brien, Sharon. (2003). Controlling Controlled English – An Analysis of Several Language Rule Sets. *School of Applied Language and Cultural Studies.* Dublin City University. Dublin, Ireland. Retrieved from http://www.mt-archive.info/CLT-2003-Obrien.pdf

Ō Broin, Ultan. (2009). Controlled Authoring to Improve Localization. *MultiLingual, Writing for Translation Getting Started: Guide*, 12-14 Retrieved from http://www.multilingual.com/gsg

Sichel, Barb. (2009, October/November). Planning and Writing for Translation. *MultiLingual, Writing for Translation Getting Started: Guide*, 3-4 Retrieved from http://www.multilingual.com/gsg