ORWELLIAN Linguistics

How using computer-aided translation tools impacts the target language.

by Else Mogensen

Else Mogensen

"The Revolution will be complete when the language is perfect." So says Syme, the philologist, in George Orwell's *1984*. Syme is a specialist in the language "Newspeak," and is part of the team of experts engaged in compiling the 11th edition of the Newspeak dictionary. He explains his work to Winston, the protagonist, in this way: "We're getting the language into its final shape—the shape it's going to have when nobody speaks anything else. When we've finished with it, people like you will have to learn it all over again. You think, I dare say, that our chief job is inventing new words. But not a bit of it! We're destroying words—scores of them, hundreds of them, every day. We're cutting the language down to the bone."

The technological revolution which is said to have taken place within the last couple of decades and is still in progress, has changed the way we translate, and as it turns out, it is also changing the language of the actual translation; it has an Orwellian impact on the target language as, I'm going to suggest, it contributes to "cutting the language down to the bone." Earlier, I would be asked to translate an English text, such as the following, into Danish: "Do not use solvent or gasoline to clean the warning signs. Solvents and gasoline could loosen the adhesive that secures the warning sign. Loose adhesive will allow the warning sign to fall." Back then, I was free, and indeed expected, to translate it into one full sentence, namely with a main clause and two subordinate clauses, because the content makes up just one single complete statement, and Danish readers would easily comprehend the cause and effect set up in this complete sentence.

Enter Yellow Language

Now, however, when I use computer-aided translation tools, and for that tool's database to be useful in translation of that text, the translation has to follow the English syntax, and should preferably be divided up into a text with three periods—three declarative main clauses—just like the source text. The resulting translation appears oversimplified, what the Danish linguist Professor Kirsten Rask calls *machine-gun sentences* and terms *yellow language*.

Professor Rask compares this simplified language to film edits that are made short to create an impression of activity, speed, and excitement, as each sequence of words lasts two to four seconds, period. She detects that this same tendency is invading the Danish language, as more and more writers are cutting up their texts into short sequences by means of periods, to appear fast, striking, and sensational. She concedes that although short sentences are not bad in and of themselves, using only short sentences results in oversimplified syntax and a monotonous rhythm lacking the liveliness provided by a diversity of long and short sentences—and lacking, as you might surmise, the spark of a normal speaker.

Computer-aided translation tools are now used extensively by technical translators. They help translators manage their material and make it easier to find and use translations of sentences and words that have already been translated (irrespective of current context), eliminating some of the monotonous work of technical translation. Databases containing translations together with their equivalent in the source language are being built while the translator is translating, so that identical sentences and phrases do not have to be typed in over and over again. The program can then search for and find identical sentences—whereby *sentence* means replicas of text (identified by sameness of spelling) from one period to the next—and by the help of so-called fuzzy searches find "similar" sentences in the database, presumably faster than a translator can think.

Economic But Not Linguistic Sense

It makes perfect sense (perhaps in economic terms), to reuse translations in this way in technical translation, because as it happens technical texts contain an enormous amount of repetition. When a new model of a product is introduced, it is usually very similar to its predecessor, with only a few new changed or added features, and the new accompanying manual is often not a rewrite, but merely a modification (or version, edition, etc.) of the old one, plus inclusion of the new features. Furthermore, it seems that machinery is "updated" and "modified" at a faster and faster pace in this day and age, which means that new versions of manuals have to be translated at a faster and faster rate.

The rather constrained environment of the computer-aided translation tools, which as a rule work with linguistic elements between periods (the program registers the words from one period to the next as a unit or separate entity), has led to an engineering or controlling of language, and not just improvement of the technology to make the programs work better, or to make them do what you would want them to do.

Ever since the notion of an "ideal language" was developed, the concept of "controlled language" never lagged far behind. The idea of controlled English emerged in the 1930s. There have been numerous attempts to design a universal language that would unite humanity and further communication in the world. But this was different from a number of other attempts such as Esperanto, Ido, Volapük, etc., in that it was *English*, and not a completely artificial language or a hybrid language constructed from several different existing natural languages.

This language was called Basic English and was in essence a simplified English. One of the ideas behind it was "to cut the language to the bone." In other words, it was deemed possible, in fact, to restrict the vocabulary to 850 words as well as simplify the grammar, and still talk and write about important issues of the world, instead of having to use or know some 75,000 English words that a fluent and well-educated speaker of English might otherwise command. As Orwell's philologist said, "his thin dark face had become animated, his eyes had lost their mocking expression and grown almost dreamy. 'It's a beautiful thing, the destruction of words.'"

