

EDITOR

Dr. Frank R. Smith  
Technical Communication Services  
910 Milldale Drive  
St. Louis County, MO 63011  
(314) 391-1002

ASSOCIATE EDITORS

**Writing / Editing:** Dr. Kristin Woolever, Northeastern University, Boston, MA 02115

**Education:** Dr. Victoria Mikelonis, University of Minnesota, St. Paul, MN 55108

**Audiovisual:** Saul Carliner, IBM Corporation, Atlanta, GA 30055

**Research:** Thomas E. Pinelli, NASA Langley Research Center, Hampton, VA 23665

**Recent and Relevant:** Dr. William E. Stephenson, East Carolina Univ., Greenville, NC 27858-4353

**Book Reviews:** Dr. Susan Dressel, Los Alamos Laboratories, Los Alamos, NM 87545

**Computers and Communication:** Liora Alschuler, Miles-Samuelson, Inc., New York, NY 10003

**Communication Management:** Patricia Caernarven-Smith, Consultant, Deerfield Beach, FL 33442

**Slaying the English Jargon:** Fern H. Rook, Consultant, Phoenix, AZ 85018

**A Teaching Tip:** Dr. Meg Morgan, University of North Carolina, Charlotte, NC 28223

**Current Research in Technical Communication:** Dr. Stephen Doheny-Farina, Clarkson University, Box 5760, Potsdam, NY 13699-5760

**Technology Reviews:** Dr. C. J. Wallia, California Publishing Institute, Box 5582, Berkeley, CA 94705

**The Wired Word:** William Horton, 1523 Ward Ave. NE, Huntsville, AL 35801

**International Technical Communication:** Fred Klein, Hollywood, CA 90046

**The Business of Technical Communication:** Dart G. Peterson, Blue Pencil Associates, Dayton, OH 45416

**Reminiscences: A Personal History of STC,** Dr. F. R. Smith

COPY EDITORS

Betsy Frick and Barbara Melton  
American Express Information Services, St. Louis, MO 63042, Sharon Boranyak, St. Louis, MO 63129

ADVERTISING AND SUBSCRIPTION REPRESENTATIVES

Williams & Wilkins Co. 428 E. Preston St.  
Baltimore, Md. 21202

ADVERTISING ..... Charles Hirsch  
(301) 528-4283

SUBSCRIPTIONS..... Maria Reid  
(301) 528-4099

CHANGES OF ADDRESS AND CORRESPONDENCE. Members of STC should send notification of change of address to the STC office. Any change of address or correspondence relating to nonmember subscriptions should be sent to the subscription representative. Subscription is included in STC membership; it is \$40 per year for nonmembers (\$60 overseas). Individual issues may be purchased for \$10 each (\$15 overseas) from the Society office while supplies last.

*Technical Communication* is published quarterly by the Society for Technical Communication, a nonprofit educational organization. All rights reserved, but reproduction rights are granted upon written request. Copyright © 1990 by the Society for Technical Communication. Second-class postage paid at Washington, DC, and at additional mailing offices.

POSTMASTER: Send address changes to *Technical Communication*, 815 Fifteenth Street N.W., Washington, DC 20005.



STC Office  
☎ (703) 522-4114

901 North Stuart St.

Arlington, VA 22203

ARTICLES

- 112 **Desktop Publishing and Resource Management**  
Richard B. Lowe  
*Examines relevant questions about time, money, work integration, and people and concludes that although desktop publishing demands a new approach to resource management, DTF will soon be indispensable.*
- 116 **Annotated Bibliography on Costs, Productivity, Quality, and Profitability in Technical Publishing, 1956-1988**  
Patricia Caernarven-Smith  
*Reviews the articles and papers of the past three decades that deal with cost estimating, cost analysis, productivity and performance improvement, and quality control, concluding with a plea for some controlled research on costs.*
- 122 **The Encyclopedic Reference as a Model for Print Documentation**  
David K. Farkas  
*Assesses the strengths and limitations of the traditional user's guide, describes the encyclopedic reference, a single alphabetically organized manual that replaces the standard documentation set, and offers some thoughts on the future of computer documentation.*
- 126 **Improving the Readability of Specialized Language in Emerging Technologies**  
Thomas H. Miles  
*Gives a case history of how one writing group devised a way to deal with the problem of author-created noun strings and long, indecipherable unit modifiers, satisfying both internal and external clients.*
- 130 **Developing a Simplified English Vocabulary**  
D. A. T. Peterson  
*Establishes the concept of a Simplified English, mentions several varieties that have been developed, and then outlines a procedure by which any technical community can develop its own version for its specific "domain of discourse."*
- 136 **The Elephant and the Blind Men: A Look at Functionality**  
Al R. Young  
*Shows how each department in a company has a different view of functionality, which is never that of the customer, and advocates a unified view founded in the user's perception of functionality.*
- 143 **Establishing Quality Standards and Trade Regulations for Technical Writing in World Trade**  
L. C. Weymouth  
*Calls for a concerted effort by publication professionals and government to establish quality standards and trade regulations for technical documents and discusses various quality-assurance mechanisms.*
- 148 **The Technical Editor and the Non-native Speaker of English**  
Ilona Leki  
*Summarizes current views of second-language acquisition, focusing on the types and origins of the problems of non-native writers, and offers suggestions to help editors avoid or alleviate the problems.*

