Two German Books
About Machine Translation

Reviewed by Alex Gross

Machine Translation: Theory, Applications, and Evaluation, An Assessment of the state-of-the-art, edited by Nico Weber

Evaluation of the Linguistic Performance of Machine Translation Systems, Proceedings of the Convens '98 in Bonn, edited by Rita Nübel and Uta Seewald-Heeg

(comprising Volumes 1 and 2 of the Series "Sprach-wissenschaft, Computerlinguistik, Neue Medien," Series Editor: Nico Weber, DM 49.90 per volume, published by Gardez! Verlag, Meisenweg 2, 53757, St. Augustin, Germany; TEL: 0 22 41/34 37 10; FAX: 0 22 41/34 37 11; E-mail: gardez@pobox.com)

These slick, green paperbacks could not be more business-like in their appearance. They are clearly serious books intended to deal with serious issues. And their twenty assembled authors carry out this intent in an uncompromising fashion without a hint of the history behind their subject. And herein perhaps lies the chief fault in these competent but circumscribed volumes.

For almost fifty years, the promise—even the certainty—of machine translation taking over from humans was a recurrent part of the grand computer dream, merely one component of an all-enveloping "artificial intelligence" destined to organize our menial tasks, our language problems, and even our daily driving. But during the past ten years—or perhaps only the last five—this dream has slowly receded, as even MT and AI experts have come to grasp the true scope of the problems they had undertaken.

These two books can barely reflect this overwhelming reality—perhaps the closest they come to mentioning it is the very first sentence of Volume I:

"Machine Translation (MT) has, somewhat unexpectedly, made a come-back during the 1990s."

What Series Editor Nico Weber most probably means here is that during this period MT developers finally gave up on trying to persuade translators to adopt their systems and seized the Internet and other publicity outlets to bypass translators and make an end run in favor of the uninformed general public. Defeated in their original aims, they decided to proclaim total victory instead and rope in as many ordinary citizens as possible as hobbyist users. Which is not to say that MT cannot be integrated into small subsets of language, such as specific knowledge domains, parts catalogs, or predetermined questions and answers—it may in fact work best here, though it is a far cry from the vast scope originally claimed for this field.

Certainly translators have not been averse to working with computers during this period—they have in fact been among the most avid users, scouring the Web for all manner of glossaries, editing tools, and translation aids. But to the extent that translators and translation companies have truly switched over to computer techniques, they have tended to abandon "Machine Translation" in favor of "Translation Memory," an approach that bears about as much resemblance to MT as does a lexicon to a log table.

So essentially what we have in these two books is the account of a solemn retreat from MT's bygone days of would-be glory. The main topic of both volumes is something called "MT Evaluation," essentially a euphemism for trying to discover and explain why these systems have on the whole performed so poorly. The entire second volume is devoted to this topic, with two of the first volume's six papers sharing the same theme (and another two aimed in much the same direction).

This leaves only two papers dealing with other topics: one by Isabelle Schrade on "cognitive" aspects of translation, and another by Jürgen Rolshoven about using object-oriented programming to improve MT systems. The first is almost a parody of Chomskian acolyte Steven Pinker's "Cognitive Neuroscience," encouraging an author to string profound bromides together almost endlessly, as is done here.

Translation, Dr. Schrade tells us, embraces seven essential qualities (and she devotes a few pages to each of them): Memory, General Knowledge, Linguistic Knowledge, Understanding and Analyzing, Recipient-Oriented Reformulation, Human Intuition, and Creativity. As for Prof. Dr. Rolshoven, he treats us to little more than a tantalizing—though familiar—exercise in Chomskian diagram-juggling.

None of these criticisms is intended to deny the high seriousness of the task being undertaken nor of the authors' sense of loyalty to their aims. The reader watches in awe as they painstakingly explain their quest for a valid methodology, one that will provide the surest and most scientific means of testing and comparing first six and later four different off-the-shelf MT systems.

But in what is already an enormous compromise, they decide that their tests "should be based on a number of grammatical phenomena which are prominent for text types which in turn are commonly considered *typical* MT text types" *(editors' italics)*. If only they could succeed in their quest, perhaps it might lead to a small but significant improvement in MT quality. After much discussion, seven types of phenomena are proposed for testing, but only three are finally selected, providing perhaps some notion of the authors' style and rigor:

"Request forms," the editors' term for typical imperative verb forms found in computer and automotive documentation;

"Compounds," comprising a vast array of noun-verb, verb-noun, adjective-noun, and noun-noun composite words;

"Coordination," their term for converting English ellipticisms into more structured German forms.

