M(A)T IN THE U.S.A.

Muriel Vasconcellos
Head, Translation and Terminology
Pan American Health Organization, Washington, D.C.

<u>Resumen</u>: Los traductores en los Estados Unidos ya están utilizando toda clase de ayuda automatizada en su trabajo. El procesamiento de palabras es la herramienta más usada: en tres muestras de miembros de la ATA, los porcentajes de utilización son 53%, 76% y 89%, respectivamente. Algunos traductores también usan <u>desktop publishing</u>. En lo que hace a las telecomunicaciones, en las mismas muestras, los porcentajes de disponibilidad de facsimile y <u>modem</u> son 31%, 49% y 58%. En cuanto al uso de programas especiales para el manejo de terminología, es difícil determinar las cifras exactas; se supone que corresponde a no más del 10% de los traductores. La traducción automática, generada por la computadora, todavía no llega a manos del traductor independiente pero yá se usa en varios servicios de traducción.

<u>Résumé</u> : Les traducteurs des États-Unis ont déjà largement recours à l'automatisation dans leur travail. Le traitement de texte est l'outil le plus utilisé. Les pourcentages d'utilisation révélés par trois échantillons de membres de l'ATA sont les suivants : 53 %, 76 % et 89 %. Certains traducteurs offrent des services d'éditique. Pour ce qui est des télécommunications, 31 %, 49 % et 58 % des membres des mêmes échantillons disposent de télécopieurs et de modem. Il est difficile de donner des chiffres exacts pour l'utilisation de programmes spécifiques en matière de terminologie; on suppose que moins de 10 % des traducteurs se servent de tels programmes. Le traducteur indépendant n'en est pas encore arrivé à la traduction automatique mais divers services de traduction l'utilisent déjà.

1. INTRODUCTION

It's safe to say that translators in the United States are not very dif-
ferent from their colleagues in Canada, Mexico, or for that matter Europe or
Japan. When it comes to applying machine technology to the translation proc-
ess, some are using it all, some are using it some, and some have yet to take
the plunge. Among those who are already using it, many have taken the initia-
tive themselves as part of a growing awareness of the importance of these
tools.

We can think broadly of four types of machine aids that are now avail-
able to lighten the load, speed up the job, or help to improve the quality of
the product: word processing (WP) and postprocessing software, communication
tools, terminology managers, and machine translation. The first three come
under the heading of machine-aided translation (MAT) because they are aids
that support the traditional translation process; the translation is generated
by people as opposed to machines. Machine translation (MT), on the other hand,
is defined as "translation generated by a computer, with or without human in-
tervention." It can be coupled with the other aids, which potentiate its ef-
fectiveness.

2. WORD PROCESSING AND POSTPROCESSING SOFTWARE

2.1 Importance for the Translator

Word processing was the first form of computer technology to directly
benefit the lives of translators. Its multiple facilities for the manipu-
lation of text—for inserting, deleting, copying, moving, saving, searching,
and replacing—have relieved translators of immeasurable amounts of drudgery
and freed them to think about the translation itself instead of the mechan-
ics of getting it on paper. WP is considerably faster than typing on a type-
writer, so that more work can be produced in the same time, and the process
becomes quite painless. Translators no longer have to worry about running
into margins or correcting mistakes. They can quickly go back and change a
word throughout the text, or search for specific occurrences and make substi-
tutions selectively. They can use abbreviations and replace them afterwards
with expanded forms. Their texts are better because improvements can be added
easily without concern for retyping. They can try out different versions and
compare them. And texts can be saved for revision at a later date. Many
independent translators who previously dictated their work are now using WP
and saving the cost of a transcriber. Similarly, in-house translation serv-
ices are able to reduce or eliminate this step.

Word processing packages are now available with almost all the foreign
characters that translators need, and they can be harnessed to printers that
reproduce these characters in hard copy. They often come with spelling check-
ers for English, and foreign languages are gradually being added.

Clients, as well as translators, have a lot to gain from the use of WP.
They save considerable expense if translations can be fitted into their text-
processing chain without rekeying. More and more they are asking—often even
insisting—that jobs be delivered to them in machine-readable form. One of
the first major translation consumers to set this requirement in the United
States was the Joint Publications Research Service.

