12 Software Review

"French Assistant" by Microtac

The Product

"French Assistant" (Version 5) is produced by Microtac Software, 4655 Cass Street, Suite 214, San Diego, CA 92109, U.S.A. Price $99.00. (We believe it might have been on sale at Dixons in the UK for £39.95)

Requires IBM PC/XT/AT or 100% compatible with minimum 640K RAM, DOS 2.1 or above and about 2.8MB of space on a hard disk.

This software is one of the "Assistant" series which includes German, Italian and Spanish.

In the words of the Introduction in the User's Guide;

> "French Assistant is a powerful translation, reference, and writing package. It provides bi-directional, sentence-by-sentence translation to help you translate your documents into French and translate French documents into English. The program generally produces understandable, but not perfect, translations. When a perfect translation is required, please use French Assistant in conjunction with a human translator.
>
> French Assistant is also a great help when studying or writing in French. You can pop up translations for individual words, refer to grammar help, conjugate French verbs, and easily add foreign characters. You can create documents in French Assistant's integrated text editor or you can import text files from other applications."

In this review we are mainly interested in the translation facilities referred to in the first paragraph but will comment on the other facilities in passing.

Introduction

This is not a review in the conventional sense when you end with a recommendation to buy, or not to buy. It's more of a technical inspection, and as it turned out, a rather limited inspection.

When I first offered to review this package I did not realise how difficult it might be, not just preparing the input and testing it, but in deciding what to do in the first place. We wanted to be fair, we wanted to be helpful to all parties, and we wanted to produce an interesting report for you illustrating some of the problems and pitfalls of MT.

How then was I to review this package?

First of all, since we do not have access to the source code nor any detailed system description apart from what the User's Guide tells us, we are obliged to "black box" test, that is, feed it with sentences and examine the results. Then to find out why certain things happened it might be necessary to modify the original sentences or even construct new ones, and try them again to see if the program's internal rules of behaviour could be deduced.

The next question was what sentences should we start with?

There are, I suppose, two main approaches, namely;

1) Just select random text; like the User's Guide itself, perhaps.

 The problem with this is that it could be rather inefficient in that you will probably get a lot of text of the same type, and might not get examples of some quite common constructions for a long time, and the vocabulary might be unsuitable for the system's dictionaries.

 The size of the vocabulary is of course very important because it is well known that the problems of disambiguation rise exponentially with the size of the vocabulary. However, at this stage we were more interested in seeing how well the system translated per se, rather than testing the coverage of the dictionaries.

2) Construct sentences from a known vocabulary, but being free to choose words we could really test the system with each sentence.

 This is technically more efficient but does take longer since so much thought has to go into making up sentences to provide suitable syntactic and semantic problems in an ordered manner.

 However, it is well known that you can always find unusual sentences which any particular MT system cannot deal with and the creation of sentences can be very unfair.

In the end time ran out and all we could do was a superficial examination based on the sample sentences provided in the package. At least we cannot be accused of maliciously selecting difficult material!

The Package Facilities

The package provides the following basic facilities;

1) An editor to enable the user to enter text in English or French, including accents, of course.

 This also includes facilities for exchanging files with a number of popular word-processors.

2) Easy access to reference material during editing;

    a) Dictionary in either direction
   The package claims 100 000 terms.

    b) Verb Conjugations
   This shows fully listed forms for 15 tenses, including the imperative, and also provides guidance for the use of each tense.

    c) Grammatical information
   This provides short notes on 42 topics, such as Adjectives, Direct Objects, Gender, Possessive Pronouns, Questions, etc..

3) Translation in two modes;

    a) Semi-automatic
   In this mode the system appears to translate what it can when there is no choice in the dictionary, but stops to allow the user to choose from a dictionary entry when there is a choice.

    b) Automatic
   In this mode the system doesn't stop, but takes the first choice where there is a choice.

4) Display syntactic structure of translated sentences.

This displays detailed information about the structure of the translated sentence including information about which words 'agree' with each other. This feature might be a useful pointer to the way the system works but I did not have time to make use of it.

