

Full Tilt into Industrial NLP for SITE

By Andrew Joscelyne

Leading French technical communication firm SITE (pronounced 'seet') is in their own words "getting ready for the year 2000". The company has recently set up a dedicated Natural Language Processing research unit and potential profit center to be known as TILT (Traitement Intelligent du Langage Naturel) under the direction of Bernard Séité in the company's Maison-Alfort premises south-east of Paris. This brings to fruition a plan mooted last year in which SITE would take a major share in B'VITAL, the hitherto cashless Grenoble-based machine translation company, as well as play a key role in providing French high technology industry with some of the infrastructural tools it needs for its massive communication needs—mainly translation.

As a service company, SITE has been positioning itself over the last 5 years to provide the ancillary engineering needed to store, manipulate, access and transmit large-scale technical information for industrial end-users, especially in the aeronautics field. This has involved developing both hardware and software for anything from digitizing huge blueprints to stitching hypermedia front-ends onto text-n-graphics held in databases. The TILT project takes this process one step back by providing, not end-products, but some of the industrial programs and utilities needed to aid others to develop end-user NLP products.

The plan is clearly designed to capitalize on the extensive but disparate results of two decades of French NLP research by transforming them into a genuine industrial platform. This will mean that standards are respected, interoperability is guaranteed and cost-effectiveness and robustness emerge as key criteria. The prime development track, therefore, will be the full testing of the translation software of the B'VITAL Ariane MT rig built out of Grenoble-based university research. This will involve porting the programs from an Assembler to a de facto standard RISC-UNIX platform, thereby reducing operating costs for the translation engine and upgrading the whole system to meet industrial software engineering conditions. One target is to reach 10 Fr (\$2) a page for computing costs, as part of an overall 150 FFr (\$30) a page translation package.

The second, and perhaps more far-reaching objective, is to develop over the next 2 years a "Plug & Play" NLP toolbox, a set of software packages (minus the linguistic data) allowing anyone to develop language-specific applications from a set of 'open' programs. These will partly derive from the clutch of modules already employed by Ariane, and will include an morphology analyzer (theory-independent), a tree-transducer tool (eg, for mapping transformations in syntax processing, or during the transfer stage in translation), finite-state automata of various kinds to run transfer and selection operations, and specialized ambiguity-processing modules such as Prolog-based Q-systems, as used already in the machinery for TAUM, the up-and-running Canadian MT system.

With Plug & Play, end-users will be able to customize indexing systems, develop special-purpose MT applications, use a variety of tools to write a special dictionary, or insert hypertextual links into textual data for intelligent processing of various kinds. In other words, TILT will be attempting to deliver heavy-duty shells for 2nd generation language processing modules, keeping the data separate from the programs, and allowing linguistic data operators to concentrate

on what they are good at.

Since undertaking this kind of research is costly and time-consuming, SITE are negotiating in all directions for interested European (?) partners (don't all rush!) to participate in research on the tool-box venture. This has meant exploring the possibilities offered by the Eureka programs, whose constitutive projects have an industrial and commercial vector, unlike the pre-competitive projects run under the Esprit IT program. Prime partners would be hardware manufacturers such as Texas Instruments, service engineers like Cap Sesa Innovation (who already have a number of NLP projects underway) and large-scale dictionary makers who could benefit from powerful processing tools.

COPYRIGHT © 1991 BY LANGUAGE INDUSTRY MONITOR