In the beginning word processors were dedicated machines. Some translators still use these and swear by them. More recently, however, with the mass availability of personal computers, colleagues have found that they get a lot more for their money when they buy WP software and combine it with other PC packages. These may include, in addition to spelling checkers, word counting software, style critiquers, spreadsheets for keeping track of accounts, and perhaps even desktop publishing. The usefulness of these tools will vary depending on the circumstances.

2.2 Usage by U.S. Translators

For our present purposes, we would like to know the real extent to which translators in the United States are using WP. It is easy to conjecture on this point, but difficult to come up with hard data. The ATA has various sources of information on services offered by working translators. Although none of them refers to the entire membership, the samples are of interest because they represent translators who wish to make their services known.

The first source is the Translation Services Directory (7th revised edition, Ossining, ATA, 1988). Of the 704 translators in the Directory who reside in the United States, 369, or 53%, advertise that they have word processing. This percentage may be low, however, because it has been noted that some members who are known to have WP have not included this fact in their profiles. Also, the Directory undoubtedly lists part-timers and translators who are just breaking into the profession, who may not have enough volume to justify investing in a computer. At the same time, many members who use WP do not advertise because they are already fully committed. For these reasons, the Directory may not be typical of the ATA membership.

A much higher percentage is found in the Professional Services Directory of the National Capital Area Translators Association (NCATA) for 1989-1990, where 167 out of 187 translators, or 89%, offer word processing. Yet another source of information is the Equipment Directory of the Science and Technology Division. Here we see 78 out of 103 Division members, or 76%, reporting that they have WP. These two sources may be more accurate than the general ATA directory because they have cells specifically designated for information on the availability of a computer.

Table 1. WORD PROCESSING AVAILABILITY, 1988-1989

Source of data	Membership listed	WP (abs)	WP (rel)
ATA Services Directory	704	369	53%
NCATA Services Directory	167	187	89%
Sci-Tech Division	103	78	76%

A number of translators in the Professional Services Directory also advertise that they do desktop publishing—an indication of their awareness that translation is part of the larger production picture.

3. COMMUNICATIONS

3.1 Importance for the Translator and the Translation Market

Once the translator has a computer, many advantages are to be gained by establishing electronic contact with clients and colleagues. Texts can now be received by fax from any part of the world, translated, and returned to the client either by fax or better yet by modem in machine-readable form. Modems can be built into computers, as can facsimile software as well. These tools are already in worldwide use, and they can be expected to have a significant impact on the profession. International boundaries are disappearing: clients are seeking translators wherever the work can be done best and at lowest cost. Soon translators will be able to live anywhere they want to and provide their services over an unlimited range. Compensation may become more uniform: just as clients are no longer bound to local translation sources, it is hoped that translators in marginalized markets will learn that they can charge the same rates as anywhere in the world.

3.1 Facilities Offered by U.S. Translators

As with word processing, we would like to know the extent to which electronic communication is offered by working translators in the U.S. Unlike the case of WP, the fact that a translator advertises the availability of communication tools may not mean that the equipment is in regular use. However, advertising is probably the best source of information available. A survey on the subject would probably not be representative because the respondents would tend to be those who have and use the equipment.

The Translation Services Directory shows that 217 translators have fax and/or modem communication, corresponding to 31% of the U.S.-based translators in the Directory. The NCATA Directory, in turn, shows that 91 of the members listed, or 49%, have these tools. The Sci-Tech Equipment Directory has the highest percentage: a total of 60, or 58% of the Division members. The details are as follows:

Table 2. COMMUNICATIONS AVAILABILITY, 1988-1989

Source of data	Membership listed	Commu- nications	Fax	Modem	Both
ATA Services Directory	704	217 (31%)	161 (23%)	134 (19%)	78 (11%)
NCATA Services Directory	167	91 (49%)	63 (34%)	72 (39%)	44 (24%)
Sci-Tech Division	103	60 (58%)	27 (26%)	52 (50%)	19 (18%)

An interesting footnote to these figures is that in each group and under each heading women represent approximately half the members reporting availability of communications equipment. Women are not just sticking to word processing; they have rolled up their sleeves and taken on communications as well.

