

[*Language Industry Monitor* 3, May-June 1991]

MT Evaluator's Forum

From our correspondent

Arrange a conference in a splendid city like Berlin and you can be sure that half the participants will cut half the sessions. A move to Scunthorpe certainly slashes absenteeism: few feel like pacing the town and even fewer bother to turn up in the first place. However, as the first ever international MT Evaluators Forum (Les Rasses, St Croix, Switzerland; April 20–24, 1992), demonstrated, the perfect compromise seems to be a smart hotel stuck on top of a Swiss mountain. Everyone came (Europe, US, Japan, China) and nobody missed a minute; the late spring snows and complete absence of night life — unless you're into squirrels — effectively nipped the extra-curricular activists in the bud.

Among the fifty or so participants, translators with hands-on experience evaluating systems were out in force. Others present included sundry academics and one developer. Guiding spirit behind the Forum was the intrepid Margaret King of the ISSC, Geneva.

All agree, MT is back with a vengeance (METAL, LOGOS, and, er... SYSTRAN) and a lot of people are thinking hard about buying it. Unhappily, both existing and potential customers are having to face up to the fact that MT systems don't actually translate texts — they merely help to translate texts. Moreover, without any kind of standardized quality metric, it is wretchedly difficult and expensive to find out just how much help — if any — they will provide.

MT manufacturers talk blithely of a "100% increase in throughput", but since these are the people whose toys turn the phrase "...les agriculteurs vis-a-vis le EC..." into "...farmers live to screw the EC...", scepticism is rampant. If you're running the hard-pressed language department of a Swedish electronics conglomerate, you want to know whether a 150,000 dollar MT investment is going to pay off sometime before the Second Coming. Some of the better MT systems can prove cost-effective in some circumstances — one translator said she really had doubled her output with METAL and, no, she hadn't been paid to say that — but this is far from guaranteed.

The bottom line is it costs a bundle to evaluate MT systems and their output. Various reported evaluations had clearly cost many tens of thousands of dollars, only to provide equivocal results. The EUROTRA UK team at Essex University say they were asked by a British publication to put German–English and French–Spanish versions of Globalink's PC-based rigs through the paces. After hiring native speakers and other expenses, they said their evaluation cost many more times than the modest few thousand dollars the packages cost. Likewise, the Union Bank of Switzerland had a team of two testing MT systems for nine months, a costly undertaking calculating salary alone.

Developer's have evaluation headaches, too. Elliott Macklovitch (Canada) reported on a series of trials done with successive "new versions" of a well-known MT system; overall translation performance actually got worse. That particular company, in concert with researchers and developers elsewhere, is now putting considerable effort into developing "test suites" — batteries of thousands of special sentences, each of which contains one potentially troublesome linguistic construction. Each time it sputters, you lift up the system's bonnet and take a spanner to its grammars.

If Forum attendees could agreed on one thing, it was this: the folks buying MT systems should not have spend a fortune evaluating them. The second big meet is tentatively scheduled for 1993.