Document generated on 07/17/2025 11:19 p.m.

Meta
Journal des traducteurs
Translators' Journal

META

Some Notes on Machine Aids for Translators

Thomas Schneider

Volume 28, Number 4, décembre 1983

URI: https://id.erudit.org/iderudit/004048ar
DOI: https://doi.org/10.7202/004048ar

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

Schneider, T. (1983). Some Notes on Machine Aids for Translators. *Meta*, *28*(4), 344–351. https://doi.org/10.7202/004048ar

Tous droits réservés © Les Presses de l'Université de Montréal, 1983

This document is protected by copyright law. Use of the services of Érudit (including reproduction) is subject to its terms and conditions, which can be viewed online.

https://apropos.erudit.org/en/users/policy-on-use/

érudit

This article is disseminated and preserved by Érudit.

Érudit is a non-profit inter-university consortium of the Université de Montréal, Université Laval, and the Université du Québec à Montréal. Its mission is to promote and disseminate research.

https://www.erudit.org/en/

SOME NOTES ON MACHINE AIDS FOR TRANSLATORS

THOMAS SCHNEIDER

Now and then technical translators are admonished by the powers that be that their main purpose in life is not to play around with expensive machinery but to translate — in which case "translate" is usually understood to mean the literal rendition of a text in a different language. Should the translator ever encounter any difficulties, there would always be an abundance of highly respected, time-proven dictionaries in the book-stores if not on the office shelf.

The absurdity of such conceptions needn't be elaborated on in this forum. On the other hand, there is an aspect to be considered : machine aids which are not geared to the practical work of the technical translator may well turn out to be nothing but expensive toys. So the translator has to determine which specific tools out of the scope outlined below are suited to his particular environment and whether their use is cost-effective for his particular application.

The indispensable basis for other, more elaborate machine aids are word processing systems. They permit the generation of source texts in machine-readable form, i.e. on a medium suited for further processing, be it floppy disk or magnetic tape. There is no compelling reason why tech-writers should always have to compose their documents on hard-copy first and then, in an additional step, have them stored on an electronic medium. This additional step is of course a potential source of random errors, not to mention the fact that corrections and other changes are much more easily incorporated on the video display terminal than on the typewriter.

A possible alternative to creating the original text directly on the screen would be to read in the typed paper copy via optical character reader. Without access to a large programmable unit with self-learning capabilities, however, one would have to rely on standardized font and format of the typed or printed pages, and that will not always be easy to coordinate, especially when dealing with a multitude of heterogeneous sources. Otherwise there would be additional expenditure: machine-reading the text, proofreading the printed output, correcting recognition errors, plus the problems associated with moving back and forth between systems.

Like the writer of the source text, the translator can benefit from composing his translation directly on the screen rather than on paper. Apart from general ergonomic considerations, one of the basic requirements for the screen as a translator's tool is that it be large enough to show sufficient context. 60 lines on a high resolution screen seem satisfactory for most applications. A large screen permits the translator to use several text files at the same time on the same screen, e.g. source text, translation in progress, terminology search etc. This

does not mean that the translator *must* work with a split screen if he prefers to work from a hard-copy source but at least he is free to choose.

Another necessary feature of such a word processing system is the facility for comfortable text editing, i.e. the system should offer possibilities to correct not just individual letters but to skip on a word, sentence or paragraph basis, to exchange words singly or by global edit, to rearrange phrases, to reorder paragraphs etc — and all that by pushing a button. If the translator has to enter long chains of commands for the purpose, the capabilities of the word processor will remain largely ignored and the system will prove uneconomical.

Another prerequisite seems trivial but is frequently ignored since a high percentage of data processing devices originate in English-speaking countries. English is actually somewhat of an exception among languages in being devoid of diacritical marks, but word processing systems for translators must be able to show on screen and process *all* characters of various languages: accents, umlauts, Spanish ñ, etc. The expenditure for manual insertion of special symbols or complete reinput can only be saved if after translation the texts are available again with all special characters and diacritical marks.

Having texts available in machine-readable form can save a lot of unnecessary work. A large part of the translations in industry consists of product documentation. Often new versions are produced which deviate from the earlier ones in only a few details. If the translator has access to a machine-readable file of the earlier document he can start a machine run to find the text segments that are actually different. If, instead of the total one thousand pages, he needs to translate only those forty pages that are new and can copy the rest of the document automatically he has made a lot of headway. Helpful for such purposes is a sophisticated filing system for source texts and finished translations that permits easy retrieval of the relevant passages and recomposition of the whole document. Such systems are commercially available (e.g. SCCS in the IS/1 environment).