5) Printing facilities

These include the ability to print, with all necessary accents, in a number of different formats on a wide variety of printers including the well known laser printers.

6) Dictionary building and maintenance

The advanced user can add new dictionary entries, and alter and add to existing entries.

Words may be given attributes (feature markers) to assist correct selection where there is a choice, and advanced users can allocate their own attributes.

7) Automatic installation

Installation was straightforward.

8) Use

The system was menu driven and easy to use.

9) User Guide

For a software manual this is quite good, though as a seasoned computer user I have already added quite a few entries to the Index, such as Auxiliary, Defaults, Dictionary - restore entry, Lexical phrase, Lexicon, Model codes, Parsing, Rule code, Slot Phrase, and Syntactical display.

Translation

As mentioned above I only found time to run the sample sentences provided plus a few experimental sentences derived from them.

First of all I tried translation in automatic mode. In this mode when the system is faced with a word with more than one meaning shown in the dictionary, it will select one, usually the first in the entry, though sometimes it seemed to choose the first noun or the first verb when they weren't at the beginning of the entry.

There are, however, a number of mechanisms which can be used in dictionary entries for automatic disambiguation;

a) A limited set of Attributes (feature markers) supplied by the system, such as, +Animate, +Place, +Proper (name) etc. for nouns, or +Quantifier, +Qualifier etc. for adjectives, etc..

b) User allocatable attributes which can be used to develop and/or improve existing dictionary entries, e.g. MUSIC could be attributed to musical instruments, e.g. *violin*, to enable the dictionary entry for *play* to distinguish between *jouer à* and *jouer de*.

c) Lexical phrase matching e.g. in the entry for *ice*, "*ice cream*: glace* (f)*" would match *ice creams* to produce *glaces*.

d) Node phrase matching e.g. in the entry for *tomorrow*, "*the day after tomorrow: après-demain*" would match the English adverbial phrase if present, to produce the equivalent French phrase.

e) Thread phrase matching. These are mainly used to deal with idiomatic expressions, e.g. in the entry for *hungry*, a special rule called *Be-Have* can be specified with *"be hungry: avoir faim"* which would match *We are hungry* to produce *Nous avons faim.*

f) Pattern matching. These are used to deal with similar but more complex situations, e.g. in the entry for *following*, *"the ˜ <Noun+Time>: le{Art+3} <3 Noun+Time> suivant{Adj+3}"* would match *the following year* to produce *l'année suivante.*

In the source pattern the symbol ˜ stands for the head word, *following*, and <Noun+Time> stands for any noun with the +Time attribute, that is, a period of time.

In the target pattern *le* stands for any definite article, the {Art+3} says that this word is an article and it has to agree in number and tense with the third word in the source pattern, i.e the noun. The <3 Noun+Time> stands for the noun in the source pattern, and the *suivant{Adj+3}* says that the next target word is *suivant* and it is an adjective which must agree in number and gender with the noun.

As the automatic (MT) mode gave some odd results I then tried the normal translation mode to see if I could improve on them. In this mode, which we could call MAT (Machine Assisted Translation), the system stops when it finds a word which has more than one translation shown in its dictionary entry, for the user to move the cursor to the most appropriate meaning available shown in the display and then to press ENTER, when the system takes over again.

On completing the translation of each sentence in MAT mode, the user then has the option of amending the result quite easily, using normal word-processing facilities.

In the following I will compare both the MT and the MAT versions of some of the sample sentences which showed problems.

It was difficult, and probably would be misleading, to count the number of errors in the translation of the sample sentences, first of all because some were repeated in a group of sentences with a similar verb or construction and some of those were repeated in a group of questions, and because some errors are more significant than others. I will therefore only refer to representative problems.

The correct choice of prepositions is the most obvious difficulty, e.g. *My brother used to study in the library* generated *dans la bibliothèque* in MT, but this could be corrected in MAT by choosing the preposition indicated by use with *"proper names of places"*, i.e. *à*.

Similarly *had they visited Paris in 1987?* produced *... dans 1987* in MT but this could be corrected in MAT to *en 1987* by choosing the appropriate preposition used *"before dates and feminine countries"*.

