

WORKING WITH THE WEIDNER MACHINE-AIDED
TRANSLATION SYSTEM

Michael G. Hundt
Department of Translation
Mitel Corporation
Kanata, Ontario, Canada

I have been asked to speak to you today about the Weidner machine-aided translation system. To do this I shall first tell you a little something about the company I work for, Mitel, and why we decided to purchase a machine-aided translation system.

Mitel is a large manufacturer of telephone switching systems which are sold in Canada, the United States, Europe, the Caribbean, the South Pacific, the Far East and in South America. Being able to provide documentation in a number of different languages is therefore a key to our success worldwide. However, this is not as simple as it may appear on the surface. Because of changes and improvements to our products every six months or so, our documentation must be changed accordingly — some 4500 pages or approximately 2 million words per language every six months. Of course, not every word in the documentation is changed, but if a paragraph here and a paragraph there are changed, it can still result in a sizeable amount. Add to this the documentation for our new products and one can come up with a very large figure. The problems associated with this are numerous: the costs of hiring four or five or six translators per language are too high, using freelance people is expensive, and the problems of finding qualified people with telephony as a speciality and of terminology between different translators arise. Add the costs of retyping, outside typesetting and printing and translation becomes almost unaffordable for a company like Mitel. It must therefore be kept in-house.

For this reason we decided to purchase the Weidner system, which is designed for a mini-computer such as the Digital PDP 11/34. Because we have three languages on the system, we use a slightly larger computer, the PDP 11/70. We have hooked this up to a number of other machines we have at our disposal in order to help things move along quickly and cost-effectively. For example, we have at Mitel a Xerox 9700 laser printer to do our printing for us and an Addressograph Multigraph/ECRM digitizer, which scans illustrations and transfers them to the new document. These machines allow a document to be printed, including all illustrations, charts and tables, without the need for regular typesetting. A text which is entered into the computer is coded in such a way that it is formatted, the illustrations added at the correct place and the desired font or typeface used. All the coding is done when the English text is input — usually by someone in the publications department and not by the translator. All that is left for the translator to do is to recall the required English file at the desired time and do the translation. Since the text has already been formatted, the translated version will be printed in the same manner. It should be noted that the typesetting codes have no effect on the translation function of the computer. It should also be noted that all of this is done completely electronically — the translator does not need a written document unless he personally prefers to work with one.

All of which brings us to the Weidner system and how we do our translation at Mitel. As mentioned, we have three language pairs in our system, English-French, which has been running for about 1½ years; English-Spanish, which has been running for about 8 months; and English-German, which we received from Weidner only a very short time ago. The following functions are available to us in the system:

A. The Amender, where we can enter texts to be translated and revise translated texts. It is quite a powerful word processor.

B. Vocabulary Search is used to ask the computer which words in the source text are unfamiliar to it.

C. Deferred Vocabulary Search has the same function as Vocabulary Search, except that the operator instructs the computer when to start the function. This is valuable if the computer is being used heavily during the day and the text to be searched is very long. The operator can instruct the computer to do the vocabulary search at night when it will not interfere with or slow down the work of other users. Conversely, if there are other people using the system, the vocabulary search itself will take longer than it

does during off-hours and a lot of time will be wasted. It is therefore to everyone's advantage if this is done overnight.

D. Dictionary Update. From the translator's point of view, this is the key to the whole system. This is where the translator enters his vocabulary. This ensures that his terminology remains consistent throughout his translations because once a word is entered into the computer's memory, it will always be translated in the same way. This is also where the computer is "taught" the grammar of the language. The translator tells the computer the word's gender and plural if it is a noun, its inflection if a verb, its agreement if adjective or adverb. This must be done precisely if the translation is to come out properly in the end.

E. Listing Utility. This is used to obtain a print-out of one's dictionary. The dictionary can now be carefully studied to see if something has been entered incorrectly.

F. Translate. This is self-explanatory and is probably the easiest function for the translator himself and the most complicated for the computer.

G. Deferred Translation has the same function as Translate, but like Deferred Vocabulary Search, it can be done during off-hours so as not to interfere with other computer users.

H. Synonym Update. This enters synonyms for words in the target language into the computer's memory. We will look at this more carefully later.

I. Translation Process Monitor. This option allows the translator to see what the computer is working on. Active, future as well as finished jobs can be displayed. When an active job is finished, it disappears from the screen and the terminal "beeps" to let the translator know that the job is finished. The language pairs as well as the file names are displayed.

