

# TRANSIT

## The ideal working environment for translators

As part of our series on translation memory-based translation tools (Language International, 4/1, 1992), this article introduces TRANSIT, 'The ideal working environment for translators', released in May this year by the Swiss-based software firm STAR AG (Software - Translation - Artwork - Recording) who also have three offices in Germany.

TRANSIT is a PC-based, multilingual translation software package which incorporates a word-processor, dictionary databases and referencing systems. It requires a dongle\*. The Interface languages are German, English, French and Spanish, but TRANSIT can be used for any language pairs supported by the extended ASCII character set (for instance, Hungarian). The user interface is based on window techniques with pull-down menus, activated by using the mouse or key combinations.

As with similar applications, TRANSIT is a tool: it does not generate the translation, it assists the translator in creating it. All the solutions offered are directly taken from previous translations or from the chosen dictionary databases.

The idea behind the system is that there are certain aspects of the translator's work which can be done more efficiently by the computer: cross-referencing between comparable texts (different versions of a manual, for instance); looking up words in dictionaries; formatting.

A source text in either ASCII, RTF or INTERLEAF format is imported into TRANSIT from diskette whereby formatting instructions are filtered out and replaced for the course of the translation by TRANSIT-internal tags. These can be protected in order to keep the prescribed format without having to format the text again. A WordPerfect filter is also planned for the future, possibly linked to the existing RTF filter.

Secondly, the imported text is divided into segments, according to user-defined rules. Although formatting tags are regarded as segments and contained in the numbering, the segments of interest to the translator would usually be sentences. The segments are numbered automatically, and the numbers in the target language text must remain exactly as and where they are in the source text.

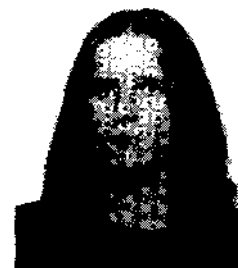
When the translation is completed, the text is exported and the internal tags are replaced again by the original ones. Thus the translator can concentrate fully on the translation itself.

Work with TRANSIT is project-based: within a project, the translator selects

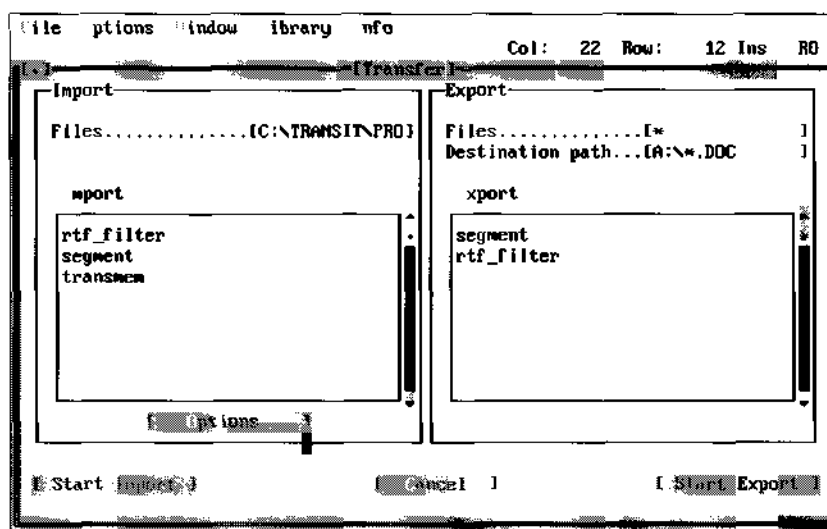
- a) the interface and text languages
- b) the source text(s)
- c) the dictionaries
- d) the rules according to which texts are segmented, filtered and matched against texts in the reference directory.

