San Diego's Other MT Company

From its humble origins, MicroTac's Language Assistant series has the potential of becoming one of the best PC-based translation packages. How far will the company be able to go?

Mention MT and San Diego in the same breath to most people and they will usually think of Systran, located in the adjacent town of La Jolla. But San Diego boasts a second translation software company, MicroTac. With some US$3 million in sales last year, MicroTac is about the same size as its venerable mainframe-based counterpart, although it cannot be said that they are in any way competitors – at least not yet. "Systran is about three miles up the road," says Michael Tacelosky, founder of Micro Tac, "but we've never shared technology or people." Tacelosky attended a Systran open-house last year and "laid hands" on the IBM 370 running there, but that is as far as the interaction has gone.

Since introducing the first Language Assistant in 1988, MicroTac has done very well, with total sales of Language Assistant packages of over 150,000 making it far and away the best-selling translation program of all time. With the four Language Assistant packages (French, Spanish, Italian, and German), MicroTac has penetrated the mainstream of mass-market software. That means you'll find the MicroTac software on the shelves of nationwide chains like Egghead Software and in the glossy advertisements and catalogues of the big software mail-order companies. With a list price of US$99.95, it is inexpensive and well within the reach of the casual – or curious – buyer. Language Assistant began life as a simple verb conjugator for French and Spanish aimed at students and other people learning languages. With each subsequent release, MicroTac has enhanced the software, first with online dictionaries and an editor, then with an automatic lookup and substitution routine. Version four offered a rough batch translation from English. Version five, which was released in late 1992, offers bidirectional translation and represented a signficant advance over the previous versions, with an expert system engine based on "thousands" of transformational grammar rules selectively firing to build a sentence tree. While version four was coded in the underlying programming language, version five has an external language that can be coded by linguists, not just programmers. As a result, the Language Assistant can divine shallow syntactic structures from source texts and generate intelligible if not flawless output in both directions.

In its current form, Language Assistant offers a pleasant environment for writing in a second language and generating rough translations of existing texts. The package sports a WordStar-style text editor and a series of reference tools, which can also be loaded separately from the main program and used from within other packages. The reference tool program provides access to the bilingual dictionaries at the heart of the system, a conjugator, and an extensive reference on grammar. The package also includes a utility for entering accents which is the most elegant system we have seen for entering accented characters on a DOS computer.

The Language Assistant also offers automatic and interactive translation modes for translating texts. To get just the bare essentials of a text, you can whizz through it in automatic mode and let Language Assistant use the default selections of its lexicons. Or, you can step through it sentence by sentence in interactive mode, whereby the program

prompts you to select a target word where there are multiple options. It would be useful to be able to enter a word on the keyboard that isn't offered as an option in this mode but that probably would not be technically feasible.

You can format the output in a variety of different ways, such as source and target sentences arranged side,by, side or line-by-line. With the Word Scan function, you can scan a text for unfamiliar words which can then be added to the Language Assistant dictionaries. The dictionary update screen allows you to enter a variety of informa, tion about a new entry, such as morphology and gender, but also semantic information, such as animate, place, or proper, in the case of nouns, and reflexive, separable, and takes do in the case of verbs. The Language Assistant dictionaries are surprising in the "depth" of their coding, but it is not clear exactly how much of this information its parsers use. Like the text proofing package RightWriter, Language Assistant also lets you take a look under its hood. In interactive mode, you can view the parse tree that the program has generated for the current sentence, with a fair amount of information about each mode. Unfortunately, the parse tree shown is a synthesis of the source and target sentences, making it somewhat less useful for didactic purposes. The possibilities and shortcomings of Language Assistant being highly instructive, it and a copy of Hutchins and Somers *Introduction to Machine Translation* would probably be about the best introduction to MT you could find.

While the language learning and educational markets are the obvious target for the Language Assistant series, Tacelosky also sees the packages being helpful for English, speakers who don't command a given foreign language yet require "informal" translations of materials. A typical case might be an office worker who receives correspondence in Spanish and just needs to get an idea about what a letter says. The caveat here is that this person would have to be prepared to type or scan the letter in, preferably with accents intact.

In the arena of informal translation, Tacelosky says that he believes that MT and email will be the "perfect match." Earlier this spring, he began circulating an experimental version of Language Assistant that could be hooked into the CompuServe offline reader OzCIS and used for scanning the contents of forum messages. CompuServe has a growing presence in Europe and there is a huge potential market for Europeans "informally" scanning CompuServe's exhaustive forums and online resources. CompuServe is itself aware of this potential; it hired a linguist last year to evaluate the possibility of offering online translation of its materials. Moreover, it is reputedly gearing up for eight-bit operation, which will ease, among other things, ad hoc scanning with a translation package like Language Assistant which works best with properly accented materials.