A number of sentences demonstrated the correct positioning of object pronouns but *will she give it to him?* produced *Est-ce qu'elle donnera le à lui?* in spite of the fact that *She will give it to him* produced *Elle le lui donnera.* MAT didn't help here because the system doesn't give the user any choice.

... before you make a final decision produced *... auparavant vous faites une décision dernière* in MT. As *prendre* was not listed under *make* and *avant de* was not available under *before* and *finale* was not under *final* the best I could achieve in MAT was *... avant vous faites une décision définitive.*

Whilst demonstrating that the system could distinguish between two meanings of *left* the system produced *J'ai laissé la réception tôt* for *I left the party early.* Although it had recognised that a verb was required it chose the first verb in the entry for *left.* In MAT

the user could choose between *laisser, quitter, partir, sortir,* and *abandonner. Quitter* was shown as meaning *quit, exit from.*

In the same sentence the words *another meeting* were automatically translated to *un autre réunion* but *réunion* is actually feminine. MAT does give the user the choice between *un autre* and *une autre.*

A more difficult problem was shown by *He was sitting under a tree* which produced *Il s'asseyait sous un arbre. He was sitting* indicates a state and therefore the verb *s'assoir* which indicates an action cannot be used in French. Instead *être assis* in the appropriate tense should be used here, *Il était assis ...* .

The jewels around her neck were inexpensive produced *Les bijoux autour son cou était bon marché* in MT because *around* matched the first word in its dictionary entry which was the adverb. In MAT the user could choose the preposition *autour de* which also corrected the verb to *étaient*, presumably by allowing *bijoux* to become the sentence subject.

John's heart beats rapidly when ... produced *Les battements du coeur de John rapidement quand ...* in MT because *beats* matched the noun *beat* in the dictionary first and when two nouns are found together the system assumes the form *second-noun of first-noun.* This was correctable in MAT when the verb form for *beats* could be selected.

Similarly in French to English, *Un garçon intelligent lit un livre* became *A boy intelligent bed a book* in MT mode because for some reason *lit* matched the noun before the verb, but this was correctable in MAT.

Donnez-moi le livre became *Give I deliver him* because the verb *livrer* was selected before the noun *livre* in MT, though the noun form meaning *book* was written first in the entry. The system seems to prefer the choice of a verb if the word follows a pronoun of any sort. In this case the system didn't recognise *le* as a potential article.

Je n'ai pas d'argent became *I don't have of money* in MT. In MAT I got *I don't have some money* by choosing the preposition *some* instead of *of* but the system ignored the need for the negative transformation of *some* into *any, I do not have any money.*

In *Les étudiants n'étudieront jamais leur leçon* the system did not seem to recognise that *leur* was a possessive adjective modifying *leçon*, and chose the pronoun form instead to produce *The students will never study them lesson* in MT. Again, this was correctable in MAT.

John est plus vieux que Mary generated *John is plus old than Mary* instead of *... older than Mary* although the system could produce *plus vieux* from *older.* This was not correctable in MAT.

Cet étudiant est le plus intelligent de la classe produced *This student is most intelligent of the class* in MT, in spite of the fact that the missing *the* in front of *most* was printed in two other similar sentences. This was also not correctable in MAT.

Le propriétaire se débarrassait de la mauvaise voiture gave *The owner got rid of the bad car* in MT. In this case *... was getting rid of ...* would have been better. The system seems to translate the French imperfect into the English present perfect, though not the other way round. This sort of problem cannot be corrected in MAT because the user isn't given a choice of tenses.

Conclusion

The above catalogue of problems unfortunately hides the fact that most of the other sample sentences were correctly translated, but we hope this brief survey has given you a flavour of

some of the difficulties that can be found in the practice of MT.

I think it would be fair to say that French Assistant is well named and should be used as such; as an assistant in the process of reading and writing French, by someone who knows some French already.

All this also goes to show that the evaluation of MT systems is more difficult than it first appears and we look forward to hearing more about it from the professionals at the CL/MT Group from Essex University in May.

I would like to thank Ian Fantom and Monique L'Huillier for their invaluable contributions to this 'review'.

J.D.Wigg