GET SMART!

Industrial Strength Language Processing from Smart Communications.

Text by Jeffrey S. Mann
Photos by Dyana Van Campen

Every year multinational corporations around the world put out hundreds of millions of words designed to instruct, inform, warn and convince. The ability to achieve each of these goals effectively can mean the difference between chapter eleven and Easy Street.

If manuals are not clear and understandable, they do more harm than good. They not only have to be understandable by someone around the corner, but also around the globe. Your Canadian customer might even demand manuals in his own language and maybe you don't speak French.

A small company in New York has been helping the huge multinationals with their language problems for well over a decade. Smart Communications Inc. provides economical on-line help for writing clear, safe documentation and translating it into several languages. This small company has been using artificial intelligence (AI) techniques since 1972 to make text analysis and machine translation software.

Over 30 companies are using SMART software, including Citicorp, Chase, Ford Motor Company, three divisions of General Electric, Hyster (the forklift manufacturer) and the Canadian government (see accompanying article).

The emphasis is on concrete results for industry. "We only deal with big documentation projects," notes John Smart, president of Smart Communications. "At least a million words, which adds up to four Love Story's. We don't deal with the corner grocery store or individuals. It's strictly big projects for us."

The SMART AI products are aimed at the densest, most turgid prose being produced in

the world today. Smart Communications's customers use them to write installation manuals, maintenance guides, operating instructions for heavy machinery and other types of writing which no one likes to read, but must be understandable.

SMART COMMUNICATIONS INC.

Smart Communications is a very small company, just eight people. They build very specific, tailored solutions for their customers. There are no off-the-shelf consumer products in the SMART catalog. In fact, there's no catalog. With two products, that's hardly necessary.

Smart sells the "SMART Expert Editor" and the "SMART Translators." Both products are built around an expert system designed by the team at Smart Communications. An Australian who came to the United States via Viet Nam and a Rockefeller foundation grant, John Smart provides the computer science and artificial intelligence background. He and his programmers worked together with linguists and translators in the New York area, notably Albert LaMothe.

They distilled the rules and concepts recognized by translators to form the basis for the SMART software packages. The approach is not research or academic-oriented. "We've got no academics, because they all said it couldn't be done," said Smart.

"The 'warmware' is an important part of any program," says Smart, which is why each installation is specially tailored. It takes around three months for an installation to be customized for a particular application. That's the time it takes to build up the specialized databases and tune the programs to the individual company. Some of the staff at Smart Communications also teach technical writing and how to use the SMART system.

THE SMART EXPERT EDITOR ("MAX")

The first product in the line is the SMART Expert Editor. This is a batch-oriented text analyzer which uses a rule-based expert system and specialized terminology knowledge bases. The rule base contains 2,500 generalized grammar and syntax rules for technical writing. For example, subjects and verbs

---

SMART MACHINE TRANSLATION IN USE: THE CANADIAN MINISTRY OF EMPLOYMENT AND IMMIGRATION.

The Canadian government is Smart Communication's biggest customer. The Ministry of Employment and Immigration has 5000 terminals which can use the English-French and French-English translators. They have been using the system since 1982. "It paid for itself within six months," according to Eric Davies, director of advanced systems at the ministry's office in Ottawa.

The Canadian installation is customized to translate descriptions of job vacancies. Clerical workers across Canada type in 180-200 character job descriptions in either French or English. The SMART Translator converts it to the other language within five seconds. Over 100,000 descriptions per year are processed on Unisys mainframes located in six regional centers across Canada.

"We're in a unique situation in that we aren't looking for perfect French or English. 90 per cent is acceptable, though we do get complaints, of course," said Davies. If the typist happens to be bilingual, he or she can post-edit the translation to iron out any imperfections. "In the Toronto area, it's pretty much a black box which puts out unedited French text," he admitted, "due to that area's paucity of French speakers."

The system does make mistakes. Davies recently received a complaint where the abbreviation for Manitoba ("MAN.") was translated into French as "homme." The typist forgot the period. It has, however, learned the difference between someone who installs "Venetian blinds" and "blind Venetians," an embarrassing mistake made in an early installation.

Officials in Canada worked closely with SMART to tailor the system for this particular application. They maintain the knowledge base themselves, adding words and contexts judged to be important. As it processes a piece of text, the translator flags any terms it doesn't recognize, to be corrected later. Once a month they look at the list of flagged words and add any terms used more than ten times to the knowledge base, which currently contains around 10,000 words.

"We are considering expanding the project to translate the documentation for the computer systems developed here," he said. "That would require developing a new knowledge base, however." The system they now have is specifically customized to translate job descriptions.

should agree. These rules form the fuel for the inference engine at the heart of the software. The knowledge base is a set of vocabulary specific to a particular application.

Smart currently has 22 different knowledge bases, such as automotive engineering, aerospace technology, construction, and computer programming. The computer terminology database will know that the word "run" refers to starting a computer program. In a knowledge base for the textile industry, that same word can mean a defect in a woman's stocking. The knowledge base includes "real world" knowledge as well. Like the fact that gasoline is dangerous and should have big warning labels.

The program essentially does what a good copy editor does, but much faster (about one page per second) and more consistently (you can't sneak things through late on Friday afternoon). After writing several pages of text, the writer runs his document through the program. The editor generates a report which prints each line of text from the document, and directly underneath, comments and suggestions.

Each suggestion refers to a message manual where the particular point of grammar or usage is explained in more detail. Where possible, the program makes specific suggestions about how to improve the text. After processing, the writer takes the report and uses it to improve the original text.

