

Translation: Aids, Robots, and Automation

Alexander Andreyewsky

Volume 26, Number 1, mars 1981

L'informatique au service de la traduction
Machine Aids to Translation

URI: <https://id.erudit.org/iderudit/001888ar>

DOI: <https://doi.org/10.7202/001888ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Andreyewsky, A. (1981). Translation: Aids, Robots, and Automation. *Meta*, 26(1), 57–66. <https://doi.org/10.7202/001888ar>

Translation : Aids, Robots, and Automation

ALEXANDER ANDREYEWski

...Man becomes all things by understanding them (*homo intelligendo fit omnia*)... (or)... man becomes all things by *not* understanding them (*homo non intelligendo fit omnia*)... When man understands he extends his mind and takes in the things, but when he does not understand he makes the things out of himself into them... Man in his ignorance makes himself the rule of the universe... he has made of himself an entire world.
G. Vico, *Scienza Nuova* (1744)¹

INTRODUCTION

This paper examines electronic aids to translation, or computerized services and facilities, both as ways and means to automate translation, and as an approach to finding a solution to the problems resulting from the shortage of qualified translators in the face of a growing information overload. At present, electronic aids to translation can include *tools*, like word processors, *systems* for *machine translation* (MT), and MT-less systems for *machine-aided translation* (MAT). While tools are effective in automating translation, MAT systems are additionally useful in meeting many of the other challenges. The ability of MT systems to be an aid in meeting either of the two objectives is unclear. The proposal advanced in this paper is that MT is by its very nature *robotic*, and ideally suited for limited applications with a finite number of clearly defined states or situations. Since new situations (relative to the system) cause robotic MT systems to fail, in applications with a sufficiently large number of new situations MT systems cease to be aids, or even tools, and may *impede* translation. This has far-reaching implications examined below.

The proposal advanced in this paper is to view a spectrum of MT systems ranging from *robotic* MT to non-robotic traditional MT. An important new development in MT are various efforts limited to sub-languages of the type currently being developed at the University of Montreal. Although such efforts are not necessarily considered robotic, they approach the goals of robotic MT and can be called *semi-robotic*. Historically, MT is the “oedipal mother” of all efforts to automate translation. Yet, logically, the MT-less MAT must precede MT in the development cycle and continue to serve as a back-up and feeder system for all MT efforts.

1. G. Vico, *The New Science of Giambattista Vico*, Anchor Books edition cited, p. 88, paragraph 405.

ROBOTIC TRANSLATION

Robotic MT, or Bar Hillel's "Fully Automatic High Quality Translation" (FAHQT), is possible in environments otherwise conducive to robotization. The applications must be clearly defined, well-worked out, and otherwise designed to involve a finite number of states or situations. For robotic MT to become fully developed, one would need in many cases an *artificial intelligence* (AI) component. At the present time, lacking AI components, *applied* robotic MT can be a possibility in all applications which admit non-MT automation, for example devising a menu-selection network program. The closest to a working MT program resembling robotic MT is Canadian METEO, which translates weather reports. One can think of applications much more limited than that but with a higher frequency of usage, for instance ticket sales, or other more ambitious with lower frequency of usage but greater volume and precision of results required: translation of manuals for the Volkswagen "Bug" may have been, in retrospect, one such example. The purpose at this time is to establish the notion of robotic MT which will require further study at some future date.

Taking some hypothetical robotic MT system, it is important to recognize that one must expect it to be *reliable* and produce some output for a given input. For every output, it is important to know how *accurate* it is. If *reliability* can be measured on a hypothetical scale of zero to 100 per cent, the *accuracy* of each output can be measured on a scale of zero to 100 per cent or it may be additionally weighted to account for the possibility of machine-introduced errors and have even negative values. The accuracy index multiplied by the reliability index (using decimal equivalents) results in the system's *performance evaluation index* (PEI).

For a true robotic MT system, the *hypothetical* value of the PEI indicator must be 100% in the ideal case and can be somewhat lower in practice. If the system's performance justifies this, the PEI number can drop to zero, with possible negative values disregarded. In this connection, it is important to note several likely thresholds. For example, at PEI = 80%, for the system to continue to perform (this being the *sine qua non* of robotic system) considerable human intervention will be required. At PEI = 50%, it may prove *cheaper* to ignore the MT system output and let the human backup team take over entirely. At PEI = 20%, except in experimental situations, it may turn out that continued reliance on system-produced output will actually result in an instance of *machine-impeded translation*.