Email and conference forums are an obvious application because all the material is sitting there in electronic form; it doesn't have to be typed or scanned in. On the other hand, this kind of informal speech can be notoriously difficult to translate automatically. Email users send messages which have a lot of the attributes of spoken speech: misplaced modifiers, ellipses, agreement errors, not to mention that translator's bugbear – humor. Like most translation systems, the Language Assistant is best suited to simple, declarative sentences. Whether the Language Assistants would appeal to European users for scanning CompuServe may depend on whether the packages are as good at generating foreign languages as they are at generating English. A casual look at the German Language Assistant revealed that it did considerably better going from German into English than vice versa, but the Romance language versions may be better because MicroTac has been working on them longer. Tacelosky says the team found German a particular challenge. Compounds, separable verbs, and extensive inflections make German translation more difficult than the Romance languages, which more often follow English word order.

As Tacelosky had discovered, if there is one group that is decidedly not the target for Language Assistant, it is translators. Tacelosky attended last fall's MT Evaluation Workshop and ATA conference in San Diego but chose not to exhibit at the latter, saying – rightly so – the translators were not his market. Tacelosky is nonetheless a frequent visitor to the Foreign Language Forum of CompuServe (FLEFO), a popular haunt for the on, line denizens of the translation

world, in particular various members of the ATA. T acelosky uses this FLEFO as an informal channel for providing technical support for the MicroTac software. While there are occasionally interest, ing threads of discussion there on MT, Tacelosky and others also regularly need to rise to defend MT from attacks by a vociferous anti,MT contingent there.

Much of the conflict surrounding MT revolves around the question; will MT replace translators? "This is the wrong question," says Tacelosky. "The real question is: will MT replace non-translation?" T acelosky makes the useful distinction of understandable-quality machine translation (UQMT) and publication quality machine translation (PQMT). "Professional translators produce translation suitable for publication," he explains, "while the average person just wants to understand what's going on in a text. MT is great for producing something understandable from something unintelligible."

While the relationship between professional translators and MT developers has been rocky in the past, Tacelosky sees smoother sailing in the future. "MT is evolving into an 'expert,system' environment where the translators are the 'experts,'" he maintains. "As has happened in other fields, professionals will not be replaced by this technology but enhanced by it." A useful analogy might be found in the engineering world. In the early 1980s, experienced drafts, men and technical designers fought tooth and nail against the introduction of computers in the technical drawing offices because they found the new technology threatening. Now you would be hard put to find a single one of these professionals not using AutoCAD or a similar package.

1993 has been a bumper year for coverage of MT in the computer press. *Byte*, *WP Magazine*, *Publish*, and *PC AI* all devoted space to stories on or reviews of translation software. It seems that the press is learning how to approach the subject. Language Assistant fared well in the hands of Barry Brenesal, who reviewed the package for *PC Magazine* (circulation: one million). That magazine's editors astutely avoided the issue of whether the package was "business" software – and the implications thereof for evaluating it – by placing the review in "After Hours," the section of the magazine which is dedicated to "recreational" software. Here it was compared with other language learning packages. "I was delighted with the article," says Tacelosky. "It described what the product does, not what it doesn't do, and did not over-hype its abilities."

In a more specialized forum, the French Language Assistant was also discussed in this year's annual British Computer Society's NLTSG newsletter. It was given an informal blackbox evaluation by the newsletter's editor, J.D. Wigg. While offering an even-handed evaluation of its workings, Wigg did not go on to speculate about what the package might be useful for. But regardless of the specific nature of this year's extensive editorial coverage, MicroTac and its competitors all stand to benefit from this burst of exposure.

With its two different personalities – language learning and text translation – version five of the Language Assistant series looks like a product in transition. Judging by the depth of dictionary encoding, MicroTac is setting its sights high. If MicroTac linguists are laying the right foundation, version seven or eight of Language Assistant might be an extremely impressive translation system. A lot will depend on the company's mobilizing considerable resources to do this. More important, it will depend on Tacelosky's technical and linguistic acumen and his ability to attract top-flight, native-speaking linguists for the foreign languages he wants to support. In the meantime, there seems to be plenty of takers for Language Assistant, and that is a vital first step.