CLAW 2006
5th International Workshop on
Controlled Language Applications
at conference of the Association for
Machine Translation in the Americas (AMTA 2006),
Boston Marriott, Cambridge, MA, USA
August 12, 2006

SMART Controlled English – Paper and Demonstration

John M. Smart
Smart Communications, Inc., New York City
jsmart@smartny.com
www.smartny.com

Abstract:
The trend to globalization and "outsourcing" presents a major linguistic challenge.
This paper presents a proven methodology to use **SMART Controlled English** to write technical documentation for global communications. Today, large corporations must adjust their business practices to communicate more effectively across all time zones and 80 languages. The use of **SMART Controlled English,** when coupled with Statistical Machine Translation (SMT), will become an ideal method to cross the language barrier.

Introduction:
The trend to globalization presents a major linguistic challenge for large and small companies. To add to this trend, most products require a high degree of computer literacy for operation and maintenance. For example, most automobiles are welded by robots, not humans. Also, the advent of "outsourcing" has expanded the ring of communications. The biggest problem is that most technical manuals are not written by professional technical writers, but engineers who are the subject matter experts.

Many advanced products, like those found in the telecommunications industry, update their technology every six months. Today, many cell phone (mobile phone) users in China update their handsets every four months to get new features. Unknown to most users, the information needed to control ring tones is some 250,000 pages of complex software documentation.

The instructions to repair a complex jet engine can amount to more than 500,000 pages. According to Boeing, if all their aircraft manuals where printed and stacked end-to-end, the stack would reach to the top of Mt. Everest and back. These mountains of manuals are further compounded by the need for language translations. For example, companies like Microsoft and IBM localize their software and documentation in 70 languages. A small company seeking compliance to the Economic Union directives is faced with 20 languages. The expansion of both NATO and the EU adds more languages. Unfortunately, the demand for professional technical translators far exceeds the supply.

What is the solution?
Many companies have found that a controlled language approach can reach across the language boundaries with a common language. This paper and on-line demonstration http://www.smartny.com/ControlledEnglish/CLAW06 shows how to create and use a Controlled English dictionary.

Smart Communications, Inc., in New York City, developed **SMART Controlled English** to help companies quickly create technical documentation for global communications. The use of Controlled English allows corporations and their suppliers to blend their cultures and business practices with a common language. In this age of outsourcing, Indian firms like Tata Consultancy Services, Wipro and Infosys must mesh their information technology staffs with their client staffs to communicate effectively across 24 time zones. To illustrate the importance of technical communication, the new **Boeing** Dreamliner aircraft is fabricated from suppliers in 36 countries. The expansion of globalization increases the demand for **Controlled English** as the solution to a modern tower of Babel. The use of sophisticated software offers an ideal method for companies to economically cross global and language barriers.

What is a Controlled Language

The Microsoft Encarta encyclopedia has an excellent definition.

"A simplified form of technical English based on 1,200 key words. A Controlled English vocabulary promotes increased readability and usability of English technical documentation. Any technical idea can be expressed clearly without the complexities of English spelling and grammar and complex syntax." Microsoft Encarta 2005

This definition gives the meaning to the term Controlled Languages. By contrast, the English language has more than 900,000 words and grows daily.

In August 2006, the latest word to enter the English language is the noun "*mashup*".
This term indicates the use of hybrid software to combine content from more than one source. For example, digital maps linked to web services to show restaurant locations and perhaps menus and reviews.

There is a growing use of wireless devices to send short electronic messages.
For example, "TXTING IS OK 4U, LOL". Btw, some scientists think the use of text messages weakens the grammatical abilities of the users. In case you do not know, btw means "by the way". BTW, studies in Canada have proven that text messages require good grammatical skills to make sense to the reader on a small 215-pixel screen.

A brief history Controlled Languages
In the 1930, an eccentric Englishman named C.K.Ogden developed an 800- BASIC vocabulary to help spread the use of English to Africa, India and Asia. This vocabulary formed the nucleus of future Controlled Languages.

In 1970, the Caterpillar Tractor Company (now Caterpillar) developed Caterpillar Fundamental English (CFE), now called Caterpillar Technical English (CTE) to write their expanding technical support libraries. Caterpillar embarked on a global campaign to use CTE for maintenance manuals, service bulletins and other support publications.
In 2006, this effort continues in some product lines, especially the large mining trucks used in remote locations like Papua New Guinea. The CTE vocabulary has a core of 1,500 words to which is added the names of the parts from Caterpillar product lines.