162

The effect of telecommunications can be seen in the geographical distribution of the translators whose names appear in the services directories. The general ATA directory lists, in addition to the 704 translators working in the United States, 33 from outside the country, not only from Canada and Mexico but also from England, Switzerland, Italy, Brazil, and as far away as Hong Kong. The NCATA Directory, in turn, lists 28 translators outside the Washington area—located in California, Delaware, Florida, Illinois, New Jersey, New York, North Carolina, Ohio, Rhode Island, Texas, Brazil, and Spain. These colleagues know that they can easily receive their jobs and deliver their completed work via electronic media.

4.   TERMINOLOGY MANAGERS

4.1   Lexical Data Bases and Their Purposes

For the translator with a computer, the days of the shoebox are gone. A variety of tools are now available to facilitate the tasks involved in maintaining and retrieving terminology records. The choice of tool will depend on the purpose for which the lexical data base in needed.

First, to replace the old shoebox, the translator may wish to set up a personalized file of terms and expressions that are frequently used but nevertheless sufficiently tricky or elusive so that they cannot be trusted to memory. Such a file would probably also include terms that required special research, so that the fruits of the effort can be saved for future reference. One of the main characteristics of this type of data base is that the translator already knows the term is there. The records are entered by the translator personally, and the entry and retrieval may be accomplished using either general-purpose data base management software or else packages, known as terminology managers, that have been specially designed to store, organize, and retrieve terminology records. These packages offer speedy retrieval direct from the word processor, and the found term can be cut and pasted into the text without retyping.

In another type of terminology application, computers can be used to consolidate material contained in assorted glossaries and dictionaries. In this case the translator does not know ahead of time whether or not a needed term will be found. Compared with the inconvenience of dealing with hard copy from multiple sources, a computerized lexical data base with terminology manager represents a considerable improvement for the translator—as long as the system is easy to consult and quick to respond, and as long as the rate of terms found ("hit rate") is high enough to avoid discouragement. Understandably, translators will stop consulting the data base when they repeatedly fail to find the terms they are looking for.

Moving up another notch on the scale of complexity, electronic glossaries can be a valuable support for translators working in teams. These colleagues stand to benefit not only because the lexical data base brings together terminology from multiple sources but also because it helps to promote uniform usage. Also, since the glossaries are easier to modify, they are more likely to be kept up to date. Another advantage is that they can be easily copied and exchanged.

Finally, massive lexical data bases that were originally developed on mainframe computers for use in large institutional settings are now beginning

to be available for consultation by individual translators working on their personal computers.

4.2 Products in Use in the United States

Two terminology managers are currently being marketed in the United States which are specifically designed for the translator working at a stand-alone personal computer. Mercury was developed by the American firm LinguaTech International of Provo, Utah. TermTracer, which comes from INK International BV in the Netherlands, is being distributed by UniWorld Business Systems of Minneapolis, Minnesota. A new product, InterLex by InterLinguistics Unlimited of Philadelphia, is expected to reach the market soon.

Both Mercury and TermTracer are well suited for the first and second type of application.[1] Both provide facilities for the development of a user's dictionary. In addition, they offer a series of ready-made subject-specific glossaries (Mercury has 22, ranging in size from 250 to 4,400 entries, plus several general dictionaries; TermTracer has 60, ranging from 2,500 to 13,000 entries). The user's dictionary is looked up first, before any of the others. In the United States a total of 150 packages of Mercury have been sold since 1985, and LinguaTech estimates that there are 125 active users.[2]

Mercury and TermTracer can be adapted to run on a local area network (LAN), but several features will need to be added before this kind of operation can be fully successful (1).

For the institutional translation service there is ALPNET's AutoTerm, which has been on the market a number of years. It goes a step further and provides automatic term lookup (2). From a machine-readable input text, the terms that have corresponding files in the data base are flagged. AutoTerm is available with large dictionaries in six source languages, and the user can add individualized glossaries as well. It runs on the IBM AT and clones (the latest version runs under OS/2) and should have at least 30 Mb of storage in order to be used effectively. Sales are currently restricted to translation agencies working under the umbrella of ALPNET, but earlier customers who still use it include, among others, the Assemblies of God, the Church of Scientology, the U.S. Department of Defense, the U.S. Internal Revenue Service, the International Monetary Fund, and the University of Arizona.[3]

Some institutions have developed their own lexical data bases. Examples are the World Bank and the International Monetary Fund (3).