J. Manager Utilities is a command which the translator doesn't need to use. We leave all of the system problems to the system manager.

At Mitel the translator gets the English text in a pre-publication form. The basic text is there and needs only a few revisions from the technical writers and engineers. At this point, about six weeks before the final document is printed, the translator runs the text through the vocabulary search procedure. He finds out which words are unfamiliar to the computer and enters them into the dictionary. The procedure is quite simple. From the list of options, one chooses "B", Vocabulary Search. The computer asks you which file to search through. After the search has been completed, the results can be displayed on the terminal screen or printed out. Note that the line in which the word is located is also displayed. From this, the context of the word can be derived. The words can also be listed alphabetically or depending on their frequency in the text; however, we have found that the context option is more useful.

Now the translator enters the unknown words into the dictionary. This is the most complicated and most important part of the Weidner system. The information given the computer must be exact for the translation to come out properly in the end. Let us look at a few examples.

For an example of a noun entry, let us look at the Spanish word for "house", which is "casa". The first question the computer asks is, "Is this word a homograph?" that is, does this word have more than one translation? The answer in this case is yes since the English can also be translated by "albergar", which is the verb "to house". For our purposes, let us look only at "casa". We enter the word in the appropriate space. Following this, one is asked the following questions:

— Part of Speech (Verb, Noun, Adjective, Adverb, etc.). We have just received a new version of the Weidner system. In the old version, we could only enter verbs, nouns, adjectives and adverbs. Now we can also enter prepositions, conjunctions and so on. However, the old four are still the most important and I shall concentrate on these.

— Gender (Masculine, Feminine, Either).

— Number. This is a question relating to special nouns, which are always plural in the original but always singular in the target language or vice-versa.

— Agency (Human, Group, Body Part, Animal, Inanimate, Concrete, Abstract). Weidner has yet to

satisfactorily explain these categories and the reasons for them to us, since most of them have no direct effect on the translation.

- Is this a proper noun? (No or Yes.)
- Is this a noun of time?
- Is this a noun of place?
- Does this translation present the "ING" form of the source word? "ING" nouns present a special problem because they can be interpreted by the computer as a verbal form. Nouns such as "building" must therefore be confirmed as such in this step.

Once these questions have been answered, a check step or a chance to make changes is included should one decide that there is something needing correction.

After this, the computer goes on to the next step. Here the translator is asked how the translation corresponds to the source language, that is, in what way is the translation influenced by the source word. In this case the correspondence is one to one, that is, the target word inflects according to how the source word inflects. This is more important with idioms, which we will look at later.

On the next page the computer asks how the word inflects. It gives examples and asks if the words inflect like the examples. Sometimes, three or four examples are given and the translator must choose the one which inflects most like his word. We shall see an example of this later.

After this the computer goes back to the first page and enters the word into the dictionary.

For an example of a verb let us look at the German translation of the word "bite" which is "beißen". I have purposely chosen this verb because it is irregular. A very specific question is asked for irregular verbs, as we shall see in a moment.

At the beginning we go through the same procedure as with a noun. Is this word a homograph? We answer yes for two reasons. First, the word "bite" can also be a noun. Secondly, we enter two forms for verbs. The first is the simple translation, "beißen"; the second is a variation on the translation. We enter "der beißen", which will become the translation for "the dog biting the man". The translation will appear as a relative clause, "der Hund, der den Mann beißt", introduced by the relative pronoun "der". For the purpose of this demonstration we will leave out the noun form.

After we enter the translations, the computer again asks us a number of questions about the first translation, "beißen", including part of speech, which we saw in the first example, and the following questions designed specifically for verbs:

- Agency (None, Direction, Location).
- Can the verb be transitive? (Yes or No.)
- What type of direct object does the verb take? (Accusative, Dative or Genitive.)
- Can the verb take an indirect object? (Yes or No.)
- Is the verb reflexive?
- Is the past participle formed with "haben" or "sein"?
- Character length of the separable prefix. Here one enters the number 0 through 15. In some German verbs, such as "hinterherlaufen", which means "to run after" or "to chase", a prefix exists, which is separated from the main body of the verb in certain tenses. In this case, the 9-character prefix, "hinterher" is separated from the main part of the verb, "laufen", and placed at the end of the sentence.
- Does the verb include "ING" adjectives before a noun? In this case, the answer is yes, because the adjective, "biting", does occur before a noun: "the biting dog".