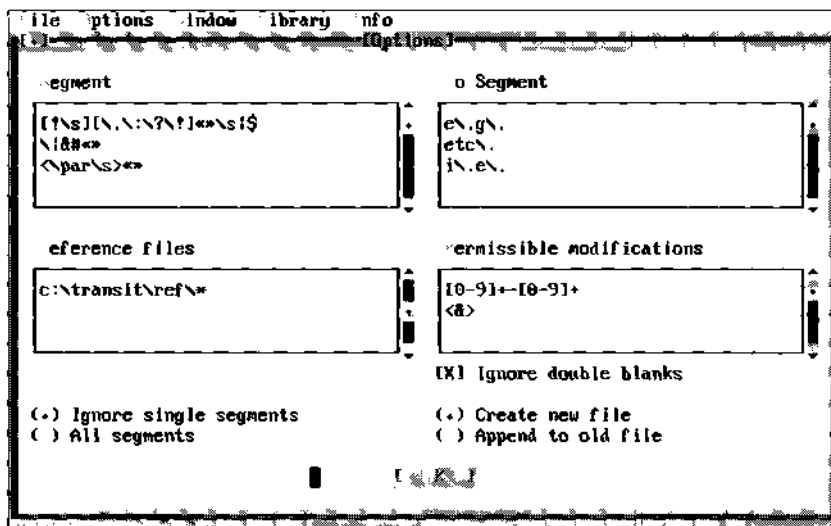
In addition, the user interface (the colours on the screen, for instance) can be redefined for each project. The advantage of the folder or project concept is that it saves time: several documents can be imported in batch without having to go through the process for each one, words will not be looked up in every dictionary available, but only the chosen ones, and different segmentation requirements for different texts/projects can be observed. Should a translator decide to use a newly received dictionary, this can just be added. It will be included in future sessions unless it is removed again.

The most important feature of using a translation memory-based program in comparison with normal translation is the possibility of automatic cross-referencing with previous translations: subject-re-



by  
**Sabine Bell**  
German translator  
at Praetorius  
Limited





lated previous translations (source and target language texts, segmented according to the rules to be applied to the new translation) are copied into a reference directory, and while a text is imported into TRANSIT, it is compared to these previously translated originals. If the program discovers a matching segment, it takes the translation for this from the target language text in the reference directory, and automatically inserts it into the current translation. These segments are marked with a tick.

As part of the import, the translator can choose to have an 'extract' file created, containing only the segments which have not been translated. As an extra option, any duplicate segments within this range can be made to appear only once, thus minimising the translation time. When this extract file has been translated, it is added to the files in the reference directory, and the complete source text is imported again with the translation memory option activated, so that it will be translated automatically and fully.

There is also the possibility of finding segments which contain a change, if this change is a model number, for instance. This can be very useful in the translation of manuals, for which the program was designed: the translator defines 'Permissible Modifications' which will not prevent the matching of otherwise identical segments. The definition is in the form of 'Regular Expressions' defined for each project. 'Regular expressions' can also be used in searches or search/replace functions when editing the text.

It is important to note that whilst these expressions are very useful in making the program flexible according to the project's requirements, they require technical/logical thinking to a degree which cannot be expected from every translator. This suggests that there should be a project manager who prepares the import stage for the translators. On large projects, it would also be a good idea to have somebody main-

tain the databases centrally and choose the reference materials. TRANSIT can be used within a network, which would be ideal under such circumstances.

The matching process in TRANSIT is based solely on character codes (pattern matching) and is not morphological, as in other comparable translation memory-based applications. The disadvantage is that the program will not recognise different forms of one verb, for instance, but the advantage of being able to use TRANSIT on such a wide range of languages is equally significant.

The process of automatic translation has two implications:

- a) The translator must decide beforehand which translation s/he would like to use for this purpose, as s/he is not offered a choice when it comes to replacing matching segments: this function, the translation memory, can be deselected, but if it is activated, it processes the texts in batch, without further interaction by the user.
- b) As the automatic translation is a batch process linked to the import of a text, recurring segments within a text which had not been translated previously are not registered, so in order to achieve consistency here, the translator relies on her/his memory. To avoid this problem in texts of high volume, the partly translated text could be copied into the reference directory, and the original source text could be re-imported. The translation of a line which has been translated and is repeated throughout the text would thereby be inserted in the appropriate places.

A question which often arises when the installation of a translation tool like TRANSIT is considered is: if parts of an ongoing translation project have already been completed, will I be able to use these for reference purposes within the new application?

In the case of TRANSIT, the answer is positive:

- Previous translations (in ASCII, RTF or INTERLEAF format) can be used if they are suitable, that is, if the source and target text can be split up into matching segments. For this purpose, there must be the same number of sentences, and the formatting must be the same.