The four categories rejected by the evaluators "because of time constraints" were "participial constructions, adjuncts, nominalizations, and idiomatic expressions."

But how valid are their testing procedures, and how likely are their findings to reach their goal? As the editors of the second volume confess in their final summary, "testing the linguistic coverage of an MT system is a tedious, time-consuming task." And a note of unintended comic relief is provided by the one MT developer invited to take part, when he points out first of all that:

Methods for evaluating machine translations and machine translation systems have been proposed, discussed, and applied for more than 40 years now, including numerous attempts at defining objectively measurable criteria to capture aspects of translation quality.

Nevertheless, a worryingly large number of evaluation reports have more or less explicit disclaimers as to the absolute value of the results, or confess to flaws in the procedure.

and then draws the precise conclusion one might expect from an MT developer:

The obvious solution to these problems is of course to avoid translation quality as a direct object for evaluations and to stay with a general impression of the role which quality plays for the overall acceptability of a MT system.

And there are other moments of unintended comic relief. For instance, the abstract for one paper tells us that the reason for these labor-intensive researches is because "these systems require small-scale evaluation methods which can be carried out without the developers' cooperation." And we learn that the advent of the latest and cheapest systems has spurred even the mighty Association for Machine Translation in the Americas to discuss a so-called "MT Seal of Approval" at their 1998 conference.

And amid all the precious examples of MT output, a few more fully certified gems emerge:

"It is a pity that I can't speak French." becomes in German

Es ist französisch ein Mitleid, das ich nicht kann sprechen.

While "The dog that had eaten the hamburger ran away." is truly turned into hamburger:

Der Hamburger lief der Hund, der gefressen hatte, davon. (which in English might become "The man from Hamburg ran the dog...")

It is a relief to report after all these testing procedures, graphs, tables, and countless examples, that the editors do finally reach a conclusion about the six principal systems that have been evaluated. Based on their experiments, they determine that "Logos, Personal Translator Plus 98 *(Linguatec/IBM),* and in many cases Systran belong to the top three. T1 Professional *(Langenscheidt)* is in the middle field, sometimes also Systran, and Transcend *(Intergraph)* and Power Translator Pro *(Globalink)* always come last."

The first volume is almost entirely in English, while the second volume weaves quite seamlessly between German and English. In so doing the editors inadvertently show something of their own basic linguistic orientation by inventing two new English abbreviations (or at least new to this reviewer) on the basis of familiar German ones. Thus, in Volume 1 we find "resp.," no doubt a German stab at "respectively," presumably on the basis of German "bzw.," *(beziehungsweise)*, while Volume 2 yields "a.o.," evidently an attempt to duplicate the German "u.a.," *(unter anderem)* for "among others." Both of these are certainly good tries and perhaps ought to exist in English, but they do raise certain doubts as to the overall English capabilities of the authors, especially when they confess that "advanced students of English (all native German speakers)" performed all the English post-editing in one task supposedly evaluating how long this should take.

This linguistic orientation is perhaps also revealed in the paper I find most interesting, the first volume's final offering: *The Automatic Translation of Idioms: Machine Translation vs. Translation Memory Systems* by Martin Volk. This piece comes down firmly on the side of Translation Memory as being superior to MT for translating idioms. But I question its basic dichotomy, that there is a clear and discernible difference between what we call "idioms" on the one hand and the "more predictable" parts of language on the other. I am not altogether sure that this dichotomy will stand up to any truly close analysis, particularly if we begin to consider more exotic languages, which even MT developers claim they will one day be able to include by using an "Interlingual" approach.

It might be supposed that this is merely a linguistic quibble, and that surely what appear to be simple sentences of the type "You are beautiful" must be much the same the world around. But I can easily conceive of languages and cultures—and I believe many of our readers can as well—where the words "You," "are," and even "beautiful" might be up for grabs and pose unexpected problems even for human translators—and certainly for machine translation systems as well. It could yet turn out that all—or almost all—of language is unpredictably and close to arbitrarily idiomatic in nature. And that only the coincidence of two languages, such as English and German or English and French, growing closely together over several centuries, has persuaded us that this may not be the case.