There are also the large terminology banks that run on mainframe computers. The Canadian Government's TERMIUM is available for on-line consultation in the United States through the library at Brigham Young University (4), in addition to arrangements made with some of the international organizations. TERMIUM is now on CD-ROM, and it is being made available to the general public on this medium (5). The disk can be read by any IMB XT/AT clone that has a CD-ROM reader.

5. MACHINE TRANSLATION

The status of machine translation in the United States is probably not unlike the situation in Europe and Canada. In a word, MT is being used increasingly in institutional settings, particularly for the translation of prod-

uct manuals from a single source language into multiple targets, but only in exceptional cases is it already a tool for the independent translator.

5.1 Types of MT Systems

At the one extreme, there are MT systems which have been tailored to very specific applications that involve limited vocabulary and a small set of grammar rules. They have been called "sublanguage systems." At the other extreme there are general-purpose systems that deal with a broad range of texts, to the point that Lawson has given them the name "try-anything" systems (6). In between these extremes are specialized applications in which the vocabulary is still not very large but is tied to a more complex grammar, so that at least some variety of input can be dealt with.

The general-purpose systems, historically the first to be developed, were designed to run on mainframe computers. The prime example is Systran, which has been being used as an information-gathering tool at the U.S. Air Force Foreign Technology Division in Dayton, Ohio, for nearly two decades. The size of its dictionaries gives an idea of the range of texts that are covered: for Russian-English there are more than 365,000 entries, of which 210,000 are stems and 155,000 are expressions.[4]

Although all MT systems are gradually migrating to microcomputers, a general-purpose system the size of Systran's Russian-English is still more effective when it runs on a mainframe. However, intermediate systems are now being ported to smaller machines, or developed directly on them, and they function quite well. In 1983 Weidner Communications Corporation made MT history when it introduced its personal computer product, MicroCAT. That system has proved itself in several settings where it has been customized by the user. Another company that produces a PC system is Globalink.

MT research is being undertaken at a number of institutions in the United States. Work on METAL is still going on at the University of Texas. The University of New Mexico is developing a system, XTRA, based on "preference semantics." There is a large project at Carnegie Mellon University in Pittsburgh, where the focus is on knowledge-based MT. IBM is also doing some work in MT, but mainly at overseas sites. Executive Communication Systems (ECS) in Provo, Utah, is working in lexical functional grammar on projects for Tokyo-based Bravice International.

Several low-end products are also being launched on the U.S. market which differ considerably from the systems mentioned above. They have a lexicon of fully inflected forms and a small set of rules for the rearrangement of output strings, with few if any resources for disambiguation. There are also multilingual packages containing phrase dictionaries for the generation of business letters. And of course there are hand-held devices which offer small vocabularies and no linguistic rules at all. These products have different objectives and are targeted at a different usership than full-fledged MT systems. They hold little interest for the professional translator. Usually the input text has to be keyed in, and postediting requires more effort than translation from scratch.

5.2 Role of the Translator in Machine Translation

MT may also be categorized in terms of the type of human intervention

required. In all cases, the task is best performed by a professional translator, although sometimes subject-area specialists are used for this work. The input can be pre-edited, either by customizing the style in which it is written or by reviewing it manually or automatically for ambiguities; the intervention can be interactive, as with ALPNET's TransActive; or the output can be postedited. Postediting, in turn, can be minimal in the case of "gisting" or increasingly more thorough depending on the purpose of the translation.

Translators can also get involved in dictionary-building and sometimes in system development. Certainly for the dictionaries they are in the best position to suggest glosses and lexical rules that specify the environments in which idiomatic or context-sensitive expressions are used.