The second category of machine aids for translators concerns the problem of terminology. Translators are not usually paid high salaries to translate trivial everyday smalltalk letters into a foreign language; for most secretarial positions, such tasks are nowadays included in the job description. Technical translators, on the contrary, have to work in subject areas and at a level of exactness that could be mastered only after a lengthy degree program. To give a measure of the complexity of the task: Siemens Language Services count on university graduates needing at least another two years of specific training before they are considered adequately prepared for the normal translation work in the department.

First of all, the translator must understand the content of what he is to render understandable to foreign users. His readers may sometimes have radically different levels of education, and though it is an impossible task to fulfill, the translator is expected to improve on the original text and clarify the content for a heterogeneous group of readers. But then: how is he supposed to transfer the many, often new concepts into equivalent symbols of the target language? Even if we exclude "exotic" languages with their innate problems, in

most cases dictionaries are — for such specific uses — not available or simply outdated the very minute they reach the bookstore. In some sectors, technology is advancing so rapidly that most technical products are less than five years old, and keeping up with all the technical terminology of a field is next to impossible for an individual.

Keeping in mind that a technical translator may spend up to 70% of his time on terminology research, one would invite disproportionate waste if one were to allow more than one translator to research the same term independently of one another. Such duplicate effort however cannot be prevented if every translator individually administers his own index cards; the problem can only be solved by a common terminology data base which is utilized collectively.

Such a terminology data base must fulfill several requirements, *inter alia* it must be

— current (new entries have to be available for interrogation immediately)
— correct (no ad-hoc entries should be included without due scrutiny)
— flexible in regard to interrogation (it should be possible to retrieve the information on the basis of string comparison, by language pair, source, subject area etc, as hard copy glossary, through direct access via terminal)*

A terminology data base in turn is a prerequisite for further machine aids: As an alternative to direct interrogation of individual terms, a *text-specific glossary* may prove useful. The machine-readable text is lemmatized automatically and compared with a system dictionary (a subset of the data base). All terms in the text that are found in the dictionary are marked with a special symbol; at the end of the text a glossary of these terms is printed out (or shown on the screen).

Such a procedure has the advantage that, in contrast to a subject area list, the terminology offered actually does occur in this specific text, that in this manner the size of the glossary is greatly reduced and more manageable. Also, in working on the text, the translator knows immediately whether a certain term is listed in the glossary and whether it is worth his while looking there. The number of "failures" in dictionary lookup is reduced which may well have an effect on the translator's attitude towards the data base.

Such a program has another application. Especially in the realm of technical documentation, translators are often victims of the verbal creativity of tech-writing engineers. In such cases, the supplied text can again be compared with a system dictionary, e.g. one containing the set of authorized terminology. The program will output all deviations from this standard dictionary. On the basis of such a list, one can check individually whether indeed a new concept need be expressed in a new term or whether the writer followed his own maldirected literary ambitions. Of course, spelling errors can be found in this manner as well.

Such a test not only improves the source text but also lightens the load on the translator who doesn't have to spend as much time on verifying and

* Since the Siemens terminology data bank (TEAM) has been described in detail elsewhere, it won't require a lengthy treatise here. Cf. J. Schulz, "A Terminology Data Bank for Translators (TEAM)", *META* Vol. 25, No 2, June '80, p. 211-229.

searching for pseudo-synonyms. Moreover, checking source and target language texts against an obligatory set of terminology puts teeth into recommendations to standardize terminology on a larger scale (which would probably be the greatest contribution to communication this side of the Bible translations).

Similarly, the translator's work is greatly hindered if the source text is unintelligible. Many a writer seems to view himself as an heir to Kantian complexity (in style, that is), striving for elongated constructs overflowing with prepositional phrases and a multitude of subordinate clauses, thereby occasionally losing the thread, forgetting the verb or giving birth to semantic absurdities.

A program to test for the *complexity of sentences* can be put to good use here. With a simple algorithm, it checks sentences for the number of words, for overdoses of prepositions and (in German) length of the span of the verb frame, e.g. between modal and infinitive. In such a way, sentences which are probably unintelligible on the basis of syntax alone, can be marked and sent back to the writer.