Although the PEI numbers *are* hypothetical, the difficulty of calling an MT system of the sort just illustrated an instance of MAT is quite real. If one considers the role played by the human backup team in the PEI = 100% case and the PEI = 20% case, the change in Marxist terms is one of quantity becoming quality. More plainly put, the two situations are no longer commensurable. While *semi-robotic* MT research may help establish the actual boundary

between robotic and non-robotic MT, the interim criteria must come from a PEI or a measure like it².

MACHINE TRANSLATION AS AN AID

There is an obvious way in which robotic MT can be helpful in *replacing* translators and reducing volumes of translation and, therefore, become an *aid* in some sense of the term. However, it aims at the complete automation of the translation process and is the *total solution*. The possibility to consider is the extent to which non-robotic MT can be of help in obtaining a *partial* solution by having *human-assisted machine translation* (HAMT)³ or a combined operation involving humans and robots. Inasmuch as the second possibility seems to be emerging as an alternative to the first, it will be dealt with next briefly.

Man-Robot Interaction?

The use of tools to automate *tasks*⁴ such as using a rolodex file to help find telephone numbers or having a telephone answering device that activates a tape recorder, is not the same as having a robot which performs a certain *function*, even if the robot requires only what C. Evans⁵ describes as the "intelligence of an earwig". *Robots are intelligent machines*. In this sense, much of the robotic MT without an AI component is really not even robotic. Building an MT system to automate foreign-language exchanges in conjunction with ticket sales at a box office (leaving the incongruity of the task aside) when one can have alternate solutions is a case of using elephant guns to shoot mice. Adding a "grammar" component may well be the beginning of AI. This, however, is a matter that requires further study. The usefulness of an MT system depends not only on the *level* of reliability, but also on the ways in which it can fail. The situation can be compared with the case of a bank employee equipped with an electronic calculator which does not accept any sequences of "4" and "9" and which sometimes mysteriously fails to work at all and, when it does, requires checking the results by hand, anyway. If there are no other devices around with a high enough PEI, the odd calculator may serve to double-check the results obtained by an employee, provided s(he) is constantly on guard. The minute manual calculations become careless or undue reliance is placed on the calculator because it is "almost" always right. In order to be able to judge whether or not this is cost-effective one would have to develop a PEI

2. A forthcoming paper, not released for circulation at the time this paper was written, deals in some detail with the subject of MT performance evaluation and a comprehensive discussion of the subject should be deferred until the paper is published.
3. The HAMT acronym is technically correct, but it never caught on and is generally replaced in this paper with references to *traditional* MT.
4. The present writer's understanding of automation as it applies to translation has greatly benefited from the ideas M. Zisman of MIT, President of Integrated Technologies. His interest is sincerely appreciated; however the present writer is solely responsible for any errors. See, for example, his "Office Automation in a Machine Translation Environment", *Computer Support to Translation* (seminar proceedings), Foreign Broadcast Information Service (May 1978).
5. C. Evans, *The Micro Millenium*, New York, The Viking Press, 1980.

which would make it possible to compare operations so equipped with all-manual operations and weigh the risks involved. The standard response of MT advocates is that the translation field is fast running out of "bank employees", the translators. In this case, the hypothetical bank may well be concerned with its *data* processing methods, if a perfect balance at the end of the day is the kind of *information* it needs.

The devices just mentioned suggest that robotic MT is *not* fail-soft. The human backup is not interacting with the machine in any real sense of the word and there are other problems to be mentioned only briefly for lack of space. Currently, robotic MT can hope to target on a specific application or a class of such applications. As soon as one attempts to shift a given application-specific *model* to a new application, there is an obvious way in which new situations may arise in the special sense used above. However, a shift to a different environment may involve subtler changes quite probably not reflected in the linguistic analysis, which in the case of robotic MT needs to be examined carefully at some future time. The key point is that a given model must be viewed in the *environment*, the "world", in which it operates. Thus, geographic factors and time-period differences alone can affect a robotic MT model in obvious ways⁶.

Man-Computer Interaction

The main objection to man-robot interaction is that a robot which fails is not a tool in the sense a word processor is a tool. This raises questions regarding the feasibility of semi-robotic MT: Can it be an aid? The answer, basically, is one involving the outcome of work being conducted at the University of Montreal and elsewhere. Although it would be easier to prove empirically, it seems that it would be extremely hard to overcome the obstacles mentioned. There are several ways the situation can be dealt with, including text pre-processing and significantly more advanced grammars. In many respects, the key consideration is the extent to which such grammars can move in the direction of AI. The economics of the situation is very hard to discuss, lacking specific data.

