

Computers in the language industry

Andrew Joscelyne reports on the recent conference at Tours on the language industries, and on related developments in computational linguistics.

Ever since Jean-Francois Dégremont, a CNRS research worker seconded to the Ministerial Mission charged with the question, coined the phrase *the language industries* when the French national Computer Assisted Translation programme began in 1983, there has been discussion among linguists and technologists over the relevance of the term. According to Dégremont, the development of language-teaching apparatus such as laboratories, videos and textbooks, dictionary-making, journalism and copy-writing constitute language *crafts* in which the ratio of person to language output remains stable.

Machine-translation, voice-sensitive typewriters and natural language information retrieval systems, however, are language-*industrial* products and processes, since the machines treat the raw material of natural language and generate language acts on an industrial scale without an increase in human expenditure of energy.

The problem with this concept is, of course, that natural language is not a raw material but itself a product of human information processing and has to be decomposed into its constituents only to be reconstructed as a machine version of itself. What is really meant by *language industries* is the industrial production of automata that produce, manipulate and interpret natural languages, thereby placing the specific discipline of computational linguistics at the heart of the notion, as the Japanese fifth generation investments have demonstrated.

Following the report on *The Language Industries: a big cultural, scientific and technological challenge for France* delivered to the French Ministry for Research and Technology last November by Bernard Cassen, ex-director of the Ministerial Mission, a Colloquium on the same theme, camouflaged as *Challenges for Europe*, was held at the University of Tours from February 28 to March 1, 1986, bringing together university research workers, industrial consultants and cultural pundits such as the one-man language industry Anthony Burgess, in an uneasy mix of European *angst* about the threats to languages which cannot 'industrialise' and Gallic pride in presenting French projects aimed at challenging Anglo-Saxon and Japanese domination in the field. Subjects ranging from a Hachette-led agreement amongst European publishers on the operation of Eurolexic, the European dictionary project, to the rôle of knowledge-based systems in machine translation, were aired before an audience which included representatives from CIT-Alcatel, Bull, Texas Instruments, Plessey, Olivetti, Logos (Norway), Auxa (Spain) and IBM. Marcelino Oreja, the Secretary-General of the Council of Europe, co-organiser of the Colloquium with the French Ministry of Culture, proposed the creation of a Group for Coordinating Language Industries in Europe, with the aim of constructing a network of scientific and technical co-operation in computational linguistics to which the Council of Europe would contribute.

He closed the get-together by launching a *Tours Manifesto* in which the virtues of European linguistic

diversity (*European citizens must by definition be multilingual*) were accompanied by the rallying cry of *industrialise your language or it will disappear*. But to many observers, the colloquium functioned inevitably as a forum for the host country's discussions of its own reactions to this threat and how to neutralise it.

The technological need to expand both fundamental research and industrial applications in the face of American private-sector dynamism and the massively-financed Japanese venture has led the French government to accept suggestions made in the Cassen report to implement plans for reorganising language-industrial research. Two major laboratories, the University of Paris VII Laboratory for Computational Information Processing and Linguistics (LADL), directed by Maurice Gross, and Maurice Nivat's Paris VI Laboratory for Theoretical Computer Programming Studies (LITP), will merge to form a new Institute for Computational Linguistics (IIL), to be somewhat paradoxically attached to the National Conservatory for Arts and Trades, where personnel will be increased and work on the automatic analysis of French syntax will continue.

At the same time, the speech analysis laboratories at the Grenoble National Polytechnic Institute and at the Computer Science Laboratory for Mechanics and Engineering Sciences (LIMSI) at Orsay will receive further funds. Some 25 to 60 research posts will be opened, with an annual recruitment of around 25 to 30 posts in future years. It is generally agreed that the recent change in government