

The Sietec Connection

The Metal development group at Sietec may not be the largest or most heavily endowed industrial NLP development group in Europe but it is one of the most enduring.

For many years, the linguistics team at Siemens toiled away in comparative quiet, turning Metal from the promising German-English university prototype Siemens acquired from the University of Texas in 1980 to one of the world's best MT systems, one that enjoys a small but growing contingent of faithful users.

These days, however, things are perceptively changing. The group, now known as Linguistic Systems, has been moved from Siemens to Sietec Systemtechnik, the long-time Austin development site has been closed, and the development team has completed the arduous task of porting the Metal system from the now obsolete Symbolics LISP machine to the Unix environment, making it easier to develop new languages pairs. And Metal is now being more aggressively marketed, both in Europe and in North America. At Sietec Systemtechnik, a systems integration and service group within Siemens-Nixdorf (SNI), Metal and its progenitors are closer to prospective customers and further away from R&D environs, where budgets are exceedingly tight these days. Like many of Europe's venerable IT companies, hard times have fallen on SNI; it racked up losses of DM600 million last year; however the Sietec Systemtechnik wing is profitable.

With no immediate prospect of directly recuperating the estimated DM50 million invested in Metal through sales of the system, Sietec is now actively looking for other ways to exploit the Metal technology. Says Thomas Schneider, director of the Linguistic Systems group, "we're now setting our sights beyond the small and very difficult translation market." Sietec would like to see the Metal engine deployed for a variety of language processing applications, from high end full-text information retrieval, message routing, and fact extraction systems, to end-user tools, such as grammar and style checkers, extended spelling checkers, and translation memories. With the substantial experience of SNI in large systems development behind it, the linguistics group is in an excellent position to develop custom-built NLP systems for large customers, and Schneider hints that negotiations are well under way with one very large prospective German customer to build an information retrieval system based on Metal technology. Another intriguing direction is that of Controlled English applications. The Metal group in Liège (Belgium), which developed Dutch-French system at the behest of the Belgian government, is prime contractor in an EU LRE project called SECC of which the goal is to develop a controlled English checker based on Metal, a tacit acknowledgement that Sietec has one of the best English parsers in the industrial world. It may seem odd that the Commission is funding the kind of development work which has already been done by other groups in industry but it certainly is one way of be assured of getting results. However, as always with the Commission's cost-shared funding efforts, it is never exactly clear who will be allowed to exploit the fruits of such an endeavor.

Large customers, however, are comparatively few and far between, so Sietec is also looking at ways of deploying its technology in more modest arenas. Off-the-shelf packages are, after all, the bread-andbutter (and the future) of the software industry. Here Sietec has less experience. Four years ago, it launched a terminology management package for the PC, called Term-PC. It was flawed in a number of ways and suffered from Siemen's lack of a clear distribution

channel for small packages. It was withdrawn a short time later.

A more recent step in the right direction is LingTools, a Unix-based client/server package which Sietec has introduced as a successor to TermPC. LingTools is a set of four modules for manipulating and managing terminology. LexikonAbfrage is a look-up tool which enables users to develop their own glossaries in conjunction with standard lexicons. Lexikon TextVergleich is a tool for comparing a text against a lexicon. It flags terms which are used, thereby generating a glossary, it flags terms which may be used incorrectly, and it flags unknown terms. IndexGenerierung automatically indexes a text, and LexikonAdministration is a tool for creating and maintaining terminology lists. Sietec supplies a number of supplementary trilingual (German, English, and French) lexicons for LingTools, including general (ca. 80,000 entries) and those for trade-finance-law (ca. 70,000 entries) and technical (ca. 37,000 entries) domains. These can be used in conjunction with a definition dictionary (ca. 25,000 entries) and a synonym dictionary (ca. 45,000 entries) for German. At the moment, only a Unix/Motif version of LingTools is available, but a Windows client is currently being beta-tested and a Windows server is under development. A modest application of Metal-based linguistic technology, LingTools is an attractive package which, when available for Windows and pitched at a modest price, will be a welcome addition to the wordcrunching arsenal of translators and technical writers.