— Does the verb include "ING" adjectives after a noun?

— Does this form include "ED" adjectives? In this case, the answer is no, but an example is the verb "to desire" where the adjective is "desired".

Again the chance to make any necessary changes is given before we go to the next step.

Here the computer asks for a cross reference, or the correspondence, between the two languages, which we will come back to.

In the next step, three questions are asked.

— Is the verb weak or strong, that is, regular or irregular? If the answer is weak, or regular, then the next two questions are left out. However, if the answer is strong, or irregular, then the following two questions are asked.

— Does the verb have an inseparable prefix, as is the case with some German irregular verbs?

— Without a prefix, this verb inflects most like which of the following? Forty-two examples are given and the translator must decide which, if any, of the following verbs inflects like his. In this case it is No. 5, "leiden". In both verbs the "ei" in the infinitive changes to "i" in the imperfect and perfect: "leiden-litt-gelitten" and "beißen-biß-gebissen".

After this, the computer goes to the next form of the verb, "der beißen", which is entered in exactly the same manner as the simple verb translation. Most of the questions which have just been answered are left out, so that not too much duplication of answers and time-wasting takes place.

For an example of an adjective let us take a look at the French translation for the word "happy", which is "heureux". Again the question concerning homographs is asked, after which the translation is entered.

After entering the part of speech into the computer, the following questions are put to the translator.

— Does this adjective reorder? That is, if the adjective precedes the noun in the source language, does it follow the noun in the target language? "The happy man" becomes "L'homme heureux". The answer is yes.

— If the adjective precedes an infinitive, insert: (Nothing, A, or De).

— Is this adjective always plural? An example of this is "plusieurs" (many).

Again the translator is allowed to make any changes at this point. From here he again goes to the next step, which is cross-referencing. After this he is asked one question concerning the declension of the adjective:

— This adjective declines most like: (Doesn't Decline, Heureux, Vieux, Faux, Doux, Index). The computer wants to know how the adjective inflects in the feminine and plural. "Heureux" becomes "heureuse" in the feminine, but "vieux" becomes "vieille" and "faux" becomes "fausse" and so on. In our case, "heureux" declines like "heureux", so it is not a difficult choice at all. Different examples are given for adjectives not ending in "x".

After this has been answered, the translation is written into the dictionary.

Adverbs are entered in the same way and are very simple for the computer to handle.

Let us now have a look at an example of an idiom. An idiom in this case is any phrase of two words or more. Since I am more familiar with German, we will use the example, "to rain cats and dogs", in German: "in Strömen regnen".

First of all, all individual words in the English idiom must be separately entered into the dictionary, whether this individual meaning has anything to do with the idiom or not.

Then we proceed as with any other entry. In this case, the idiom as a whole is a verb. Therefore, the same

questions as for a verb are asked by the computer.

The next step is an important step for idioms. This is the cross-referencing, which was mentioned a number of times earlier. Here we ask ourselves, which word, if any, in the translation corresponds to the source language. Looking under "CR-REF", we decide that "in" does not correspond to anything, so we enter a "0". We also decide that "Strömen" does not correspond to anything and we enter another "0". However, "regnen" corresponds to "rain", so we enter a "1", which is the number next to "rain", in the appropriate space. Now "regnen" will do in German what "rain" does in English; that is, if "rain" is plural in English, then "regnen" will be plural in German. After this, the computer asks us what type of word each individual word in the translated idiom is. Beside "in" we enter "other" because it is not a noun, verb, adjective or adverb. However, beside "Strömen" we also enter "other". You may now say to yourself that this is a noun. That is true, but in the idiom it never changes; it doesn't inflect in any way whatsoever. If we told the computer that it was a noun, it would try to inflect it and probably make a mistake in the translation. Beside "regnen", however, we must enter "verb".

On the bottom left of the screen the computer then asks us for the key word in the idiom. It does not want the most important word in the idiom, but the word which occurs less frequently than the others in the computer's dictionary. This helps to speed up the computer's search for the idiom during translation. It has no direct effect on the translation itself and any word can be selected without causing any translation problems.