In 1988, the European aircraft alliance, called Airbus Industrie, followed Caterpillar and developed a Controlled English vocabulary to write aircraft maintenance manuals. Today, the **ASD-STE100 Simplified Technical English** is widely deployed by 800 airlines and mandated by Airbus, Boeing, Embraer, Bombardier and other aerospace companies. To avoid product liability, these companies must mitigate the risk of human error in maintenance tasks. For example, ASD-STE100 Simplified Technical English does not contain the common English auxiliary verbs, *might*, *should* or *may*. These verb forms do not exist in other languages. The ASD-STE100 Simplified Technical English allows the use of valid part names and terms in addition to the basic 985 words.

You are not reading a typographical error. There are only 985 basic terms. However, the dictionaries expand by component with terms for hydraulic systems, avionics and engines. An average STE vocabulary can reach 8,500 to 12,500 terms. Later in this paper, there is an example of ASD-STE100 Simplified Technical English in use for the new Airbus A380 that can carry 550 or more passengers.

In 1995, SMART Communications, Inc. (SMART) was commissioned by Nortel Networks, a large global manufacturer of telecommunications equipment, to develop a controlled language. At that time, the global demand for telephone switches expanded rapidly for the Internet. Nortel did not have enough technical translators to keep up with the demand for multilingual documentation.

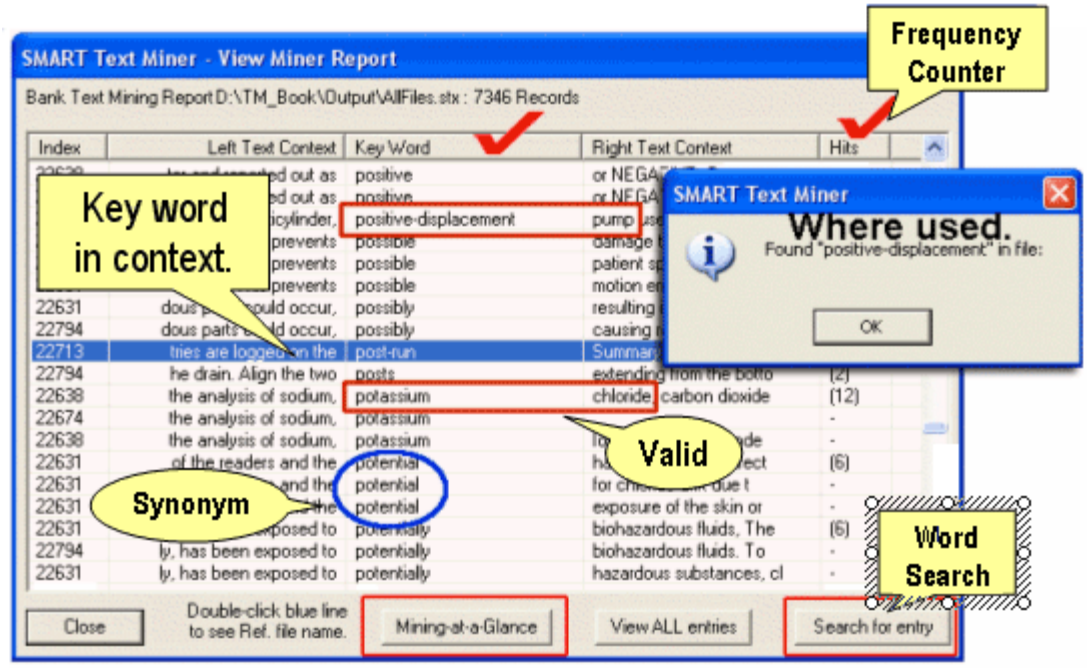
SMART used the steps outlined in this paper to develop Nortel Standard English. All of the technical terms were developed with the help of the SMART Text Miner. The subject matter analysis was done by Dr. Ralph Calistro, of Nortel Customer documentation staff, in Ottawa, Canada. During the project, SMART and Nortel saw the need for a sophisticated grammar checker to help their busy writers. To meet this need, SMART wrote the **MAXit Checker** that uses *artificial intelligence* to read, understand and electronically critique Controlled English sentences. The **MAXit Checker** now has a decade of use by thousands of writers around the world. The **MAXit Checker** is a plug-in to Adobe FrameMaker, WORD, Arbortext Editor and other tools.

