EUROTRA goes public

by Tony Whitecomb, Netherlands

Review of the EUROTRA Special Issue of "Multilingua" (5-3/1986)

The latest issue of *Multilingua*, Journal of Interlanguage Communication published under the auspices of the Commission of the European Communities (EC), devotes 17 articles (some 27.000 words) to the EC's machine translation (MT) project EUROTRA, "the best funded effort to develop MT in the Western world today".

In his introduction, Harold Somers,from the University of Manchester Institute of Science and Technology (UMIST), calls it "the project's most comprehensive ever presentation to the outside world". It is good to know this. It released me from the feeling that there might have been publications around which I happened to overlook. Equally comforting is Somers's hint that the various internal documents referred to can be obtained by the interested reader. From a couple of years ago I remember that practically all Eurotra documents were considered confidential, "to prevent any unfair competition by private companies", as I was told.

Eager to devour all kinds of unsightful information that I missed, I went through the 45 pages of neatly ordered Eurotra descriptions, and finally felt relieved - because I seemed not to have missed much.

Whatever criticism I have, I must compliment the initiators of this special *Multilingua* issue: it is very useful to have all this information together under one cover, even if the reader already knows much of the facts from the press, EC bulletins, conferences or personal communication. One wonders why the list of research groups involved in Eurotra does not give the names of the members of the Central Team, but this might be due to the volatility of that membership.

After a clear *Management summary* by project leader Sergei Perschke, the general design of Eurotra is discussed in five articles on 13 pages, the authors of which presumably are members of the Central Team. This design overview mainly consists in philosophy on levels of description or representation, with a very meagre amount of concrete examples. It must be unbearable reading for applied linguists, let alone for lexicographers or translators.

The scarcity of translation fragments illustrating how an advanced system as Eurotra is designed to work, exposes a serious shortcoming of some of the Eurotrian authors: the inability to reach and capture a broad audience of (undergraduate) students of language. From a project carried out by universities, one should expect more concern for the educational aspects, the more so because a secondary purpose of Eurotra is claimed to be (in Doug Arnold's article): "to promote the development and spread of expertise in MT and related fields through the Community", and also Perschke underlines "Eurotra's impact on the future of computational linguistics in Europe".

Not only the lack of language material serving as examples of translation problems, but also the minimal number of references (eight, mainly by Frank van Eynde) to general linguistic literature, make the Eurotra design framework not very convincing. On the other hand, the honest way in which the ongoing controversy on the interface structure (IS) is admitted and briefly explained, deserves appreciation. The shown need for structure-changing rules in the *simple* transfer stage appears to become the Achilles' heel of the ambitious Eurotra design.

Even in a *small-size experiment*, as described by Bente Maegaard, one would expect some figures (number of sentences involved, sentence lengths etc.) A clear indication of the size of the *corpus* would have been useful too: some references suggest that it contains all CEC texts published in 1983, but a closer look reveals that it contains only 20.000 words of text per language (the 1983 ESPRIT document). Nevertheless, Bente Maegaard's account of the experiment clarifies how premature and arbitrary the sophisticated formal systel of levels and representations is, as long as it has not been exercised on substantial amounts of language material ("Only language groups can provide sufficient empirical feedback"). One gets the impression that the engineers of the Eurotra framework have gone much too far in their preoccupation with *modularity*: with their extremely abstract *legislation*, they have turned themselves practically into language-ignorant mathematicians, instead of inspiring and guiding their fellow linguists by a comprehensive body of pioneering work in one or two languages.

In this respect, isn't it symptomatic that there is no account of work done on English, in spite of the facts that English is notably the most extensively analysed language and the British group the largest in Eurotra? Careful reading of the lines written by Jeanette Pugh discloses that the group's energy is largely absorbed by participation in the Central Team and establishing communication between Essex and Manchester.

Of the national research teams (the *crucial role* of which is emphasized in the introduction), not only the British, but also the French group failed to present their linguistic groundwork. Only four language-groups managed to report on their activities, and only the German group in a more than cursory way. However, all four reports contain important comments on the general design and status of the Eurotra project:]"... fruitful work by language groups presupposes that linguistic insights remain at least half a step ahead of conceptualizations within the Eurotra framework" (Tsitsopoulos). This observation from the Greek may be equally applicable to the situation of a Portugese or Spanish group.

Haller from Saarbrücken makes the following remark on the feasibility of a common IS (including a basic set of semantic relations and properties): *Voraussetzung dafür sind monolinguale Analysen auf der Grundlage gemeinsam erarbeiteter Kategorien, die nicht nur nach ihrer mathematischen Theorienrelevanz zu beurteilen sind, sondern die auf Vorarbeiten der europäischen Linguistik zurückgehen.* Indeed, the substantial research of the German group shows ties with prominent valency and dependency-grammar work *Engel, Helbig* outside Eurotra, in both the Federal Republic of Germany and the German Democratic Republic.

The report from Copenhagen mentions a "thorough description of Danish noun phrases" but gives no reference to any publication or internal document on this work. The transfer experiments reported by the Danish group confirm that lexical transfer has always been a neglected area *Melby, 1986*: in linguistic theories, in previous MT projects, and now in Eurotra again. Apart from the obvious notion that compositional translation is a suitable safety-net operation for unknown words, no conclusions seem to have been drawn.

The Belgo-Dutch group, mainly involved in *basic research*, echoes a statement that it is better to include complex words in the dictionary than to attempt to decompose them over and over again. Van Eynde's account reveals some of the Eurotra pains such as: many widely differing proposals, preliminary clarification of criteria for evaluation of legislation, etc. This report ends with the hopeful thought that "descriptive and applied research will be given more attention in the near future".

The presentation of the Eurotra project continues with an article on dictionaries. Though no details of actual progress (in terms of lexical entries completed) are given, the intended approach makes a sound impression. Of the four software articles that follow, and on which I have no particular remarks, I found the one by Boitet *Environments for Eurotra* the most interesting.

The final article by Margaret King *Organisational Aspects* is real fun! One only needs to read the first and the last sentence of it: European taxpayers will be glad to hear that (thanks to some 25 million ECU) "a diverse set of people will have invested much effort into learning to talk to one another and to work together".

The feasibility study named Eurotra has not only been very expensive, it seems to have also been very incomplete: little has been heard about various techniques for disambiguation, nothing at all about AI (knowledge-based translation), hardly anything about discourse analysis.

It is for sure that computational linguistics, AI and language technology face a big future in the industrialised countries of the world. In that future, Eurotra will probably be reacted upon as the cultural revolution that threw away the multilingual baby with the model-theoretic wash water.

Reference:
A.K. Melby, *Lexical Transfer: A Missing Element in Linguistics Theories*, Proceedings of COLING '86.