Simplified, or controlled language as it is now called, produced a much better result with machine translation and machine-aided translation tools than our ordinary, ambiguous, and varied language. And the point is almost self-evidently true. The less ambiguous the meaning and syntax, the easier it is to engineer a system or program that has to deal with exactly those problems. "One word, one meaning" is its basic principle. In practice it means that the verb *to advise*, for example, can only be used in the meaning *to give advice*, so to use it in a sentence such as "Please advise us of the availability of parts at your earliest convenience," would not be allowed because here *advise* means *inform*. Orwell's philologist says it clearly in his explanation:

> "Of course the great wastage is in the verbs and adjectives, but there are hundreds of nouns that can be got rid of as well. It isn't only the synonyms; there are also the antonyms. After all, what justification is there for a word which is simply the opposite of some other words? A word contains its opposite in itself. Take 'good' for instance. If you have a word like 'good,' what need is there for a word like 'bad'? 'Ungood' will do just as well—better, because it's an exact opposite, which the other is not. Or again, if you want a stronger version of 'good,' what sense is there in having a whole string of vague useless words like 'excellent' and 'splendid' and all the rest of them? 'Plusgood' covers the meaning, or 'doubleplusgood' if you want something stronger still. Of course we use those forms already but in the final version of our Newspeak there will be nothing else. In the end the whole notion of goodness and badness will be covered by only six words—in reality, only one word. Don't you see the beauty of that, Winston?"

Now, however, when I use computer-aided translation tools, and for that tool's database to be useful in translation of that text, the translation has to follow the English syntax, and should preferably be divided up into a text with three periods—three declarative main clauses—just like the source text. The resulting translation appears oversimplified, what the Danish linguist Professor Kirsten Rask calls "machine-gun sentences" and terms "yellow language."

Does the Linguist Control the Tool, or Vice Versa?

In other words, language is now being changed to fit the tools, instead of the other way around. When you translate in a translation program, you do have some control over the tool. It normally allows you to expand a segment to include the next segment, that is, the text between the next two periods, making you free to combine the two segments into one complete sentence with appropriate subordinate clauses if you so choose. But you are restricted to expand with one additional segment, and aside from the fact that you are syntactically constrained, my experience is that technical translators, myself included, tend to use the tools the way they were designed to be used, most often producing texts with a syntax close to that of the source text.

Controlled language has invaded the corporate world; companies create their own standards for language use. The motivation, we are told, is that controlled language counters the tendency of technical writers to use jargon and unusual styles and grammatical constructions or inconsistent language. AECMA Simplified English, for example, was developed to facilitate the use of maintenance manuals by non-native speakers of English. It limits the length of instructional sentences to 20 words (as do Microsoft's grammar checkers, by the way), and requires that sequential steps in an activity be expressed in separate sentences.

This has been adopted by an entire business sector, the aerospace industry. Aerospace manufacturers are required to write aircraft maintenance documentation in such simplified English. Another controlled language, PACE, has been developed by the British company Perkins Engines. (PACE stands for Perkins Approved Clear English.) The American company Caterpillar has its CFE (an acronym for Caterpillar Fundamental English).

Controlled Language: Outgrown Its Mandate?

Now, however, it is no longer used only for its original purpose, namely to make manuals easier to read by people whose native language is not English, but also for the benefit of the translation industry, making it supposedly easier for a text to be translated. And one difference now is that it no longer concerns just the English language, but via translation it applies to the rest of the world's languages as well.

Although the Danes to a certain extent embraced artificial languages such as Volapük and Esperanto (the well-known Danish linguist Otto Jespersen even created one called Novial) there have not, as far as I know, been any attempts to create a simplified or basic Danish language. But by using computer-assisted translation tools, as they are configured at the present time, a simplified Danish is emerging. Danish technical manuals have traditionally been written in an impersonal style with complete sentences, telling the user what to do in order to accomplish this or that. This style is quickly being replaced by short commands in the imperative verb form without the explanatory connections that flowing text provides.

There have not, as far as I know, been any attempts to create a simplified or basic Danish language. But by using computer-assisted translation tools, a simplified Danish is emerging. Danish technical manuals have traditionally been written in an impersonal style with complete sentences. This style is quickly being replaced by short commands in the imperative verb form without the explanatory connections that flowing text provides.

Supposedly, controlled language improves readability and comprehensibility by reducing lexical and structural complexity (such as ambiguity) and specifying standards of style. That economizes maintenance, since a text that is easier to read and understand is also easier to update. It makes the computational process easier due to the reduction of lexical and structural complexity (again, such as ambiguity, etc.) and prescription of stylistic rules. In short, it makes machine translation feasible.