It should be clear that the purely "engineering" solution of reducing the scope of areas covered is not an answer. The problems inherent in what has just been noted are a matter of the *design* of MT systems and the underlying model of the *translation* process adopted by MT developers. The effort to solve the translation problem by eliminating translators overlooks many alternative solutions and uniquely duplicates the problems of a single translator working in a vacuum and engaged in little more than vocabulary substitution using a dictionary.

6. The first who failed to see this among the economists seems to have been none other than Adam Smith, who while writing about the Invisible Hand failed to notice one of the biggest monopolies of the day prospering a few miles away. In addition to the fact that the "same" thing can be done differently at different historical times or different geographic locations, robotic MT may suffer from such relatively recent ills as rapid obsolescence of terminology.

The consequences of the oversimplified approach to the *model* of the translation process inherent in MT are well known and do not require elaboration. It should be emphasized that robotic MT is not a viable possibility, at least for the next several decades. In addition to everything that has been brought up so far, an evaluation of the MT potential should take into account the facts considered next.

MACHINE-AIDED TRANSLATION

MT-less MAT is important as a system of integrated tools or equivalent capacities and as a model of the translation process which can be developed around it. While offering practical solutions and serving as an aid to translators, MAT can open new opportunities for increased automation, if desired or necessary. Although it can be argued quite validly that MAT does not help to alleviate the shortages of qualified translators, it opens the way to looking for alternative solutions to the translation problem in a changing environment. The need, it will be seen, is to move from the "microeconomic" view of translation within the context of a single operation to the "macroeconomic" environments of the translation field and related areas of language — and information processing. Much of what follows has been discussed in other sources and is given here only by way of update and elaboration of topics relevant to the present discussion⁷.

A Model of the Translation Process

Taking a black box approach, the distinction must be made between the translation *process* or processes going on in the black box and the resultant product. The facility with the black box at its core is a *translation operation*, which generates, in addition to *translations*, an important by-product, i.e., *terminology*. Since only the production of translations is of interest here, it is assumed that each operation has other functions relating to administration, sales, and so on. For the purposes of what follows, the concern is with only one other operation, the *terminology data bank (TDB)*⁸ which is involved in transactions relating to terminology on a quasi-commercial basis.

a) *Translation*

Ideally, the black box involves a team effort of at least one translator, one terminologist, one editor, and one graphics expert all of whom are backed by appropriate computer facilities. Although graphics operations, which include all of the word processing operations as well, are not directly related to the ques-

7. A. Andreyewsky and D. McCracken, "Automating Translation of Languages in the 1980's." (technical report), Center for the Study of Translation, Carnegie-Mellon University (August 1979), (to appear) and other publications of the Center. The prototype model for MAT below is the Center's system of Computer Aids for Human Translators (CAHT), which represents a class of such systems.

8. In the sense CAHT is a model for MAT, a possible model for a TDB is the Carnegie-Mellon University TDB. The most recent description is in Andreyewsky and McCracken, "The Carnegie-Mellon University Terminology Data Bank" (paper presented at the Summer 1979 Dartmouth Conference on Computers in the Humanities).

tion of computer aids, they do account for a sizeable share of the costs, and any economies in this regard can go toward solving the translation problems.

The terminologist-translator-editor triangle represents schematically the process of translation and the tasks involved in it. Even this oversimplified view suggests the need to break down aids by categories. While an MAT system insures access to word processing capabilities and a terminology database optionally supported by a TDB, providing for an improved communication capability between all concerned, there are special considerations in each case.

Improvements in the work of a translator can be accomplished by an improved terminology and reference service⁹, access to specialists more generally, and access to previously translated materials. The productivity of a translator can be improved by changing his work patterns and training in the use of dictating equipment, among other things. Most importantly, the right kind of translations can be given to the right person, taking into account his or her training, past experience, and language skills.

The aids to terminologist and editors, who are typically expert translators themselves, include much of what was just mentioned but may involve other aids. This is a generally neglected area which requires further study.

The facilities of an MAT system provide a model of a translation operation designed to improve the *efficiency* and *effectiveness* of the operation and of the people involved through automation. The MAT system provides the vehicle and a "second-order" tool to monitor the results achieved in this regard, and to direct future efforts¹⁰.

b) Terminology

Terminology is important not only as a by-product of translation activities (and thus as a valuable form of feedback) but also as a resource with a considerable potential exchange value. Although, it seems that most translation operations do not single out the value of terminology, treating it a by-product, it is likely that the replacement cost of each entry, consisting of a word or a phrase and its equivalent, is currently about \$10 each, on the average.