Both texts are filtered into TRANSIT through the import process and then copied into the reference directory. The translator must check that they contain exactly the same number of segments - otherwise they will not work.

- Dictionaries can be imported if they can be dumped as an ASCII file and edited to obtain the required number of lines per entry etc. I found this process rather tedious, but it was well worth the effort. Alternatively, one could recreate the dictionary within TRANSIT, but this would be even more time-consuming.

Once a dictionary is imported, it can be added to and

edited in other ways as any other TRANSIT dictionary.

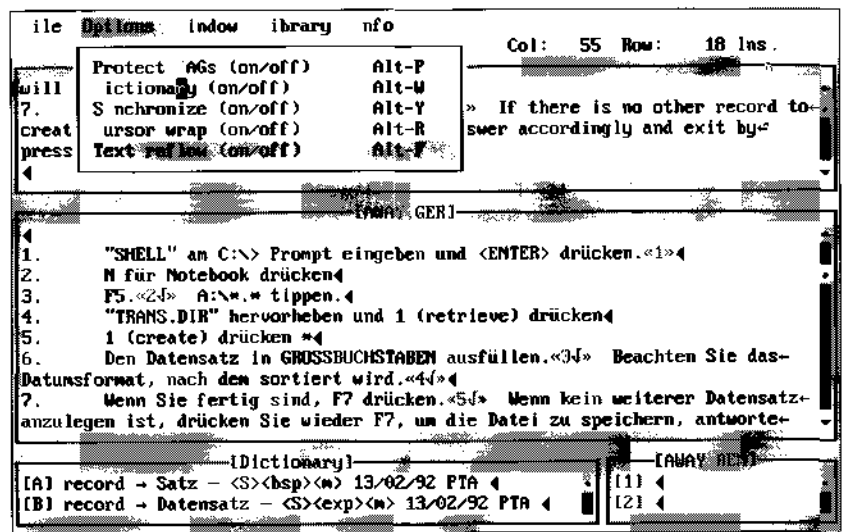
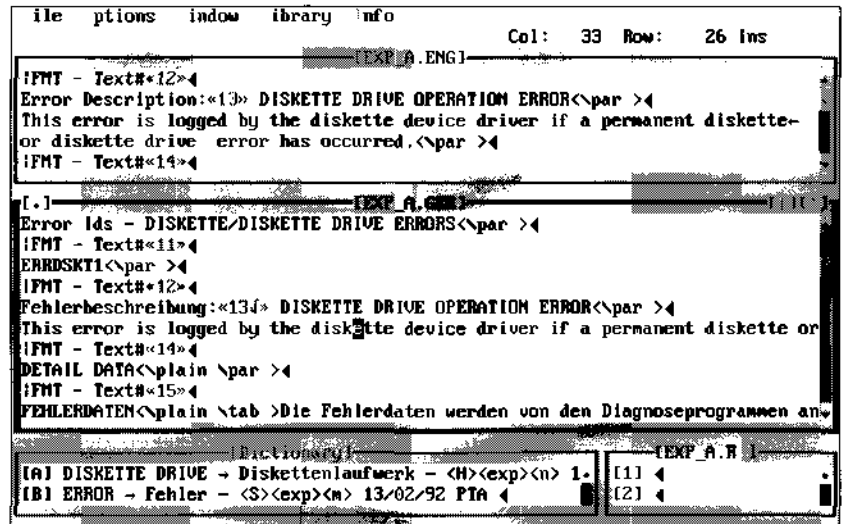
Within the program, three or four windows are constantly displayed: the source text window, the target text window, the terminology window and, optionally, a notepad window.

The source text window displays the original text, split into segments, highlighting formatting information and words found in the dictionary. The source text window can be synchronized with the target language window, which means that they scroll simultaneously. This is the usual setting, as it is useful to have the source text always displayed for reference. The translator can customise the windows, for instance, by activating 'Text reflow' (word wrap).

The target language window is much the same as the source language, as it initially displays the same segments in the same format (apart from those which have been translated automatically). The translator overtypes the original text, or inserts the translation before deleting the original. Adherence to the segment number is essential for the translation to be used later for reference.