The writers at Nortel write their documentation in structured (Extensible Markup Language) XML, then ask the **MAXit Checker** to check for compliance. This inexpensive, quality-control technique has eliminated some 42% of procedural errors and made complex telecommunications documentation easier to read and understand in 208 countries. There is an example shown later in this paper.

How to Develop a Controlled English Dictionary

There are three easy steps, which are also shown in the demonstration:

Step 1: Text Mining: This task is the extraction of knowledge from existing documents by a computer. The documents are processed by the **SMART Text Miner** to generate a report, as shown below. This report allows the user to quickly sort the good terms from the bad terms and store that knowledge in a database. These databases are distributed on a server to all writers and authors.



In the picture above, from left to right, column one holds an index to the original document. Column two shows the left context. Column three shows the key word. Column four shows the right context. Column five shows the number of occurrences the key word is found in a set of documents. The higher the frequency of occurrence of word in the text, the more important the term.

This software is easy-to-use by persons who are not lexicographers or linguists.
The terms are extracted and checked by subject matter experts. In this example, the user can quickly identify a **"positive displacement pump",** which is a valid term with a valid NATO stock number. Another term that is easy to identify is "**potassium chloride**", a valid chemical name. The **Text Miner** also identifies ambiguous terms like "potential", which can mean a measure of electrical force or an adjective. These terms are put in a reverse dictionary for use in the **MAXit Checker**. The **Text Miner** has a convenient browse function, term harvester and Mining-at-glance feature. The objective is to discover the hidden terminology that must be included in a **Controlled English** dictionary. It is possible to develop dictionaries manually, but the computer-approach is fast, accurate and more efficient.

Step 2: Electronic Critique – MAXit Checker

It is possible to manually check texts for compliance to a **Controlled English** vocabulary, but the manual method is too slow. The key to a successful Controlled English program is to give the technical writers and authors a tool that does 95% of the work, but leaves the final decision to writer.

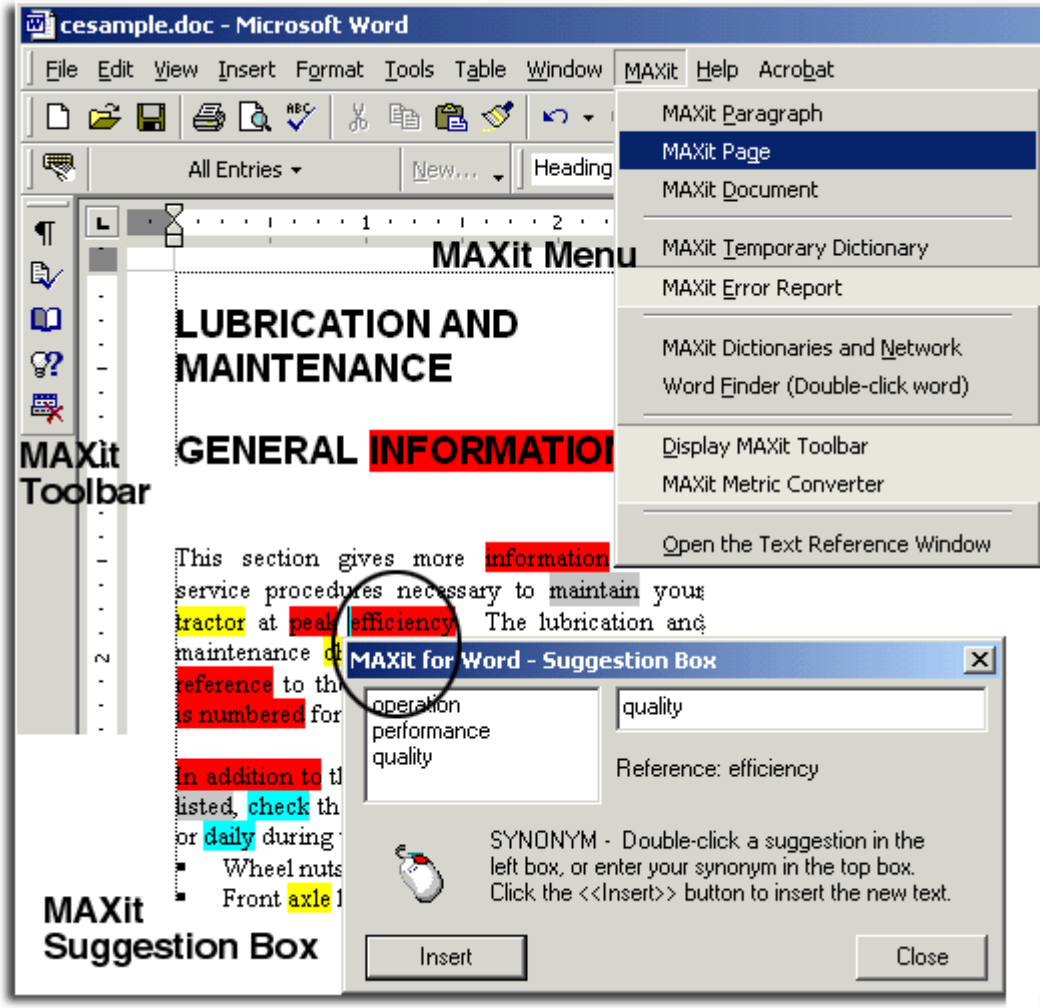
The **MAXit Checker** is such a tool and integrated in major publishing tools to extend the grammar checking functions. Because **MAXit** uses **artifical intelligence**, the writer has the illusion that computer is thinking. The **MAXit Checker** includes some 9,500 grammar rules, based on style guides from public sources and the Chicago Manual of Style. Because the **MAXit Checker** uses

