

## Cap Volmac's Tailormade MT Systems

**Exploiting restricted language is the key to success in Cap Volmac's machine translation systems. One-hundred percent accuracy with no post-editing is the reward for imposing stringent controls on source texts.**

Given that fully, automatic, high, quality translation is not likely to be achieved in our lifetime, what are the alternatives? One is restricting source text in terms of grammar, style, and vocabulary. Various organizations exploit restricted English for translation purposes – notably Xerox, Bull, and Perkins Engines – but only one company currently builds such systems for third parties. Lingware Services of Cap Volmac (Utrecht, the Netherlands) has developed custom MT applications for a number of European customers in the fashion, insurance, automotive, and pharma, ceuticals branches. In addition, Volmac has built a correction and verification system to enable aircraft manufacturers to deliver AECMA Simplified English-compliant maintenance manuals. And suitably enough, a Lingware application is also used internally within Cap Volmac's Benelux organization for translating the curriculum vitae of its personnel from Dutch into English.

“Cap Volmac is not in the business of developing products,” emphasizes Lingware's Gert van der Steen. “That's simply not part of our corporate culture. We are a software house, and hence we develop tools for building applications for our customers.” Being part of Europe's largest systems integration house means that Van der Steen and colleagues have had to acquire a project-oriented, software engineering approach to building applications. System integration work is characterized by an accurate evaluation of needs, available resources, and careful planning; building custom MT systems is no different. Fortunately, the Lingware group can build on the grammars and lexicons developed in previous applications; each new system is a matter of extending existing work. A key element is being able to gauge quickly whether a company's texts can be handled by the Lingware grammars and exactly how much work it will take to expand the coverage of the Lingware system for a given application. Van der Steen can calculate just how much time it will take to code a new verb, noun, or adjective, or a new grammar rule. Van der Steen points out that his colleague Bert-Jan Dijenborgh's two-year stint with Eurotra Nederland has stood him in good stead by enabling him to quickly analyze text corpora prior to developing a new application.

The Volmac Lingware tools include a DOS-based, restricted language editor (for Dutch and AECMA Standardized English) and a translation system (DOS and Macintosh) which has been deployed for Dutch-English (both ways), Dutch-to-Spanish, Dutch-to-German, and English-to-French. The basic tools are driven by grammars, lexicons, and thesauri which are customized for each application. Restricted language is not universally applicable and Van der Steen acknowledges that there can be a certain amount of resistance to it among technical writers. But for companies willing to take the plunge, the rewards are worth it: Volmac maintains that proper correction and standardization of texts means one-hundred percent correct translations can be produced – that is, no post,editing is required. Running on DOS machines, the Lingware editor can correct and standardize texts faster than the user can type; the translation module is rated at 50,000 words per hour. For a number of applications, Volmac Lingware has implemented SGML for the

underlying file structures.

The Lingware services have been largely marketed by Cap Volmac's 125-strong sales force, who present it to the company's customers as one service of many in the company's portfolio. However, Van der Steen says that the Lingware group has grown to the point where it can handle multiple projects simultaneously and may soon be pitching their wares directly to documentation departments, who may as yet be unaware that such systems are indeed feasible and who need to be acquainted with restricted language, not yet a universally understood concept. On the other hand, while Cap Volmac management is naturally happy to see the group attract external projects, there is also a considerable internal market for Lingware applications within the Cap organization. In any event, being part of such a large organization has its advantages; Van der Steen points out that external customers are also concerned about the supplier they are dealing with. "Customers want to be sure that they can get support and upgrades in three year's time," he points out. This places large enterprises like Cap Volmac at a natural advantage.

Van der Steen joined Volmac in 1990 as part of Vleermuis Research, a five-year research program launched by Volmac in the 1980s which encompassed research in NLP, CASE, neural networks, and office automation. The Lingware group emerged from Vleermuis and is now an integral part of Cap Volmac services. The Lingware group is no longer engaged in research; officially speaking, Cap Gemini Innovation (Paris and Utrecht) is the research wing of the Cap empire. The latter is incidentally one of the members of SECC, a new project in the second round of the LRE program which will develop a grammar and style checker for restricted English based on the METAL parsing engine. Lingware is not participating in this project, because as Van der Steen puts it, "Community funding brings a lot of red tape with it." Moreover, he feels it is not to Cap Volmac's competitive advantage to participate. However, Van der Steen is a member of the industrial interest group which is being set up around SECC.

Devoted to computational linguistics for many years, Van der Steen has not always been able to make a living from it. The Cap Volmac Lingware group is an ideal arrangement for him, Dijenborgh, and their colleagues to pursue their interests – and see concrete results. "The big research programs like Eurotra are good," he says, "provided that the people who are involved can later find jobs in this business. Unfortunately, I know of a number of people who have had to find work in completely different fields."

Cap Volmac Lingware Services, Postbus 2575, Utrecht, 3500 GN, The Netherlands; Tel: +31 030 526148, Fax: +31 30 543244