FEATURES

- 133 **Desktop Publishing—Just What Isn't There?** Ken Pachmayr
- 139 **The Mentor's Apprentice** Morris Dean
- 152 **Feedback: A Multifaceted Approach to Assessing User Needs** Linda Siemers

before any word will allow you to use it as a noun: *The good; the poor; the ups and downs; the ifs, ands, and buts; his kindly and his pleases were learned at his mother's knee.*

There are, of course, sentences without stated subjects, imperative sentences, but

these will be dealt with in a future section on verbs.

### Summary

As soon as writers recognize that certain categories of words, phrases, and clauses

are regularly used in different ways in sentences, subjects being one, they can begin to realize the real scope of English. They can concentrate on the message they must convey, and not worry whether it is *correct* to use a word in a way that is not listed in today's grammar books.

---

# International Technical Communication

Fred Klein, Editor

---

*This forum is open to professionals from all over the world. Different, challenging, controversial, and thought-provoking viewpoints will be presented. Readers are encouraged to send contributions (500 words maximum) to Fred Klein, 1628 N. Courtney Ave., Hollywood, CA 90046, U.S.A. Original articles written in any major natural language will be accepted subject to editorial review and space limitations. Articles not written in English will be translated at no cost and submitted to the author for approval, with the final text published in English only.*

## MACHINE TRANSLATION IN THE 1990s

Muriel Vasconcellos

As we embark on the new decade, it's a good time to take a look at machine translation (MT, defined as "translation generated by a computer with or without human intervention" [1]), and to find out where this technology stands as a tool for technical communication. Has quality improved? Have people's attitudes changed? What's happening in the marketplace? Are there new developments in store? And finally, what is being done to ensure that we learn from our mistakes of the past?

Innovation comes at a snail's pace in this field, but the bottom line is that progress is steady and sure, and clear trends are emerging that can be expected to set the pace from now until the end of the century.

### Quality

Just as MT quality is difficult to measure, so too are the improvements that have taken place. Hutchins [2] recently ventured the opinion that on the whole the quality of MT output has not improved very much in the last 20 years. While it may be true that in the old days we occasionally got "accuracy" rates of 80% and that these levels have not been bettered by much more than 10% in general

texts, the numbers conceal tremendous differences.

To begin with, today's high rates of "accuracy"—if one believes in such a yardstick—are more meaningful because the volumes of text being machine-translated are greater by orders of magnitude. To cite just a few examples in which MT is used to translate texts on a broad range of subjects, the U.S. Air Force reports a volume of 50,000 to 60,000 pages a year [3,435]; the European Commission, 3,000 pages a year [4, 161]; and the Pan American Health Organization, 2.1 million words (8,400 pages) a year. METEO 2, a more specialized application, produces 10.5 million words (42,000 pages) a year [5,453]. Today, high percentages of accuracy—for METEO 2, 97%—are being maintained over thousands of pages, compared with the occasional "show-and-tell" piece 20 years ago. In other words, a single lucky sample with an 80% rating in 1969 doesn't begin to compare with the steady streams of production that we now see in which levels of 90% and higher are a consistent phenomenon.