customized dictionaries with linguistic tags, the software can make sophisticated checks for terms like ExxonMobil and mHz.
The **MAXit Checker** can also disambiguate terms like db for decibels and database.
A metric conversion function helps authors express dual dimensions.

The next page is a picture of the **MAXit Checker** inside a WORD document. The colors, added by the **MAXit Checker,** indicate the possible errors in 40 message classes.
For example, the words marked in yellow must be examined for possible addition to the technical dictionaries. A like term "peak efficiency" is both difficult to define and translate. The writer is prompted for a definition, perhaps revolutions per minute, number of operating hours or mean time between failures. These numbers tell the reader more than just a vague expression. The controlled language authoring approach forces better quality information.

This table shows the **40 MAXit Message Classes** into which the 9,500 rules are divided by five colors. The writer soon learns that a GERUND is green and must be removed.

Abbreviation	Gerund	Number	Punctuation	Translation
Adverb	Hyphenation	Parallel	Missing Ref.	Verb tense
Adjective	Meaning	Passive Voice	Rephrase	Wrong Choice
Apostrophe	Measurement	Phrasal verb	Sent. Length	Wrong Verb
Awkward	Metric	Prefix	Spelling	Word Cluster
Capitalization	Name-Label	Preposition	Style	Change2Verb
Comma	Negative	Product Liability	Terminology	Change2Noun
Delete	Not-in-Dictionary	Pronoun	That-which	Long <para>



What is difference between ASD-STE100 Simplified Technical English and SMART Controlled English?

The SMART Controlled English is a more advanced version of ASD-STE100 Simplified Technical English (STE). For example, in STE there is no verb to "click" for the action of the computer mouse. In STE, the phrase is "press the left mouse button". Controlled English allows the writer to use "click the left mouse button". The reason why "click" is not yet added to STE, is because STE is a global standard and controlled by a committee where decisions are not easy to make. By contrast, **SMART Controlled English** was developed by SMART Communications so that a customer can customize the verbs.

Technical Nouns:
Both STE and **SMART Controlled English** allow companies to add technical nouns to their vocabularies. For example, in the packaging industry, the term Radio Frequency Identification tags (RFIDs) is important. An automotive company must have all the part names of an automobile. A software company must include field names and concepts.
To manage the terminology, SMART offers a simple tool called the **SMART Lexicon Manager**.

Step 3: Writer and Technical Author Training
In 2-3 days of training, Controlled English is very easy to write. The writers soon learn what verbs to avoid, how to write short sentence and the tricks. SMART offers either a 20-hour eLearning course or on-site seminars. The motivated writer can be productive in a week. For writers who use English as second language, the task of writing in **Controlled English** is easier than writing unstructured English because there are less choices in terminology, verbs and structure.

The Future of Controlled English - A New Global Language

The STE standard was developed to write aircraft maintenance manuals. **Smart Communications, Inc.** has developed more technical vocabularies for use in many products and services.

Today, SMART offers a **Controlled English** vocabulary that has a "core" word list of 1,200 words to which customized product terminology is added. All dictionaries and rules are easy to customize for a product line. Unexpectedly, China is fast becoming a large user of **Controlled English** to write service instructions for their advanced exports to new markets, like Western Europe.