As an obvious growth path for Metal, Sietec has been extensively involved in the EuroLang project, to which it has dedicated some twenty engineers, thereby making it Site's major industrial partner in the undertaking. Although the partnership has had its ups and downs, partly due to substantial changes in direction, Sietec representatives give every indication of being solidly behind the EuroLang Optimizer package. Initially – and up until quite recently – the consortium spoke of building a “EuroLang MT System,” with Metal as the translation engine at the system's core. This would have been the upgrade path for both current Metal users and new customers. However, Sietec and Site have tempered their collective ambitions somewhat. Sietec will continue to develop Metal in its current form and offer the system as an optional back-end to Optimizer, while Site will offer Optimizer as a frontend to Metal and other MT systems. Which ever way the cake is sliced, Optimizer remains of strategic importance to Sietec. While Metal boasts probably the most powerful and sophisticated userinterface of any existing MT system, it does not have the kind of facilities which make it easy to integrate into the modern notion of a “translator's workstation,” namely, seamless “desktop” integration of MT with translation memory and terminology databases and other productivity tools. In theory, Optimizer is Sietec's path to this idealized solution. In practice, however, it will not be as closely integrated as it would have been possible with Metal fully integrated into the EuroLang system as originally conceived. That “nirvana” will have to wait.

Responsible for sales and marketing of linguistic systems, Sietec's Rudolf Thiem is intimately aware of the challenge of marketing a system like Metal over the years. For a start, he notes that MT is a “support-intensive” application. He points out that Sietec is careful with the timing of its new releases, and, for example, takes care not to overwhelm a new user with a new language pair. Thiem is sceptical of colleagues who try to low-ball MT systems into the market by undercutting their prices. Short-term gain may be offset by the high cost of long-term support, he warns, and Thiem gives every indication that he is speaking from hard-won experience.

Thiem also points out that it can take a long time to win a new customer, particularly within large companies or organizations, where the decisionmaking process can be glacial. A recent example – perhaps a bit extreme – is the Swiss Air Force, which will be using Metal to translate 70,000 pages of maintenance manuals. As Thiem recalls, contacts dated from as far back as 1989; however, serious negotiations began only after a popular referendum upheld the purchase of the planes. According to Thiem, at least half of Metal's users are now outside Germany. The past year has seen three new customers in North America, where the system is currently being marketed by Sietec Open Systems (Canada).

Other directions notwithstanding, further development of Metal will continue to be a major focus for Sietec. New language pairs are under way, including Catalan at Sietec's former subsidiary in Barcelona, now called Incyta; Arabic is another possibility. Many of the new language pairs are being developed in collaboration with third parties; for German-Danish, for example, is now under development at the *Institut for Erhvervsforskning*, in Kolding (Denmark).

Being part of a massive industrial complex – one of Germany's largest – the Linguistic Systems group at Sietec may seem encumbered by its vast parent organization, but Schneider claims that this doesn't necessarily make the group less competitive. "There is a limited amount of R&D development funding available within SNI," says Schneider, "and we have to compete for it along with many other groups." He notes that the survival of the group very much depends on its employees' performance. "In industry," Schneider reminds us, "no future is *ever* guaranteed."

With its huge investment in language processing, Sietec is one of a small handful of companies with both the experience, the technology, and the internal resources to exploit the burgeoning NLP market. Its ultimate survival will largely depend on how quickly it can deploy robust high-end systems and simultaneously downsize and "package" its technology for the much larger small-systems market.

At times in the past, Linguistic Systems group has been faced with the same three-year "make it or break it" criteria imposed upon other groups within Siemens. But when push comes to shove, language processing is perceived as fundamentally "strategic" for the Siemens concern as a whole. This means that for the foreseeable future Sietec will be sticking with it.

Sietec, Carl Wery Straße 22, D-81739 Erlangen, Germany; Tel: + 49 89 636 45191, Fax: + 49 89 636 49646