-
9. The need to develop such a service in the United States, patterned after several successful efforts in other countries, has been argued in detail in sources cited in the two preceding notes. The most recent effort to develop a TDB in the United States after the Carnegie-Mellon University's TARGET project, which dates back to 1975, is the recently announced Weidner Communications effort. While all such efforts should be encouraged, the ultimate solution must involve public support and international cooperation between friendly and unfriendly nations alike, not withstanding Marguerite de Valoi's advice (In love, as in war, a fortress that parleys is half taken) which gave rise to the Italian *castello che dà orecchia si vuol rendere*, reliable and accurate communication is a matter of survival. An insight into the magnitude of the problem involved is provided in the section on terminology in this paper.
 10. By putting "clocks" on each operation, an MAT model can help to establish the best way of doing a particular job (a matter of tactics and efficiency). It can also suggest ways of best using the time (a matter of strategy and effectiveness). Putting the *what* before the *how*, efficiency before effectiveness, is not the optimal way. The same applies to efforts to improve translation before exploring the ways that one can solve the problem by eliminating translation.

The \$10 cost is estimated on the basis of an average 5,000 entries per man-year productivity over protracted periods of time. The resultant entries are the "best effort" but are not necessarily of publishable quality, because the quality and accuracy of an equivalent rises in direct proportion to the attested instances of usage, which takes time. The substitute for this is extensive and costly research which cannot be accomplished in the 24 minutes average allocated per entry. Anything produced in less time is likely to border on the marginal¹¹.

In addition to being a potentially considerable expense, the accumulation of terminology requires a great deal of lead time, which is a very important consideration in the case of MT efforts. The need to add system-specific coding and to process the entry to suit a given system increases the expense, which is not a retrievable cost. Moreover, most current MT entries are severely restricted in their meanings and are for all intents and purposes unusable by humans and not exchangeable with other systems. Because of this, any MT system must have a humans-usable dictionary and other facilities required for human backup, which rapidly becomes an equivalent of an MAT system as one moves away from robotic applications.

The prospects for the future are quite disquieting in this regard. Although it is almost impossible to predict required dictionary sizes with any degree of precision in each particular case, non-MT translation efforts are known to have reached the one million entry mark per language and are likely to rise even higher. If one considers that each entry refers to an underlying *concept*, the estimated number of concepts per culture, the *inventory* of that culture, has been put at 10 million. The inventory of the world, not counting most geographical and personal names may reach several billion by the year 2000¹². Much of this rapid terminology growth is a function of the information explosion and the resultant information overload discussed elsewhere. However, given the existence of such rapid rates of growth, it is to be expected that dictionary coverage in most areas will never be adequate. Unless one is able to develop dynamic databases, which are impossible without a TDB with a broad base of support, the costs per entry in all translation operations will climb to multiples of the \$10 price tag suggested above¹³.

11. Per-entry costs are, obviously, not discussed. The \$10 figure at current wages in a Fortune 500 company is likely to be closer to \$15. One should not be surprised at multiples of \$15 in MT operations. The additional factors to consider include : (a) the cheapest is not always the best, and poor entries will cost many times this amount to correct and remedy in the output, human or machine, and (b) the number of times a given entry is utilized, because most entries are used very infrequently, especially if frequency of use is plotted over the effective lifetime of the entry, which is probably about five years, on the average. For this reason, a \$100 entry in robotic MT can be cheaper than a \$1 entry in non-robotic MT.
12. All the figures and sources are given in Andreyewsky, "Terminology, Automation, and the Information Overload" (to appear).
13. All of the figures given for terminology apply to averages. The newer the concept, the harder the search, and the ultimate cost. Cooperation on matters of terminology may not appeal to a single firm. However, *if* the translation problem is as severe as it is portrayed, economies of scale on a national level can make quite a dent in this area alone, especially in the United States, where terminologists may have to be put on the endangered species list.

A MODEL FOR THE FUTURE

The broad base support required for a TDB, which in itself is not an unknown aid, introduces the need to consider the potentials of telecommunications networking and pooling of resources. The ability of an MAT system to put translation operations on-line and the prospects of having access to networks requires looking at some of the components outside the immediate concerns of translation tasks as these have emerged historically.

By its very nature, translation cannot be an end in itself and is done for some specific application. In this sense, there are as many kinds of translation as there are applications. If one looks at an estimated 300 fields of specialization and the need to work in as many as 60 languages, or 3 540 language pairs, one has over a million possible language-specialty combinations¹⁴.

The translation jobs themselves vary a great deal. Translating from sources that one has no control over is quite different from translations from sources that can be controlled. Ultimately, each channel of communication tends to develop its own shorthand for the most regularly and frequently used concepts, expressions, and other data. The "periphery" must be regularized as much as possible and made unambiguous and "understandable" even to computers. This can be accomplished by pre-processing texts or working toward a standardization of usage. With increased automation, the otherwise unenforceable requirements can become a reality, which can go a long way toward improving the cost-effectiveness of translation activities¹⁵.