Of course, there are certain matters which must be expressed in complex sentences, or else their meaning is distorted. Qualifications of a philosophical or logical statement, for example, must be included within the sentence boundaries. That however is usually not true in general or technical writing. Here, syntactic complexity is frequently not the result of necessity but the consequence of sloppiness and utter disregard for the reader. Thus an important text can become useless for the poor soul who has to work with it.

A psychological advantage lies in the fact that the writer of an inadequate text is alerted to the flaws not by a colleague or superior but by an "impartial", objective machine, by a tool. He can then decide for himself whether he wants to reword the sentences in question.

Such a program not only aids the translator by weeding out unintelligible passages but it also has a long-term pedagogic effect on the linguistic performance of the tech-writers, with the great advantage that the documentation becomes more useful for the persons it is intended for.

To some readers, this may seem a superfluous introduction to the topic of machine translation (MT), but it must be emphasized that without machine-readable texts, without word processing, without a terminology data base and without adequate postediting facilities, the best MT system would be useless in practical application.

MT systems are complements of the other machine aids to translation; they cannot replace them, and to jump ahead to another conclusion, they will never replace the translator. The human translator will always be called upon to improve on the output of the MT system. Therefore it should be understood that the term *MT* is only used for the sake of convenience: "machine-assisted human translation" is just too long.

MT systems can be applied to a limited scope of texts only, but not to "fiction" in the widest sense of the word, ranging from poetry to advertising and other texts in which the main content must be read between the lines,

be they political speeches or quarterly business reports. But why translate such things by machine anyway? They are usually short enough and the subject matter is usually interesting enough to challenge the human translator.

In spite of various Cassandra cries, machine translation has made some progress since the 1960's. Today it can be utilized in good conscience for several applications:

— for "quick-and-dirty" information gathering (e.g. to sort through large amounts of foreign-language texts in order to decide which documents deserve greater attention; to collect rough information for experts)
— for pre-translation of texts within limited subject areas ("pre"-translation implies that the machine-translated output will need to be postedited by a human translator).

But even this statement must be qualified. MT systems can be applied to such uses *if*

— the effort is justified (there is no need to trouble the mainframe in the computing center for a five page text without deadline)
— the quality of the output is high enough (if the posteditor spends as much time on correcting the pre-translated text as he would translating from scratch on his own, the system is likely to be in a terminal state)
— the cost is outweighed by the benefit (machine time is expensive, and a slow, CPU-intensive system on a large computer can easily become uneconomical)
— the time required is not excessive (a system which may produce translations of good quality but uses half an hour real-time per sentence is equally uneconomical)
— established procedures and structures are not greatly interfered with (if all other users of the computing center have to unload their jobs in order to make room for the giant MT system, enthusiasm for the system is likely to wilt).

In other words, a machine translation system has to deliver translations of high quality, be user-optimized, run quickly and cheaply on a small machine and be accepted by both management and translators. That is utopia.

At present there is no linguistic theory which would fully explain all utterances of even one single language; and there is not a hypothesis in sight that might prove successful in this century. Therefore one needs to compromise and try to approach a useful system by limiting lexicon and grammar to a specific type of text and a specific subject area. In such a way, a part of the highly complex disambiguation features and procedures which inflate and destabilize the system can be dispensed with.

According to the type of text, identical surface structures would need to be interpreted differently. In a prescriptive text, e.g. a maintenance manual, a German sentence like *Geräte warten* would be interpreted as a structure with the value of an imperative ("Service the devices"), in a descriptive text as an expression describing a state ("Devices are waiting").

By limiting the subject area, ambiguities in the lexicon can be reduced. The English word *pipe* for example can be interpreted as a musical instrument, as a tube or a smoking implement, depending on subject area of the text. If the

system lexicon is geared to just one subject area, there is less need for disambiguation procedures.

Unfortunately, many ambiguous cases remain to be resolved. As a matter of fact, the difficulty of translating a specific term correctly is usually inverse to the lenght of the sentence. It is not true that short sentences are necessarily easier to translate. Indeed, the less context that can be utilized for the interpretation of the string, the more problems arise. If in a German manual dealing with a data processing system there is the sentence *In solchen Fällen muss Band A neu eingelegt werden* ("In such cases tape A must be remounted") it is obvious that the *Band* in question can only be a magnetic tape and not "volume A" of the documentation.

If the MT system finds just *Band A* without further context from which to derive clues as to the meaning of the term, the question cannot be decided at the sentence level. In such cases one would probably proceed on statistical grounds, based on frequency distribution of the word's various meanings in the relevant texts — unless of course someone has arrived at a scheme to formalize the semantic structures of a whole text and incorporate them meaningfully in the analysis of the individual sentences. More utopia.