The Editor has its limitations, of course. Outside of technical applications, it's not very useful. If the knowledge base doesn't know what you're talking about, the critique won't be very good.

Because the Editor seems to have a personality, it's been nicknamed "Max." The nickname is based loosely on Maxwell Smart the know-it-all hero in Get Smart, an American situation comedy broadcast in the 1960s. See the accompanying article for a sample critique from Max.

An example of its technical orientation is the check for sexist references. Max flags every occurrence of "his" or "he." In technical documentation, "After the operator opens the hatch, he should secure the pipe" is sexist. Who says the operator must be a "he?"

However, Max doesn't know if "he" really does refer to a male person, such as "Johnson opened the hatch before he secured the pipe." The knowledge base could probably be updated to recognize most of these situations, though with that sentence, it's not clear whether Mr. or Ms. Johnson secured the pipe.

LIABILITY

In addition to checking grammar and usage, Max scans documentation for references to dangerous procedures and substances, and suggests how they should be labelled. A good lawyer can turn vague or unclear documentation into a multimillion dollar liability award.

"In the United States, a technical manual is legally considered part of a piece of equipment. As a part, it is subject to product liability — and it can fail," Smart points out. "A lot of our business comes from insurance companies looking to contain their risks."

Max has been used in a courtroom to show jurors that the manufacturer has taken reasonable efforts to ensure that their instructions are understandable. "Using this program shows jurors that the company made that extra effort to make their instructions understandable," Smart remarked.

"We have been in courtrooms where attorneys projected the word 'should' on a big screen and then handed out dictionaries and had the jury look it up," Smart recalled. "It's all over then. The poor guy reads something like 'The valve should open' so he opens it, and it blows his face off. That 'should' will stick in jurors' minds."

A famous example of product liability involves the Dana Brake Company. Their user manuals described their brakes as "failsafe." A multimillion dollar settlement proved that nothing is failsafe. Max flags that word whenever it encounters it in any publication. He remembers that it caused a problem over a decade ago.

Not just money is involved, of course. The investigators looking into the Bhopal toxic gas leak and the Zeebrugge ferry disaster are looking at documentation and training practices to see if shoddy language might have contributed to these catastrophes.

THE SMART TRANSLATORS

An added benefit of using Max is that text becomes easier to translate. "Sprinkling commas like confetti makes text hard to translate," said Smart. "When you prepare a document with the SMART Expert Editor you've already prepared it – accidentally – for translation. Many of the things which cause problems in translation also cause problems in human understanding."

"We got into translation more or less by accident," Smart says. "We were working with companies to improve their documentation with the Expert Editor, and then they asked if it could do translations too."

This need becomes particularly acute as countries pass laws requiring that foreign companies provide technology transfer documentation in the host country's language. The translator takes a piece of text, chews it over and spits out the translation within a reasonable time. Any terms which it does not recognize are left as is, and flagged in the text. Smart stresses that the translations are not perfect, about 90-95 per cent accurate.

"The effort and cost involved in catching that last 5 per cent are not really worth it," he feels. After processing, a post-editor is usually called in to clean up any problems. See the accompanying article for a description of how the Canadian Government, Smart's biggest customer, uses the translators.

The Expert Editor program is not required to use the translators. However, cleaning up a document before translation results in better translations which require less post-editing.

Translators are currently available for French, Spanish, Portuguese, Italian, and a bit of German. They've started work on a Kanji version (Japanese). All translations are between English and these languages. Smart has no plans to implement cross-pair translators, such as French-Spanish.

FUTURE DIRECTIONS

Smart sees a great number of possibilities for these products, as they stand and as a start for many other products. The obvious possibilities for expansion are more language pairs for the translators. Until now, the emphasis has been on the North and South American markets. That's why French, Spanish and Portuguese translators are available now. They are also exploring the possibility of bringing out "Le Max," a French language analyzer.

John Smart hopes to start Smart Communications Ltd. in England in the near future so as to crack the European market. Right now Smart has no European customers. Development of the Kanji module would provide access to the huge Japanese market.

He also hopes to develop a multilingual telex and message writer based on the "kit-set" analogy. Like a hobbyist builds a model airplane using a kit of standard parts, this package will allow a businessman to write a letter in any language by selecting the parts needed.

"They are model letters where you plug in the date of the aircraft, the name of the corresponding bank, the sending bank, the amount, the problem, etc.," says Smart. "Most business communications actually use quite limited language. This tends to standardize language so that a businessman can correspond with a person in another language."

There is also the possibility of developing a more interactive user interface. The products now chug all the way through a document and write a report. The Japanese are particularly interested in working with pop-up windows which show each sentence as it is translated and allows the operator to make changes, or allows the operator to choose a phrase from a list and zap it into a document. Smart remains skeptical about whether that is what users really want, however.

SMART will continue to aim at multinationals who need to process millions of words. It will be quite a while before a SMART product is used on a kitchen table, if ever. Smart doesn't see much potential in that market.

No matter what changes they make, Smart Communications will remain fairly small. "If Logos claims to be the IBM of machine translation, we intend to be the boutique", Smart said. They will continue to work closely with their clients and build tailored solutions for specific problems. "Some clients will want more specialized service and be willing to pay a bit more for it." The privately-held company has made a profit since the first year in 1972. Their turnover this year will be around US$1 million which Smart hopes to boost to US$10 million within a few years.

But making more money is certainly not Smart's only motivation to grow.

"One day," Smart confides, "I want to see the language business as big as the dogfood business."

Jeffrey Mann is a technical writer and consultant living in Amsterdam. In addition to writing documentation for 4th generation software, he is still not writing a novel.