

Whither MT?

**A report on the
TMI meeting in
Kyoto (July 14-
16, 1993) by
Colin Brace,
Editor and
Producer of the
Language Indus-
try Monitor**

The bi-annual Theories, Methodologies, and Issues in Machine Translation (TMI) conference was held in Kyoto (Japan) this year, the week preceding the MT Summit in Kobe. With the Summit this year consisting solely of invited talks, the TMI provided a complementary setting for the presentation of refereed research papers. As the premier forum for researchers involved in MT, the TMI offers a valuable glimpse of the "hot" research topics at the moment and (ideally) gives a taste of things to come. This year's conference was revealing both for what it did and did not reveal.

The MT community has been shaken up in recent years by the arrival of statistics-oriented researchers who have boisterously dismissed linguistics as a basis for MT in favour of massive word frequency tables derived from bilingual corpora. To be sure, statistics-based MT is decidedly experimental (and wholly dependent on the availability of large, high-quality bilingual corpora), and correspondingly it is still a *long* way from equalling, let alone surpassing, the quality of Systran.

While the divisiveness of their presence has been overstated, this small group of researchers have nonetheless set in sharp relief the fact that the linguistic approach espoused for the past 40 years — and at the heart of today's working systems (notably Systran and Logos) — has been more or less exhausted. Syntactic and semantic analysis as we know it has been exploited about as far as it can go, and this leaves us with MT systems that require a lot of constraints to work, either in terms of input quality, output quality, domain, or human interaction.

The problem is that there is a lot more to language than just linguistics, namely those aspects of human communication which are usually clumped together under the term "pragmatics". This includes textual phenomena which bridge multiple sentences, such as pronoun references, and a whole realm of matters concerning the context of a spoken or written utterance. Psychologists, (socio-)linguists, philosophers, and cognitive scientists are all grappling with these issues, albeit from different angles. Obviously, we will never be able to exploit such information in computers unless we understand it ourselves.

In tacit acknowledgement of this seemingly insurmountable challenge, this year's TMI (like the previous one) was noteworthy for the substantial interest in example-based MT. Very simply, example-based MT sidesteps the entire issue of linguistic and pragmatic models of language by producing translations derived from a database of bilingual translation examples. To be able to translate texts not in the database, these systems use a thesaurus-like

function to try get guess the “nearness” of the source text to the examples it has in its bilingual database, thereby allowing it to make substitutions.

Japanese researchers, who, as a whole, have eschewed the theoretical debates of their American and European colleagues, have embraced example-based MT for its obvious practical advantages. In theory, the more examples you have the better your system is. In addition, example-based MT is eminently suited for multi-processing systems, because retrieval chores can be easily divided among several computers, something more difficult with linguistics-based systems. Hence, what your example-based system might lack in finesse, it can compensate for with brute processing strength. While largely the domain of the Japanese, this approach has also been tried by others, notably the DLT team at BSO, which referred to it as “analogy-based translation”.

Mind you, no commercial MT system incorpo-

rates example-based MT techniques; it is still very much an experimental approach. But commercial developers in Japan and elsewhere are interested in it because they see it as a way of potentially improving the “naturalness” of the output. You are therefore not likely to see a system built solely on the basis of example-based MT techniques but rather you may find it used as one technique in an array of many. That’s not surprising. All of today’s working MT systems are by definition hybrids; you can’t build an MT system without being pragmatic, and that implies using whatever works.

However, example-based MT won’t solve all of the problems that MT researchers and developers are faced with. While we may have come as far as we can with “traditional” techniques, TMI ’93 did not suggest any comprehensive alternatives. To venture a forecast: there are no major breakthroughs on the horizon.