5.3 MT in the United States

Commercial machine translation was born in the United States, and even today, despite the emergence of vendors in other parts of the world, the U.S. companies still dominate the market: Systran, Logos, ALPNET, Smart, and Globalink. Also Weidner, soon to be replaced by a Bravice line of products developed in Tokyo, was originally a U.S. firm. METAL, being marketed by Siemens in Europe, was developed in Texas.

Even though commercial MT has its roots in the United States, implementation has been worldwide and probably greater abroad, where translation markets are larger, than at home. In the U.S. it is used almost exclusively in in-house translation services, with staff or contract translators revising the output. The oldest and most successful installation is Systran's Russian-English at the Air Force, where the need for postediting has diminished over the years in inverse relation to the growth of the dictionary and is now partially automated (7). Systran is used at a number of other sites at well, including Xerox Corporation, where since 1978 it has been translating product manuals from English into five languages, with others soon to be added.[5] Logos products are used by Antler Translation Services and also by Cullinet Software and Unisys, in addition to installations overseas.[6] Smart, which focuses on technical manuals, currently has a large project with Ford Motor Company.[7] The firm specializes in marrying their product with desktop publishing software. Globalink provides MT support for its own translation services and has been selling its PC product to clients in the U.S. Government, among others. U.S. customers of ALPNET's TransActive include the Assemblies of God, the Internal Revenue Service, the International Monetary Fund, and the University of Arizona.

In the non-commercial arena, the Pan American Health Organization uses SPANAM and ENGSPAN, developed by PAHO's own staff, for the translation of texts in several disciplines. In addition to in-house personnel, some 12 translators work as contract posteditors on an as-needed basis.

6. M(A)T AND THE TRANSLATOR OF THE FUTURE

There is a need, and there will be a market for, any MT product that runs on a PC and accomplishes its purpose quickly and effectively. It is reasonable to expect that the older systems which still run on mainframes will be being downsized in the near future, if the process is not already under way. Once they are available, the consumer will be faced with a bewildering array of options. Not only will choices have to be made between systems, but

166

the mode of work and the extent of human intervention will have to be decided on as well. Translators will be needed as posteditors, dictionary-builders, and terminologists, and they may be called in and asked for criteria for evaluating systems. At the same time, free lances will at last be able to try out these systems for their own work, and some will undoubtedly find that their productivity is greatly increased.

NOTES

[1]See the three-part review articles by Chapman in the ATA *Chronicle*.
[2]Alan Melby, June 1989, personal communication. Sales of the product in Europe under the name of Termex are not included in these figures.
[3]Larry Childs, June 1989, personal communication.
[4]Dale Bostad, June 1989, personal communication.
[5]Alan Portela, June 1989, personal communication.
[6]Peter Wheeler, June 1989, personal communication.
[7]John Smart, June 1989, personal communication.

ACKNOWLEDGEMENTS

REFERENCES

(1) Chapman, Don. A Look at Terminology Managers. Parts 1-3. ATA *Chronicle* Aug-Oct 1988.

(2) Good, Robert L. Automated Lookup: AutoTerm of ALP Systems. In: *Technology as Translation Strategy*, ed. M. Vasconcellos. Binghamton (N.Y.): SUNY Press, 1988. ATA Monograph Series 2. pp. 87-91.

(3) Alvey, John. Patrice Guilmard. Automation of Terminology in the International Organizations. In: *Technology as Translation Strategy*, ed. M. Vasconcellos. Binghamton (N.Y.): SUNY Press, 1988. ATA Monograph Series 2. pp. 68-72, 73-75.

(4) Melby, Alan. Introduction to Section II. In: *Technology as Translation Strategy*, ed. M. Vasconcellos. Binghamton (N.Y.): SUNY Press, 1988. ATA Monograph Series 2. p. 53.

(5) Larichellière, Monique. Termium on CD-ROM. Paper presented at the 29th Annual Conference of the ATA (Seattle, October 12-16, 1988).

(6) Lawson, Veronica. *Practical Experience of Machine Translation*. Amsterdam: North-Holland, 1982.

(7) Bostad, Dale A. Machine Translation: The USAF Experience. In: *Proceedings of the 28th Annual Conference of the ATA*, ed. Karl Kummer. Medford (N.J.): Learned Information, 1987. pp. 435-443.