Searching for Quantitative Measures

A method of measuring quantitatively a text's ease of readability, the so-called LIX index, was invented by the Swedish linguist C.H. Bjørnsson. LIX is the acronym for *läsbarhedsindex* which in English literally means *readability index*. It delineates the degree of compression the language of a text has, thereby indicating how difficult it is to read and understand (to some presumed average reader). Bjørnsson let various groups of people read 340 different texts and found, not surprisingly, that what makes a text difficult to read and understand is a large number of words between periods and, also, a large number of long words. And he established an equation for quantifying a text's degree of readability (how hard it is to read, as many would put it), namely the *number of words per period + the percentage of long words = LIX*.

But dividing or cutting up a text into short declarative sentences also removes the connectives between them, and when connection disappears, content disappears as well. The text, in fact, may become more difficult to read and comprehend. A typical example of a technical text might be the following:

> *The seawater that is circulated through the heat exchanger and the amount of operating time of the vessel affects the following items:*
>
> • *Cleanliness of the tubes of the heat exchanger.*
>
> • *Effectiveness of the heat exchanger system.*

Because of the bulleted format, which actually means that this short text is divided up into three separate paragraphs, and in a translation program into three separate segments, the translation will typically be done in the same format, i.e., in three distinct (unconnected) segments.

However, a natural rendering in Danish of the same information would have been given in a principal clause with an embedded clause, followed by two subordinate clauses that are connected with a causal connective, such as "consequently" or "therefore." In effect, you get less information in the bulleted text as the syntax of the sentence also contains or yields information and affects comprehension. Bear in mind: "Don't you see that the whole aim of Newspeak is to narrow the range of thought? In the end we shall make thoughtcrime literally impossible, because there will be no words in which to express it," says Syme, Orwell's philologist.

Questioning the Basic Assumption

The contention is that consistency in the use of terms will help to improve the overall consistency and quality of the text being translated. After all, one of the simplest and most direct benefits of computer-assisted translation tools in technical translation is that given terms are always translated consistently because they are

It is, thus, a misunderstanding that short sentences necessarily are easier to read than longer sentences, because there is a lower limit to brevity and conciseness (which is mistaken for simplicity). If sentences are too short, we inhibit the reading process, let alone the comprehension process, because we hinder the flow of information, thus diminishing the fluency of comprehension. The reader loses the sense of context.

taken to be bilingual synonyms in the electronic dictionary. We get consistency. As the Orwellian philologist maintained:

> "Every concept that can ever be needed will be expressed by exactly one word, with its meaning rigidly defined and all its subsidiary meanings rubbed out and forgotten. Already, in the Eleventh Edition, we're not far from that point. But the process will still be continuing long after you and I are dead. Every year fewer and fewer words, and the range of consciousness always a little smaller."

And this prediction or omen seems to be the rule or imperative for Rask's *yellow language*, a term derived from *yellow journalism* — the news industry of daily papers that, in an intrusive fashion, "covers" peoples' private lives, ephemeral events in momentous terms—in other words, sheer sensationalist writing. The designator *yellow* stands for hyperbolic, one-sided, superficial, intentionally simplified use of language.

But as Professor Rask points out, a sentence is not always equal to what is between two periods. The definition of a sentence is a verb with attendant qualifiers, and it continues until no more is associated with that verb, no matter how many periods it takes. You need a complete sentence to express a complete meaning, and that may be one or more principal clauses or a principal clause and any number of subordinate clauses. A long sentence is difficult to read when it has numerous dependent clauses before the main verb, whereas a long sentence may be easy to read even with many dependent clauses following after the main verb.

It is, thus, a misunderstanding that short sentences necessarily are easier to read than longer sentences, because there is a lower limit to brevity and conciseness (which is mistaken for simplicity). If sentences are too short, we inhibit the reading process, let alone the comprehension process, because we hinder the flow of information, thus diminishing the fluency of comprehension. The reader loses the sense of context. The staccato effect of oversimplified sentences inhibits the reading process and the sense of sequence, although they seemingly present clear facts.

Putting the Tools in Context

Computer-aided translation tools are here to stay (if not in their present infantile forms), and they are interesting and challenging to work with, but we have to realize that by using them we contribute to changing the target language syntactically as well as limiting the range of vocabulary within the area we are working. We want consistency and readability in both source and target languages of technical documentation, but is it desirable to "cut the language to the bone" to obtain this?

Designers of translation tools need to make room for cross-linguistic differences in order for us to hold on to not only semantics and syntax, but also the overall grammatical characteristics of the target language. *Yellow language* has invaded newspapers and pulp fiction as well as advertising and other commercial texts in Denmark, and currently we, the translators, aid and abet the process because the tools we use make it too easy to ignore the syntactical differences between the source language and the target language. We have to look at the languages as well as the tools so as not to end up with languages cut to the bone.

Else Mogensen, is a native of Denmark and holds a Ph.D. in classical philology. She is an in-house senior translator and editor at Eriksen Translations in Brooklyn, New York.