The Benefits of Controlled English

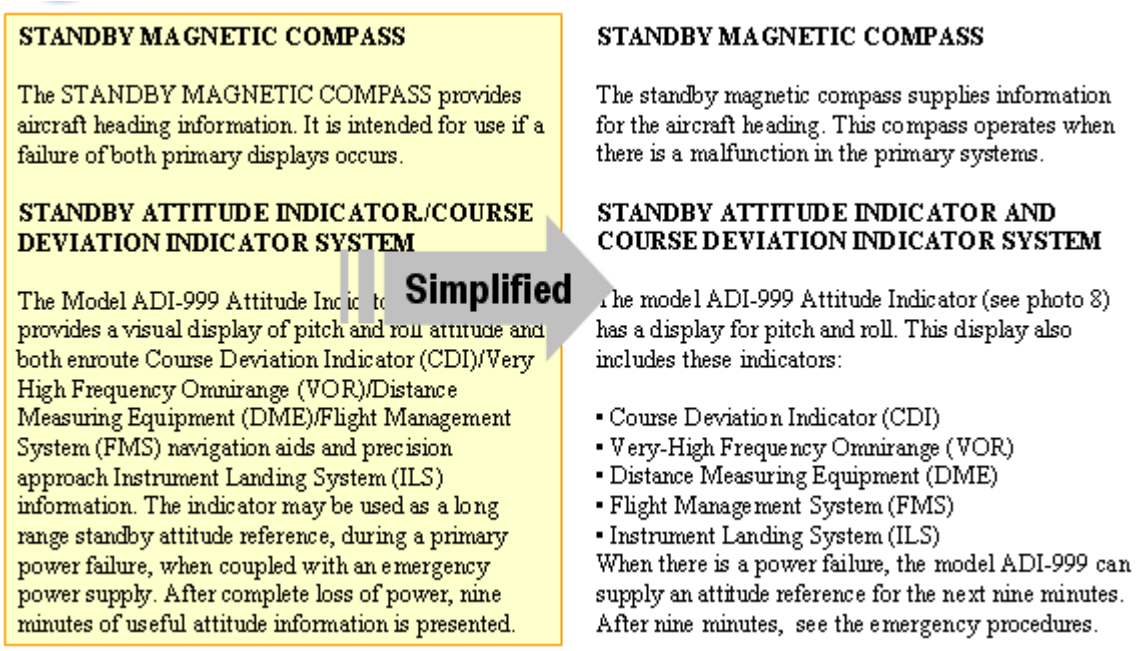
- Compliance to documentation quality standards, required under Six Sigma, ASD-STE100, ISO9000 and other standards.
- Creation of legally-defensible technical manuals, in clear, concise English to avoid product liability.
- Reduction in call center and product support costs through better manuals and web pages.
- Removal of ambiguity to create an ideal source language for translation into 80 languages by hand and computers.
- Faster training of technicians and customer support staffs in global markets.
- The use of a common language for worldwide customer communications.
- Preparation of easy-to-read help texts and work instructions.
- The use of short sentences to enhance text readability, comprehension and usability.
- Reduction in the cost and time needed to write technical documentation.

Controlled English is a Source Language for Machine Translation

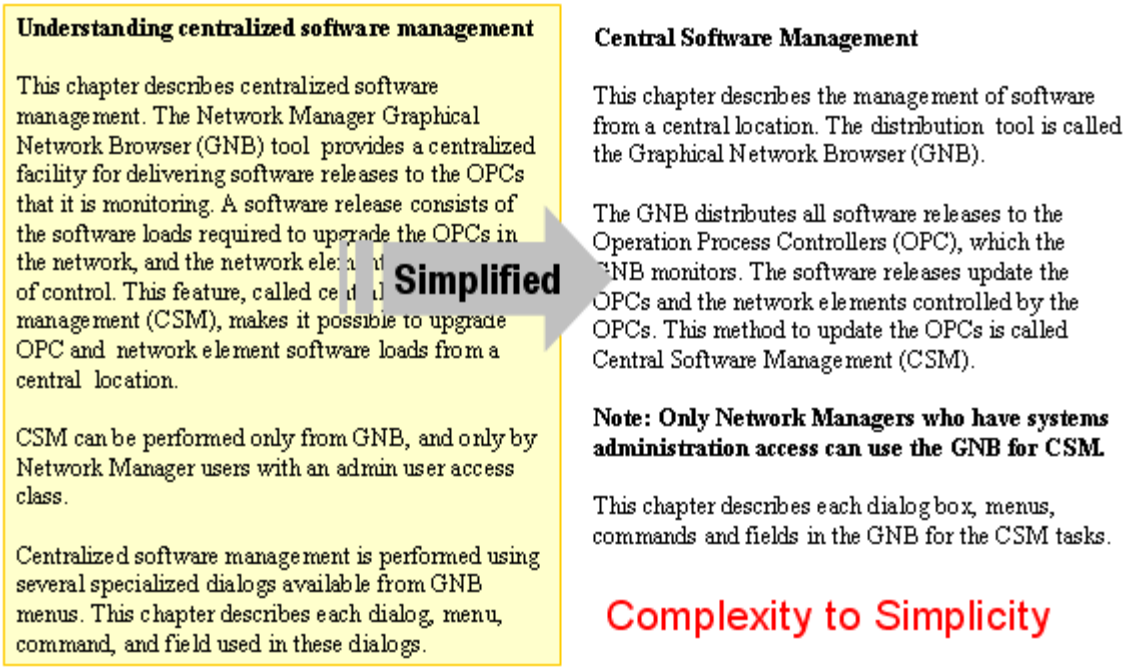
SMART has made studies that show that when the English texts are simplified, automated translation tools, like Systran, make the translations at a high level of accuracy with less than 5% post-editing. The reason is because the English texts are structured for machine translation. SMART sees further gains when the same technique is applied to the IBM Statistical Machine Translation (SMT) engine. To meet the growing demand for translations, we believe that **Controlled English** must be matched to SMT techniques in various domains.