Below this the computer asks the translator which source word governs the agreement with the target language. In this case it is no more than a confirmation of the cross reference above. We enter the number "1" because the source word, "rain", governs the agreement. However, in an idiom where there is absolutely no relationship between any of the words in the two languages, that is, where the cross references are all "0", one of the words in the source still influences the target language as to number, tense, etc. This is taken care of here.

After this the idiom is written into the dictionary.

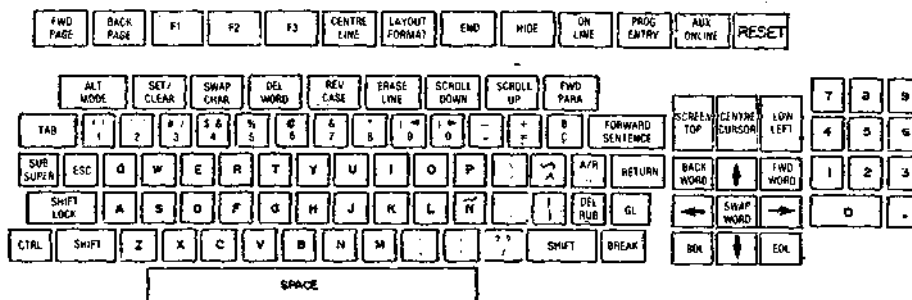
As I'm sure you can understand, the dictionary entries must be done very carefully and accurately in order to receive the best possible computer translation. At Mitel we have found that idioms are an extremely important part of the translation procedure. You see, machine assisted translation is for the most part word replacement and we all know the problems that idiomatic expressions cause, even in regular translation. Therefore, it is important to enter as many idioms into the dictionary as possible. There is only one thing to watch out for and that is an idiomatic expression which also has a meaning if the words are taken separately. An example is the phrase, "to ring a bell". It literally means "ring a bell", but the idiom means "to arouse a response", as in "that name rings a bell". If the phrase has been entered as an idiom, the first meaning will never be translated. You, the translator, must therefore decide which phrase will occur most often in your work and make the entry accordingly. That is the nice thing about the Weidner dictionary. If there is a translation that will almost certainly never come up in one's field, one can delete it from the existing dictionary or leave it out altogether. Therefore, at Mitel we would not enter the idiom "to ring a bell" because we will never need to have that phrase translated, but in the telephone business, bells literally ring all the time.

To see which words or phrases are already in the dictionary, the Listing Utilities option is selected. It arranges all the words in the dictionary alphabetically and prints them out so that they can be studied carefully. A number of words which we do not need for our kind of technical translation had been originally entered into the dictionary by Weidner and we decided to delete these or to change their translation to suit our needs.

Now would be a good time to look at the Synonym Update option. This allows the translator to enter synonyms for words in the target language into the computer. For an example we could look at one of the French words for "car" — "auto". The computer asks us for the word type, which in this case is a noun. We enter three synonyms: voiture, véhicule and cabriolet. These are now entered into the computer's memory. When we look at revising the computer translation I shall show you how to put this to use.

The translate function itself is very simple. One simply chooses a language pair, selects the translate option and enters the name of the file to be translated. Using the rules in the Weidner program and the words in its dictionary, the computer translates the file into the desired language. The progress of the translation can be followed using the Translation Process Monitor discussed earlier.

To revise the translation, the Amender option is chosen. The text is displayed in both languages on the terminal screen, one above the other. It is very easy to see what the translation is and to decide what it should be. The terminal now functions as a word processor. The keys on the top two rows, the twelve keys between the typewriter keyboard and the numbers on the right as well as the keys on the left and right sides of the typewriter keyboard allow one to manipulate the text in any number of different ways. Words can be deleted and moved around. Entire phrases can be replaced and lines deleted. Here is also where our synonym entries from before come into play. Remember that we entered synonyms for the word "auto" in French. To display the synonyms for this word on the screen, one places the cursor at the beginning of the desired word, in this case "auto". The "ESC" button on the left side is pressed and the part of speech chosen: the number 1 for verbs, 2 for nouns, 3 for adjectives and 4 for adverbs. The synonyms will now appear at the bottom of the screen. The translator makes his selection by choosing the appropriate number and the original word is automatically replaced. In this way the translator can rapidly polish his text until he is happy with the translation.



All that is left to do now is to remove the source language from the file by choosing one of the Amender functions called "Split Translated Dual-Language File". After this the translated text is ready for printing by the publications department of the company.