Not all translation jobs can be treated in the same way; custom-made solutions must be sought. Yet, one of the fundamental problems remains the attitude toward translation and the lack of familiarity with what is involved in obtaining quality results. Like medical assistance, translation can be made available "right away" at the *scheduled* time or on an emergency basis. No matter what facilities are built, one must have *reserve* capacity but no *idle* capacity. Networking can go a long way in providing some of the solutions in this area as well. One can expect an increasing number of greater and greater overloads to develop. It is a managerial decision and duty to decide whether such overloads are the result of poor planning or developing trends. Inevitably, it is a combination of the two that is responsible for overloads, with a fair share of attested instances due to lack of planning.

14. The idea to train specialists as translators is neither new nor unique. For example, it is easier to train an electronics engineer to be a patent attorney than to train an attorney in the field of electronics. Pity, there is no machine proposed for this application yet, but C. Evans (see Note 5) seems to be optimistic that it may yet come to that.

15. Standardization of usage in various technical areas is being worked on nationally and internationally, because of the enormous costs resulting from the lack of standardization. There have been many efforts to develop a World Language, starting with Galen. Presently, large companies are taking the lead in trying to develop at least the same "dialect" within a single firm (a fine example of such an effort is the *NCR Fundamental English Dictionary* published in 1978). Such normalized usage, whatever the way it is instituted, is an important aid to translators and can save a great deal of money, both in translation and in other areas, if taken as a matter of national concern.

As translation jobs numbering in the scores of millions of words emerge worldwide¹⁶, it is essential to think about the future and develop the necessary arsenal of electronic aids. However, "translation only" solutions are not economical and one must view them in the broader context of the emerging technological environment and the requirements in other areas of language — and information processing. Although one can build systems, networks, and networks of networks, it is important to look at the challenges involved as these were sketched, in part, in this paper. One of the serious challenges is terminology. Although in any one instance it is likely to account for about five per cent of the total effort in terms of manpower requirements, the lead time required to develop the terminology and the results of poorly prepared terminology can cause problems far in excess of anything comparable to that proportion. Besides, much of the work required to develop multilingual terminology will also be required to produce monolingual terminology and vice versa. This raises the matter of the need to do something about automating terminology development and aids to terminologists.

CONCLUSIONS

Robotic MT is an aid to translation to the extent that it can *replace* translators with machines. MAT is an aid to translators and *displaces* them with machines, which is what is done by various tools as well¹⁷. All MT relies on human backup and the critical need for it increases to its maximum point in the case of traditional MT. Unless there is a conscious effort to move into the direction of robotic MT, the belief that traditional MT is in any sense an MAT can seriously backfire.

The chicken and the egg argument must be settled in favor of MAT. MAT is the first step toward automation; robotic MT is one of the possible spinoffs. However, without AI components or functional equivalents, robotic MT applications also allow non-MT automation. The robotic objective is a worth-while experimental goal, but not a practical solution, except in very limited applications. Beyond that, MAT must remain the vehicle for further automation.

16. Published data regarding the size of any one specifically identified project is hard to come by. Like the cost of heavy equipment, which used to be estimable in dollars per pound, translation volumes were estimable from the dollar amounts of a given transaction. However, with the rising costs and computerization it is difficult to offer such estimates. 100 million words per translation job or a company's annual output are not uncommon. Such overloads are important not only as a technological challenge, but also as a matter of national concern, especially if the total translation output of a country is not much higher than that on an annual basis.
17. The present writer owes this distinction to B. Teague, President-Elect of the American Translators Association. While translators do not seem to object to being displaced by tools and having machines do the hard work, the same does not hold for the feelings about being replaced, or, one may add, *replaceable* by machines. The subject of translators and automation requires study, for it is much more complex than stereotypes tend to suggest. Opinions may differ, but it is the present writer's view that the impact of aids can be quite negative if automation is viewed as leading to translation factories with translators playing the role of production-line mechanics. Yet, such seems to be the trend and this a possible nemesis that needs to be analyzed carefully.

Although the distinctions between MT and MAT are important, the future will bring other challenges with translation no longer a special case, but a part of the broader goal of language-encoded information processing. Terminology development and dictionary production is one of the areas where this is becoming quite apparent.

Existing aids to translation do not exhaust all of the computerized services and facilities possible that can help. However, along with this awareness, there must be a clear recognition that just as there are aids, there are also impediments, electronic and otherwise. A thorough understanding of what is involved in translation can make the critical difference.