Examples of Controlled English

ASD-STE100 Simplified Technical English
This example shows the original text on the left side and the simplification for global aerospace markets. Note the use of a bulleted list instead of a dense block of text.
The Simplified Technical English is easier to read, write and learn as a second language.

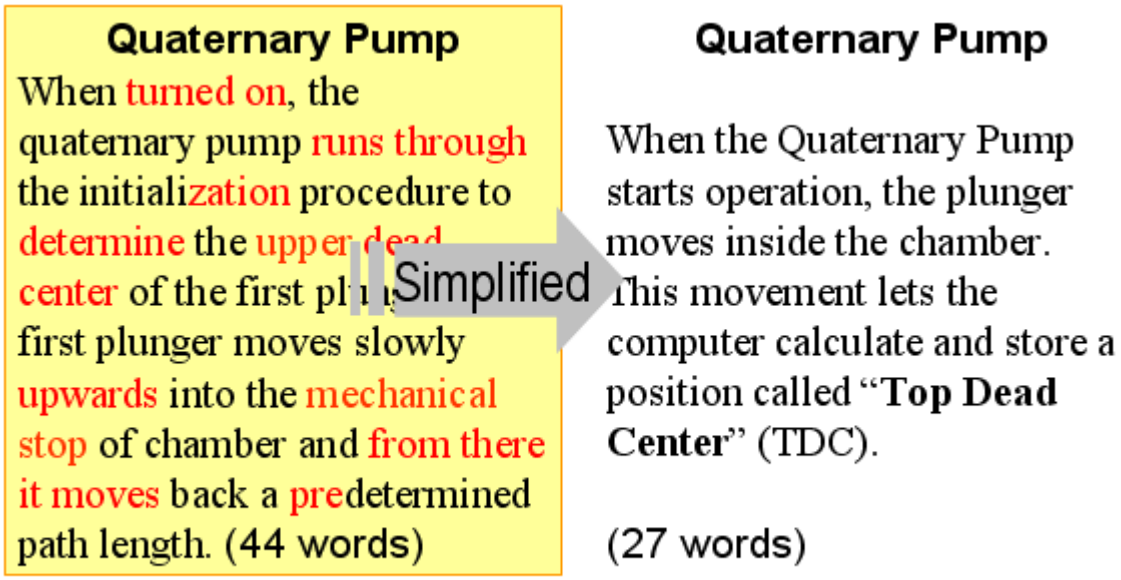


SMART Controlled English - Telecommunications Documentation
This example shows the original text on the left and the **Controlled English** for a telecommunications product on the right. In this example, the gobbledygook is removed and technical information is easier to find and comprehend.



SMART Controlled English - Medical Devices

This example shows the original text on the left and the **Controlled English** for a medical device on the right. In this example, the original is written by an engineer then simplified for a service technician. The **Controlled English** offers a 30% saving in text and later localization costs.



About the author:
John M. Smart, is the Managing Partner of the language consulting firm, SMART Communications, Inc., in New York City. The firm has 20 years of experience in text simplification, localization, language translation and globalization.
For more information contact: info@smartny.com