

# Opening up Eurotra: Trap or Treasure-house?

*Viewpoint by Tony Whitecomb, Utrecht*

In Europe, a major community effort to boost the language industry's much needed technology is underway. It consists of EUROTRA, and its newly-extended lease on life in the form of the Linguistic Research and Engineering project (see EUROTRA Continues, LIM. No.4). As the EUROTRA follow-up plans from Luxembourg gradually take shape, they are marked by an emphasis on openness, modularity, the use of common tools, and the reuse of resources.

In general, these decisions are to be welcomed. For too long, the rapidly-expanding field of computational linguistics (CL) has suffered from needless duplication of efforts, caused by an alarming lack of coordination among research groups. Reinventing the wheel (but making sure to add that personal touch) has been normal practice in a field where the work style has traditionally been much closer to academic dissertations than to industrial manufacturing processes. Moreover, the quality of method has often been blurred by difference in terminology. If the proposed CEC policy on common tools and interchangeability manages to clear up this chaos, much will be gained.

However, a word of caution is due. In the first place, adhering to a set of common methods (eg, formalism frameworks), software tools (development platforms), and resources (grammars, dictionaries, and termbanks) will only succeed if it is voluntarily supported by the majority of workers in the international CL field. Specifically, this means that the EEC initiative for a "standardized" CL working method will need the full support of not just the various European teams but also our American colleagues. Computational linguistics as an academic discipline has always been very much an American phenomenon, and the sum of the CL activities in US industry (including IBM and AT&T) is certainly no less impressive than in the Old World. As a whole, CL contacts may in be more intensive across the Atlantic than with countries at Europe's perimeter.

EUROTRA circles are, of course, aware of the American connection and its general influence on their field. They rightly support the US-instigated TEI (Text Encoding Initiative). It is essential that American CL centers are actively involved in the broader framework of NLP tools, because if the Americans join, the Japanese are sure to follow.

At the same time, the announced common toolbox approach, admittedly improving the infrastructure and opening up EUROTRA to a wide range of applications, must not veil our continuing inability to make inroads into the semantic barrier. This is the major obstacle keeping us from achieving high-quality MT and other advanced forms of language processing, such as content-based document retrieval, which incorporate at least a modest degree of contextual understanding.

The current situation resembles scattered groups of nuclear fusion or AIDS researchers announcing, after years of trying, the adoption of common work methods and laboratory standards. Using the same tools may enable them to communicate better, exchange partial results, etc. However, the breakthrough may well come from applying a method outside the common tool set the EUROTRA teams are being encouraged to develop.