At the same time, the fact that large quantities of text are being machine-translated means that more linguistic structures, more domains, and more text-types are being handled. General systems, with big dictionaries developed over the years, are coping across the board, and

high levels of accuracy are being achieved in response to linguistic challenges that were not even attempted in the late 1960s. Progress in MT is progressively revealing the true complexity of human language—the messy parts that defy logic (conjunction, ellipsis, anacolutha, fragments, unexpressed inferences, and the like).

Translation programs are now dealing with some of these, but the effect as measured in the output may be negligible. The criterion of "accuracy," if it is understood to be the percentage of text that is syntactically and semantically acceptable without human intervention, fails to reflect the transcendence of the solutions. For example, an instance of correct homograph resolution could be just plain luck—i.e., a simple default translation—or it could be the effect of any of several strategies ranging from a choice based on probability to an elaborate analysis of the context. It could be based on a single linguistic rule or a network of ordered rules. The solution will pass unnoticed if it's correct, and its accuracy will not count any more than the routine fetching of a univocal noun from the dictionary. In sum, "accuracy" does not take into account a program's versatility, which is crucial to its producing texts other than the one being assessed.

In the end, quality should be a function of the purpose for which the translation is being done. It therefore follows that, pre-

cisely because more text types are being translated, MT is useful in a great many more settings than it was in the 1960s, and thus it is correspondingly more valid as a technology.

One last point to keep in mind about the 20-year-old show-and-tell sample is that it was most probably coddled by the protective application of such strategies as careful selection of the input passage, customization of the input text, overly specific coding of the dictionaries, or undercoding so that no choices had to be made between alternatives [6]. Today the public is more aware and less tolerant of "canned" shows. As a result, MT is now often demonstrated on input selected at random.

### Attitudes

Folks in general are more willing to accept the shortcomings of MT than they were in the past, and they are more tolerant of "raw" and lightly post-edited output. It may be that the sheer numbers of people involved in the MT field, coupled with the growth in computational linguistics and AI, have had the effect of educating the public.

There is also increasing concern in the U.S. about the quantities of technical literature being produced in languages other than English. For information that is already stored in digitized form, MT is at the very least acknowledged to be a good scanning device and at best it is looked to as the ultimate solution to the bottleneck, albeit some day in the future.

Because there is a greater understanding of how complex language really is, there is also increasing recognition that MT will take a long time to be developed. Impatience for quick results and impetuous wholesale rejection of MT have been replaced by an acceptance that the task may take decades and that progress in MT is inevitably gradual.

Another change is that linguists and computational linguists are not at each other's throats as they were in the 1960s. There is a sense of *glasnost* and a willingness to work together. A number of projects are now being developed in several countries simultaneously. This cooperative spirit is important for the MT field. In fact, there is growing recognition that at all levels the technology is people-

driven: it is people-driven because MT developers need to collaborate closely, it is people-driven because progress depends on regular feedback from users, and it is people-driven because long-term success depends on the translators, editors, and dictionary updaters who work with MT on a day-to-day basis.

### Market Trends

*Language combinations.* Originally MT was used to capture technical and scientific information in foreign languages and make it available in English. This direction of events was to shift in the late 1960s when U.S. Government funding dried up and the post-Cold-War climate began to foster translations aimed at disseminating information instead of collecting it. In particular, the growth of computer technology was accompanied by a push to sell products overseas, and MT was enlisted to translate the manuals and documentation so that they could be launched not only sooner but also simultaneously in multiple markets.

While this trend of course continues, we are now also seeing a demand for into-English again, especially from Japanese, and for the development of source languages other than English—Japanese, German, French, etc. The commercial companies, understandably, tend to develop new combinations rather than dwell on the refinement of old ones. More Arabic, Chinese, Korean, and other non-Indo-European languages can be expected. *Glasnost* may lead to the marriage of systems that were independently developed [7], and in general there will be a greater trend toward cooperation between teams across international borders.

*Specialization.* The debate continues to rage over the advantages of specialized versus general-purpose systems. Some pundits [2;5] predict that limited-domain systems will ultimately prevail, since they can promise more reliable results and therefore require less human intervention and less care and feeding than do the general systems. On the other hand, the latter, even though they may not be as consistent

in their performance, are directed toward a much broader usership and are therefore more marketable—and more commercially interesting for vendors and venture capitalists.

*Integration.* MT now serves in many cases as a link in the larger text-production chain. Publishers expect the hardware and software to be fully integrated. The incompatibility of foreign character sets has been an obstacle which in the beginning impeded transmission via modem and is sometimes still a problem. Another concern is the preservation of typesetting codes so they will not have to be re-entered manually in the target language. The challenge has been met at the extra-sentential level, but codes embedded within the sentence interact with the translation program itself and are still the subject of research.