It is really not worthwhile discussing the various problems of the system in great depth because in the first place they are much too numerous and secondly, problems will always exist. After one is solved, another is sure to appear. Take the sentence, "The rabbit stole the carrots from the old woman's garden". This should be translated by "Le lapin a volé les carottes du jardin de la vieille femme." and not by "Le lapin a volé les carottes du vieux jardin de la femme." The system is not created with an inherent understanding of language, as we are, so a result such as this should not be too much of a surprise. We know from the context that "old" modifies "woman" and not "garden", but the computer knows nothing of context. In another case we receive the translation: "Ce brochure est conçu vous pour montrer opérer commensément la console" for: "This booklet is designed to show the operation of the console." This is certainly not very good French, but it is not all that bad either. It is now just a matter of using the word processor function of the terminal to change "ce" to "cette", to reverse the words "vous pour", to correct the word "commensément" and to then reverse "opérer comment". This is about as good a translation as can be expected of the Weidner system and probably of any other machine-aided translation system. What it does is give the translator a text, more or less grammatically correct, with the terminology entered into it by the translator himself, which can then be revised. Depending on the speed of the computer, it does this extremely quickly. What it boils down to is that the system is capable of producing a draft, which a translator can then upgrade faster than if he had done the whole thing from scratch. It saves time!

Finally, what is required of the user? First and foremost, patience and an understanding of what the machine can and cannot do. It cannot translate a word which is not in the dictionary, for example. The system comes with a basic vocabulary entered into it by Weidner, but any specialized terminology must be entered by the end user. Any word not in the dictionary is left in the original language and the translator can translate it himself when revising the text. The first few months of working with the system are almost exclusively spent entering words into the dictionary. If one is running a vocabulary search, which in the end results in a list of 100 or so words, it can be very discouraging. A whole day can be spent entering these words into the dictionary — not a lot of fun. But it certainly is not time wasted because these words will remain in the dictionary forever and one will never have to enter them again.

After a few months, most of a translator's "personal" terminology should be entered. Naturally, there will always be some new words to enter, but this can then be done very quickly. As well, the computer will always make mistakes, including some that it probably should not. For example, the computer still has trouble distinguishing a noun from a verb from an adjective if the words are identical in English. The word "light", for example, has been causing us a lot of problems. The phrase "the light box" is sometimes translated by "la lumière boîte" or by "la allumer boîte" instead of by "la boîte légère". The translator must accept that fact and be willing to work with the computer on those terms. The system is not perfect; it is merely a tool, an aid to help the translator do his job more quickly.

Secondly, the translator must be familiar with the operation of the word processor functions; however, this is very easy to learn.

Last, but certainly not least, the translator must be competent. This system does not and will never replace him. It is merely a tool for him to use like a pen, a typewriter or a dictionary. It gives him certain freedoms, such as translating at night at a very high rate. But in the end it still depends on him to make the system work.

In an article in the February 8, 1981 edition of "Temps Réel", a department of IBM, the DTAF, or document translation assistance facility, is referred to. They claim to have discovered that a regular translator spends only 20% of his time on actual translation. Terminology research takes up 40% and the remaining 40% of his time is spent typing, inserting illustrations and so on.

After we at Mitel bought the Weidner system we spent about 80% of our time doing dictionary entries and terminology research and about 20% of the time doing translation. Now that the majority of our technical terminology has been entered, we estimate that we spend only about 5% of our time doing dictionary entries and 95% of our time doing revisions. Much of the credit for that is due to our typesetting system. We spend no time inserting illustrations for example. However, even without this capability it is our opinion that the Weidner system would save a large amount of time.

The conclusion that can be drawn from all of this is that the Weidner system indeed works as an aid to translation. At Mitel we have found that, given proper dictionary entry and sufficient know-how on the word-processing end of things, 800 words an hour as a final figure is not unrealistic. If, as we do at Mitel, one combines the system with other devices such as our Xerox 9700 laser printer, it is a very worthwhile investment.

Personally, I feel that there is still a great deal of work to be done. Nevertheless, I am very excited to be working on this system. The Weidner system is still very much in the development stages and Mitel is one of the first companies to have such a system. We have agreed to let the Weidner people know about some of the shortcomings of their system, and when we do, they usually come up with a solution in a very short time. So the system is steadily improving. We receive an updated version of the system about every three or four months and can see the improvement each time. We at Mitel are convinced that this is the correct path to take in the future.