Siemens Language Services had long seen the need for operable MT systems. The goal was to have available in-house an MT system which would be able to translate monotonous texts (such as operating manuals for telephone systems) at such a quality level that the posteditor can greatly increase his usual output. The great advantage of using an MT system is that peak loads in the not so calm and even translation business can be managed. Contrary to all rumors, there is a dire shortage of highly qualified translators; there is no other choice but to try to upgrade the tools and working environment of the available qualified translators.

At the same time, translators are not being abused with nerve-wrecking assembly-line work for weeks at a stretch. If in an operating manual on one hundred pages, the same or similar sentences keep reoccurring two or three hundred times, the translator is so undertaxed intellectually that his interest in the work diminishes. And that again leads to errors and oversights. It is understood that such dull routines do not enhance the joy in one's profession. If the translator can limit his involvement in such texts to proofreading the output and making minor corrections here and there, he can devote more time to intellectually more challenging tasks.

After various experiences with MT systems and systems claiming to be MT systems, Language Services decided to utilize the many years of research expended in the USA and, in cooperation with the University of Texas, to develop a system which takes practical aspects of a translator's work into account. Research and development have been conducted jointly by programmers, linguists as well as translators. The MT system, known as METAL, is a language-independent system on the (modified) basis of Fillmore's case grammar; during the first phase, it is intended for translations from German to English.

It is modular in multiple ways. Programs and linguistic analysis procedures are strictly separate so that one can be modified without affecting the other.

Thus, for several types of texts, different grammars can be invoked without any change in the programs.

Furthermore, the grammar rules are linked to an index which permits the use of the whole of the grammatical analysis system or parts thereof, depending on application. If the system is to be used for quick-and-dirty information gathering, only the less complex syntax and semantics rules are called; the translated text will contain more errors than it would if the whole system were used, but the rough draft translation would be output much faster and cheaper.

Language-independence of a system is especially useful if often the same source texts have to be translated into several target languages at the same time. Since the analysis of the source language is by far the most costly step it would be uneconomical not to use it more than once. If 80% of the machine time has to be spent on the analysis of the German source text, a lot of time and cost can be saved if this same analysis intended for the translation into English can be used for the translation into e.g. Spanish and Arabic as well.

METAL is embedded in a text processing environment which automatically reconstitutes the original text format with all tables, graphs, etc. after the translation process. That is especially crucial in technical documentation. It saves the step of having to reinput everything and restructure the translated text.

A spelling corrector decreases the number of mistranslations that are due to errors in the source text. If the system is unable to analyze a sentence with a minimum level of certainty, it produces a terminology list as output. The posteditor is called upon to translate the sentence (instead of possibly having to sort through "garbage").

The quality of MT systems has often been measured in percentage points. The figures usually refer to the number of "correct" words or sentences. Such a scale is of course highly questionable. A correctness of 75% measured in words still implies that every fourth word needs to be changed in postediting and that perhaps not a single sentence of the text is correct. It does not give any indication as to the degree by which the translator is aided by the system.

During the experimental stages, the METAL German-English system has so far reached a correctness level of between 40 and 80% of "sentences"*.

Obviously, such a statement is not meaningful either. It is still no indicator as to the suitability of an MT system for practical productive use. For translators as well as administrators, the main criteria are really whether the system produces output of such quality that the translator derives a benefit from its use, whether it is cost-effective and easy to use. It is no secret that the present state of the development could be improved upon.

It must be stressed again : MT is only one of the many machine aids available to the translator, and he has to evaluate his own needs critically to be able to decide which tools would be the most useful in his particular situation.

---

* "Sentences" are all grammatically independent units; this includes headings which may consist of only a noun phrase, or table entries consisting of a single word. "Correct" means that no morphological, syntactic or semantic errors occur. Stylistic changes which a revisor might make but would not *have* to make are not considered.

Not every translator will be able to utilize an MT system, though for some applications it can be safely said that *legitimate* MT is here to stay. It will not replace the translator but it will probably change his job description.

The number of less qualified, general translators will decrease, that of the specialized highly qualified revisor will increase proportionately. Terminology and dictionary work will become more important. Dull routine work will be left to the machine whereas the more interesting things remain within the realm of the human translator.

All in all, the prospects for translators are not bad.