Soviet Union holds major MT conference

On December 5 last year hundreds of delegates, representing seventeen countries all told, gathered in Moscow in order to take part in a week-long seminar on machine translation (MT).

Many of those delegates must have recalled their attendance at the previous two seminars held there in this four-yearly series of intensive "updating" conferences. The handbook - the best part of 300 pages long - of seminar pre-prints is an eloquent testimony to the results already achieved and those in almost certain prospect. As always, there are indications of hopes, and descriptions of salutary experiences too!

statistics

The statistics are impressive enough: some 152 papers were delivered, 134 of them in Russian (including contributions from Bulgaria, Cuba, Czechoslovakia, Hungary and Poland), 14 in English, two in German and two in French. The papers in English came from Canada, Denmark, both "Germanics", Great Britain, India, Japan, Poland and the United States. MT experts from the German Democratic Republic made two presentations in German, and French was used by speakers from Canada and France.

Whether or not the above statistics induce feelings of admiration or consternation, there can be nothing but sympathy for, and empathy with, the words of Professor Marchuk, chairman of the seminar's organising committee. These words, - quoted in full but translated into English, and clarified at one or two points, - cohere to become the Moscow seminar's "watchword":

"A basic feature of this Seminar, and one which distinguishes it from the previous Moscow International Seminars, and which characterises recent developmental tendencies in MT, is the predominance of practical views of the MT problem. MT has become a part of general information systems and services, with all its inadequacies and unsolved problems.

"Many papers are devoted to the results being accomplished by operational MT systems, such as technical and operational cost-effectiveness, the refinement of algorithms and dictionaries, the improvement of system performance, new hardware facilities, the place of MT in information science and information technology (IT). Considerable experience has been accumulated and digested with respect to 'customer satisfaction' with various types of MT output, such as MT for 'current awareness' documentation, MT with pre-editing or with post-editing coupled with the use of special, restricted input languages.

"Be all that as it may, the predominance of the customer's viewpoint, or that of the operational MT system's manager, does not mean in the least that interest has waned in cardinal problems of a theoretical nature, problems being tackled by bold attempts to communicate to the computer the essence of the human translator's complex and often creative work. These problems are the subject of a good many interesting papers."

At a major conference the topics of the plenary papers are usually a sure guide to the focus and thrust of the conference. On this occasion the plenary papers fell into two more or less distinct categories: those delivered by "guests" and mostly offering "state-of-the-art" accounts with reference to particular countries and those delivered by the "home team" and concentrating on particular themes of central importance to MT.

Marchuk gave an overview of MT in the USSR and will, incidentally, be speaking on the same subject at the forthcoming MT conference at Cranfield. Exactly the same service was rendered, by Gobeil, with respect to MT in Canada. Somers - based at the Centre for Computational Linguistics at UMIST - gave an account of their impressive Bede MT system, conceived for and partially implemented on an 8-bit Superbrain. Uchida relayed to the Moscow Seminar the principal design features of the ATLAS semantics-based MT system being elaborated in Japan.

In terms of "theme" papers in the plenary session, those by Kotov, Piotrovskii, Rozhdest-venskii and Nelyubin are well worth dwelling on.

Kotov, in bravely declaring his pragmatism in dealing with "real-life" texts as input to MT systems, called for a more sustained effort by applied linguists in their analysis of such texts, notably of business and commercial documentation, with a view to elucidating - and maybe even influencing in the longer term! - the theory and practice of terminology and nomenclature.

Piotrovskii drew attention to the horrendous problem of mapping linkages extending beyond the merely denotative components of meaning.

minefield

Rozhdestvensii reported on the lexicographical "minefield" which has to be charted by MT dictionary designers: one area where - to maintain the analogy - the mines have not been laid in symmetrical patterns is, paradoxically, that of general scientific lexis. Mapping English of German into Russian is fraught with difficulties. The analysis of translations has revealed problems which appear to be successfully defying, at the present time, the best endeavours of MT lexicographers to design a sufficiently sophisticated lexical data-base as opposed to a lexical data-bank.

Nelyubin, in a subtle paper, pleads for the use in MT of techniques he has researched into and published on in the world of military communications systems. His flow-chart reads "optimisation - formalisation coding - translation", and involves a whole gamut of techniques for equivalence matching; more important, however, is the suggestion that research, in "text linguistics mode", is now urgently needed to identify and "fingerprint" large numbers of sublanguages, each with its distinctive lexis and grammar. This view that language varieties are not on a cline but have their own isoglosses and "isogrammars" is necessarily somewhat artificial but offers much of potential value to MT researchers.

The extensive comments above on the Moscow Seminar's plenary papers could give the impression that the sectional papers were less important or interesting. Such an impression would be false.

self-adaptive

Balaban's paper on self-adaptive dictionaries is extremely thought-provoking and Skorokhodko's account of the strategies needed for text analysis without a complete dictionary is encouraging. A lot more can be done than just lodge the offending item in a "not-found" file. Algorithmic methods can often close the gap to a remarkable extent.

Vannikov contemplates "loose" and "tight" bonding in text in an effort to isolate translation units successfully. Yet other pragmatic aspects of MT also receive their share of attention: Ryabtseva presents an interesting functional analysis of post-editing, and Malevich reminds us of the software engineering which must have been completed before any text can be processed at all. This often requires the definition and implementation of special programming languages. Malevich's "Osa" ("Wasp") is in this category, having been designed for the purposes of semanto-syntatic analysis.

All in all, the papers delivered at the 1983 MT Seminar in Moscow provide much information of value and interest to MT specialists elsewhere.

It is to be hoped that the Seminar's Proceedings will be published in due course. The author of these words believes - as a result of his personal contacts with many of those involved - that Soviet MT research, even after the loss of some of its "stars" by emigration, has a lot to offer in terms of philosophy, strategy, tactics and pragmatics, not to mention software or computerised dictionaries. Almost intuitive assent has to be given to the view of MT as a perfectly "normal" module - at most, primus inter pares - in multi-purpose information processing systems capable of performing a whole range of activities, from automatic abstracting to fact retrieval, either monolingually or bilingually.

F E Knowles