*Hardware.* Up to now the “serious” MT systems—those with heavy-duty dictionaries and sophisticated linguistic apparatus—have been running on mainframes or minis. When such systems are finally ported to PCs, the technology will be accessible to the general public and it will be affordable. Small translation services and independent translators will be able to buy it. The litmus test will be whether these professionals can demonstrate that it increases their productivity. If it does, MT will at last be validated. With such an endorsement it can be expected to enjoy brisk sales.

But we don't yet know whether translators will be able to turn a profit because so far none of the high-end MT systems has been ported to a PC for commercial use. If the market is there, what is holding the vendors back? There have been a number of hurdles, but they are gradually being overcome.

Text input has been a problem from the beginning—for mainframe MT as well as for PCs. High-volume MT can happen only when reliable streams of digitized input are available. OCR technology, though steadily improving and coming down in cost, remains imperfect, and false readings undermine the translation results. Only large-scale operations, such as the U.S. Air Force Foreign Technology Division [3], can absorb the cost of monitoring OCR input [8]. Today the task is still too labor-intensive for a lone PC user or even a small translation service.

Another problem with the PC has been that the large dictionaries and programs needed for general-purpose translation were too space-hungry for the configurations that were standard in the early 1980s. Today storage is no longer an issue. However, lookup on the PC is much slower than on a mainframe or a mini. Unless machine lookup is considerably faster than the human translator, the appeal of these systems will be limited. Fortunately, turnaround can be expected to improve a lot with the new high-speed chips and more efficient database-management systems.

Probably the most serious difficulties for MT on PCs revolve around the dictionaries. Customer support for dictionary maintenance, whether done by the vendor directly or by the customer with vendor hand-holding, is labor-intensive and incompatible with mass-marketing strategies. At the same time, updating on networked PCs can quickly get out of control—a situation that may soon be resolved with new software for LANs. Also, MT vendors have been reluctant to offer their large and deeply coded dictionaries on a PC because these data sets, which can be easily purloined, represent considerable investments of time and money. For some systems, a mature dictionary may account for half or more of the total development cost. Smaller MT dictionaries are not the answer: they produce poorer translations and hence make bad press for the technology. These difficulties remain to be ironed out. However, for the most part the obstacles are falling away, and it is not unreasonable to predict that all the major MT systems will be available on some form of microcomputer before mid-decade.

*Applications.* While product manuals will continue to be developed, it's safe to say that there will also be much more use of MT for information purposes, especially in database retrieval. Human translation will be too slow, too expensive, and too hard to come by to meet the need for real-time or near-real-time translation of the quantities of technical and scientific material involved. These factors will motivate the public to accept the unpolished MT product.

All these developments will mean that MT is reaching far beyond the small trans-

lator/linguist community with which it is currently involved.

## Development Trends

A number of new developments are being forecast for MT, some of them more apt to reach fruition than others.

In the practical realm, creative alternatives are being proposed to get around the eternal problem of input. To begin with, we all expect—or at least hope—that OCR technology will improve. In addition, advances in technology should make it increasingly possible to enlist modes and media other than traditional digitized files. Of course, the ultimate goal is speech input and output, or “real-time interpretive MT.” Teams are already working in this area, but the prospects are still quite remote. A low-end type of MT has been proposed in which options are selected from menus by means of a mouse [9]. The keyboard itself will take on importance once again as sophisticated processing and parsing give renewed impetus to interactive MT.

At the recent MT Summit II, Boitet predicted the advent of small, specialized “light” MT applications based on the use of only core linguistic knowledge, specific domain knowledge, and intrinsic semantics [10].

In terms of approaches to the translation problem, transfer architecture now dominates the world of practical MT. Systems based on an interlingua or pivot language so far remain under wraps, but it is to be hoped that they will finally be making their debut in the course of the decade. Meanwhile the use of knowledge bases, either wholly or just to do parts of the job, is gaining ascendancy. The “bilingual knowledge bank” that underlies DLT, scheduled to reach the market in 1993, is an innovative variant of the knowledge base concept [11].