The terminology window, once activated, will display all the dictionary entries relating to the current segment, i.e. the segment the cursor is on. Dictionary entries comprise the word in the source language, the translation, attributes like gender (if entered), the date of entry and the name of the user who made the entry, if applicable. As I mentioned above, the words which appear here are also highlighted in the source text window. The term from the dictionary can be pasted into the translation by either: pressing a key combination to select the relevant item from the terminology window, or by positioning the cursor on the relevant word and pressing a key combination which automatically replaces the source language term with the target language, no matter at which point in the current dictionary listing it appears. In other words, the program establishes a link between the translation window and the dictionary, and it can act upon this cross-referencing.

The actual handling of the dictionary is, in my opinion, a weak point in the program: although it works very smoothly as a reference in the terminology window, editing is not as straightforward. Instead of selecting within the editing window the dictionary to which the entry should be added, the translator has to activate one of the project dictionaries, using the pull-down menu, before any changes can be made. In addition, if one looks for an item which turns out not to be non-existent and one wishes to enter it (having activated one of the dictionaries), the word must be typed in again. I understand that STAR are changing this and intend to have the word or phrase inserted automatically into the appropriate field once the user has selected the 'add' option.



There is an additional feature in the currently distributed TRANSIT version (which has superseded the one discussed here): it is possible to check whether terminology offered by the dictionary has actually been used in the translation. Starting from a source text segment containing the terminology, the program scans the texts and returns all the appropriate target language segments for checking. This is a very valuable feature, since a translator may not consider it necessary to refer to the dictionary if s/he is familiar with the subject, but consistency and customer-preferred terminology may require strict adherence to the dictionary. For the same reason, dictionary entries are now highlighted in both the source and the target language window, which was not the case in earlier versions.

The user can define whether s/he wishes to work in Wordstar or Microsoft Word emulation mode. Consequently, it is very likely that s/he is familiar with the available functions. Any word-processors installed on the computer can also be accessed di-

rectly (via 'DOS') from within TRANSIT, which can be useful if new source texts have to be edited before they can be imported.

How does this differ from working with just a word-processor?

- Subject-specific dictionaries are constantly activated. A project-specific dictionary can be built up before and during the translation, whereby the translator is always reminded automatically which translation s/he used for words which may have more than one correct translation.
- As it is interactive, the dictionary database can be updated constantly, either by creating a new personal dictionary or by adding to existing ones (unless they have been write-protected).
- Consistency in the sentence structure is easier to maintain, since recurring text segments are replaced automatically by the translation used previously. This also means that if a translator embarks on a task started by someone else, s/he will be able to use the same phrases and terminology which were employed previously.
- For the translation of updates, the often time-consuming process of identifying amendments is eliminated, as the previous version may be stored in the reference directory for comparison. When the imported file is opened, a simple key combination takes the translator straight to the first new segment, i.e. a part of the text which is either entirely new, or which has been changed within the segment.

Since the release of the TRANSIT version with which I have been working, various amendments have been made, including the provision of on-line Help screens (which currently follow a hierarchical concept, but are in the process of being re-sorted) and the development of a more sophisticated matching process for the translation reference material.

The program will be able to find segments which contain differences other than numbers. This may result in more than one match being found, in which case one of these will have been inserted, but the translator is informed that there are other possibilities. These can be scrolled through at that point in the text, before the translator accepts one and edits it to achieve the correct translation. In my opinion, this feature (which other computer-aided translation programs already contain) is essential to maintain consistency within a text - if that is the aim. Technical documentation is often repetitive in a positive sense, structuring the reader's thoughts, and this example should be followed in the translation.

The manual which is supplied with the package is very thorough, and the screen models and examples are well presented. It has proved to be a very valuable reference. In some cases, it would have been useful to have a quick reference guide, listing

all the available hotkeys under appropriate context headings, or explaining in note form how to set up a project. It may be that the on-line Help screens have filled this gap. The manual is available in German and English, and a French translation is planned. The Help files have not been translated so far, possibly because STAR do not yet have a distributor in England.

Once I had familiarised myself with the program, I enjoyed using it. The structure of the system is transparent, which facilitates the management of files and projects. The fact that a variety of source text formats can be used is a great advantage. Although some aspects of the program could be improved, it obviously has a lot of potential, especially since the hardware requirements are relatively easy to fulfil. The developers are working closely with the company's translators; they understand the requirements and aim to accommodate these in the program.