The Ovum Report

Major survey – computer processing of language

Has the UK missed its chance?

Journalist and researcher Tim Johnson, who has made a speciality in recent years of the study of expert systems, has produced a 460-page survey of what the computer processing of natural language means in commercial practice. The work is an impressive drawing together of threads, drawing on reports that have appeared in various journals, the writer's own research, and scientific papers presented at any time up to and including the summer of 1985.

It is a work of interest one the one hand to those interested in expert systems and the application of artificial intelligence in computing, and on the other to linguists interested in the techniques or in the implications, which are immense, of processing natural language.

It must be emphasised that what this work covers is not only what has come to be known in the world as machine translation, but also devices which enable humans to communicate with computers using natural language, rather than abstruse programming languages. Indeed it becomes clear as one reads the report that machine translation between different languages (almost invariably English and one "foreign") is a specialist field of its own within natural language processing, and has been pioneered by companies outside the mainstream of developments in expert systems. However both machine translation and other forms of natural language processing (NLP) do depend on the same techniques of linguistic analysis developed in the last 25 years.
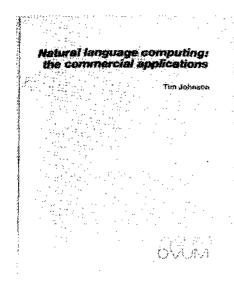
Exponential is a word that author Tim Johnson uses on a number of occasions in the course of the report, and he makes clear his belief that we stand only at the threshhold of the

Tim Johnson

Tim Johnson took a physics degree at Imperial College London, and became a journalist and consultant, working for organisations such as Logica and the *Sunday Times*. He became more and more drawn to an interest in expert systems, and published a previous report on this subject in July 1984. He is the chairman and founder of Ovum Limited, a company which will specialise in consultancy and published research on new developments in artificial intelligence technologies and digital communications.

revolution which natural language processing will bring.

He pays a lot of attention to the potential of what he terms "talkwriters", which are devices using voice recognition techniques, and he is boldly confident of their importance. "Talkwriters", he writes in his opening pages, "will be the main gateways to the natural language networks of the future, and the recent rate of progress suggests they will be available sooner than

anyone dared to predict a few years ago".

The report is geared to the practical, commercial implications of the present state of the art and the path which successful developments are likely to take. Tim Johnson is confident that once people are used to using natural language effectively to access computer resources, they are unlikely to go back to older methods, and he sets out to identify where the opportunities lie.

Early on he sounds a note of realism, and warns just what are the limitations of natural processing – for instance, that none of the systems "come anywhere near the natural language ability of a human child – nor will they for the foreseeable future", and he lists the problems. Later he goes into more detail of how linguists have discovered in the last 20 years just how much deeper than syntax or semantics goes language, drawing on many different levels in the human mind to convey its full meaning. Machines will never be able to talk to humans as humans can, or as one leading NLP theorist is quoted as saying, "In 20 years time you will be able to talk to your toaster – but not about anything very interesting".

But he also indicates how some problems can be tackled. "This report", he writes, "matches confidence in the ultimate importance and capability of NLP with a cautious view of its rate of progress. People looking for quick results from investing their time and money should not venture into this field, and this means it may have problems attracting the finance and staff numbers it needs. But the markets are

there, the means to serve them are in view, and for those who have the stamina, natural language processing will prove a rewarding technology."

Machine translation, though just one of the applications for natural language processing, does already have an established body of literature, at least as far as mainframe systems are concerned. The report therefore concentrates on two more recent trends, the growing availability of MT products on office mini and microcomputers, and a closer relationship with other natural language processing developments.

One of the main points made in the report is the comparative lack of NLP activity in Britain, with what does exist often resulting from the Alvey (British government) or ESPRIT (European Communities) programmes. "Unless something drastic is done soon," Tim Johnson told *Language Monthly* in October, "British companies will find they are shut out of the new markets before they have even started. The Alvey programme has certainly stimulated applications research in natural language processing, but something more is needed to encourage product development by British companies. One idea would be for the government to fund pilot schemes for real application of NLP in its own offices."

The social issues of introducing NLP are considered, from job losses and the hostility of some translators to the *Nineteen eighty-four*-style opportunities (Newspeak *et al.*) it might offer to a totalitarian regime to control the content of information flows or the very lexicography of self-expression within a corporation.

Section B is concerned with markets, and he posits four different scenarios of how natural language processing might develop in the market place. Tim Johnson argues in favour of a gradual spread from its existing base, and gives his reasoning. Potential sales and cash flows to the end of the century are investigated at some length.

Machine translation, he finds, has developed along somewhat different lines to the rest of natural language processing, partly because there was a period when authoritative opinion thought MT was a waste of time, and it took maverick companies or organisations like the Mormon Church (whose work gave rise to ALPS and Weidner) to choose to ignore such conventional wisdom.

"European entrepreneurs with such a weight of opinion against them", he adds in side-swipe at conservatism on this side of the Atlantic, "would normally find it impossible to raise funds".

Part C is concerned with the technology and concepts, and must rank as one of the most lucid explanations yet penned (for those of us who are not specialists in computational linguistics) of both what is involved in voice recognition, and the theoretical principles behind the processing of natural language. Here the reader will find a guide to the differences between context-free and context-sensitive grammars, definitions of transformational, phase-structure and case grammars, and what happens when a computer is faced with a badly-written original which fails to conform to the grammatical rules so carefully elucidated.

Applications are dealt with in part D, and many readers will be surprised to learn just how much is going on. The report, however, is almost entirely based on experience in the United States and the United Kingdom, with only passing references to what is happening in Japan, and virtually no reference to what is happening in Europe, such as the French telecommunications authority's ambitious plans for processing natural language inquiries.

It is suggested that the market for machine translation software could be worth about £100 million a year by the early 1990s in the United States and the United Kingdom, with a potentially much bigger market on the continent of Europe. Reports are given of claimed productivity gains – such claims will be familiar to readers of *Language Monthly* (some of the reports are indeed taken from *Language Monthly*) – but Tim Johnson goes on to evaluate these claims and compare them to those for manual translation. He also looks at other advantages of MT (convenience, consistency of terminology, integration with document flow).

Part E consists of profiles of companies involved in natural language processing, and section F consists of a glossary and references.

The importance of this report for anyone concerned with the future of cannot be underestimated. It also provides a clear warning that Britain may find itself excluded from one of the most important fields of development.