For the development of computer grammars, the notions of *declarativity* and *monotonicity* are widely discussed. Soon linguists will be able to write their programs in a declarative language without worrying about the order in which operations are carried out. In this same vein, parallel processing will make it possible to deal with simultaneous aspects of language expression, and *connectionist theory* will give new insights into how the

human brain can process along several channels at once. Whether or not these innovations will translate into improved MT products for the real-world end-user remains to be seen.

### Progress through Cooperation

Why have some MT systems failed? The technology is acutely vulnerable. It may well be in a class by itself for the long-term nursing that it requires. As I have said above, it relies heavily on people—from the development team to the marketing or promotion group to the staff in the offices that buy and install the systems.

To guard against this vulnerability, it is essential that there be cooperation on all fronts. Each sector—R&D, the suppliers, and the users—has a unique contribution to make to the advance of MT. The *public institutional* sector—government agencies, non-profit organizations, and academic institutions—by virtue of its stable existence is in a position to offer continuity and long-term perspective for policy decisions. It can therefore help to foster development and channel resources toward the highest priorities.

The *private entrepreneurial* sector, in turn, is the principal supplier of MT systems. It is in the best position to sense the real needs and the most promising applications for MT, and to offer products that can meet these needs. The market dictates the channeling of scarce resources, and the vendor has a strong vested interest in seeing that the products are successful.

Finally, it is the *user community* that makes MT happen in the ultimate sense by showing that it can be effective—through years of devotion to scrutinizing the output and building the dictionary data bases. It provides the genius and the sweat, and in the end it makes the case for MT.

There is now a proposal before the MT community to bring these three sectors together in a common leadership initiative—an International Association for Machine Translation, with three Regional Associations, one each in the Americas, Asia, and Europe. Working together, the three sectors can exchange information, educate the public, and contribute to the setting of policy and priorities. It can raise issues, and it can help to avoid duplication

while still recognizing the importance of diverse approaches. MT will be able to survive only if bonds of continuing cooperation are forged between those who develop it, market it, and use it.

### REFERENCES

1. Veronica Lawson, "A Translator's Map of Machine Translation" in *Technology as Translation Strategy*, ed. M. Vasconcellos, American Translators Association Scholarly Monograph 2 (Binghamton, NY: University Center, 1988) 223-240.
2. W. John Hutchins, "Historical Perspectives," in *Proceedings, II International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (Pittsburgh: Carnegie Mellon University, Center for Machine Translation, 1988).
3. Dale A. Bostad, "Machine Translation: The USAF Experience," in *Proceedings of the 28th Annual Conference of the American Translators Association*, ed. Karl Kummer (Medford, NJ: Learned Information), 435-443.
4. Ian M. Pigott, "Systran at the Commission of the European Communities," in *Technology as Translation Strategy*, ed. M. Vasconcellos, American Translators Association Scholarly Monograph 2 (Binghamton, NY: University Center, 1988) 159-166.
5. John Chandiooux, "METEO: 100 Million Words Later," in *Proceedings of the 30th Annual Conference of the American Translators Association*, ed. Deanna L. Hammond (Medford, NJ: Learned Information), 449-453.
6. Claude Bédard, "You Trust Your Mother. But YOU Cut the Cards," *Language Technology* 7 (May-June 1988), 26-27.
7. John S. White et al., "Application of a Natural Language Interface to a Machine Translation Problem," in *Proceedings, II International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages* (Pittsburgh: Carnegie Mellon University, Center for Machine Translation, 1988).
8. Patricia E. Newman, "Information-Only Machine Translation: A Feasibility Study," in *Technology as Translation Strategy*, ed. M. Vasconcellos, American Translators Association Scholarly Monograph 2 (Binghamton, NY: University Center, 1988) 178-189.
9. Jaime Carbonell, Untitled oral presentation at MT Summit II (Munich, August 16-18, 1989).
10. Christian Boitet, "Motivations, Aims, and Architecture of the LIDIA Project," in *Proceedings of MT Summit II* (Munich, August 16-18, 1989), 50-54.
11. Claude Bédard, "The BKB (Bilingual Knowledge Bank): A Promising New Approach to Machine Translation," in *Proceedings of the 30th Annual Conference of the American Translators Association*, ed. Deanna L. Hammond (Medford, NJ: Learned Information), 437-441.

---

Muriel Vasconcellos is Chief, Translation and Terminology, Pan American Health Organization, Washington, D.C. 20037.