Automatic Translation Comes of Age

Transit as a functional link between translation agency and translator

by Rolf Wennekes

A "tribal war" has raged for years, not simply between the makers of translation tools, but even more between their users. To a large extent liability for this virtual pitched battle–it has to be said in all honesty–must lie with the fairly aggressive marketing approach that some translation tools manufacturers chose some years ago, thereby setting the scene for a looming battle. Certainly one beneficial side effect of these campaigns was indeed a significant drop in prices across the board.

The difference between the available translation memory tools is well-known in broad terms. The Translator's Workbench, made by TRADOS, is the market leader–a fact which endows many with a comfortable feeling, even if one of the major investors has withdrawn from the

However limited or extensive though: this environment has to be prepared, set up, but most importantly learned.

TRADOS alliance in the interim. STAR's Transit is exceptionally flexible, the developers are improving its functionality constantly, and the program is popular with connoisseurs. ATRIL's Déjà Vu is the ideal tool for the smaller independent translator, cheap and simple, but in the field of terminology management it could hardly be called professional. IBM's TranslationManager has reached a stage of fossilization. With this we have covered the most important players on the translation-tool stage.

No philosopher's stone

In linguistic circles little is known about the black box, the nucleus of the products, the specific routines which the programs use–in short the *engines*. I will say nothing in this article about the differences at this fundamental level, for the simple reason that as a linguist and littérateur I know nothing of significance to say about it. I offer here a description of the translation tool Transit/TermStar from STAR GmbH of Böblingen in Germany. It is not my intention to compare it with the other packages. Let me begin by stating: Transit is also no philosopher's stone, and just like its competing products it also has its limits. When STAR introduced an entirely new Windows 32-bit version 3.0 of

Transit/TermStar at the beginning of 2000, international criticism was anything but gentle. Let's just say that in the heat of the battle to win favor with the users, STAR introduced a product to the market prematurely–an accusation which has been leveled at Windows by the experts from the beginning right through to Windows 2000, although sadly the sales figures have consistently contradicted the critics.

Translation environment

In the interim, translation tools like Transit and Translator's Workbench are no longer simple tools. One even talks now of a *translation environment*, consisting of the translation module with its required filters for processing word-processing and DTP applications, the database with translation memory, dictionaries and glossaries, an Internet environment to give freelancers access to external terminology, and the integration into 'workflow' such as STAR's ProActive into which Transit is fully integrated. Naturally this environment can be as comprehensive as the operational management requires–from individual translator to translation agency. However limited or extensive though: this environment has to be prepared, set up, but most importantly *learned*. And this latter is a process which demands weeks if not months, though this is not to say–as my practical experience has repeatedly shown–that a new user can't be translating complete sentences given half an hour and a handful of simple instructions.

For larger translation environments the construction and maintenance of client-and/or subject-specific translation data is also a factor which cannot be ignored; to
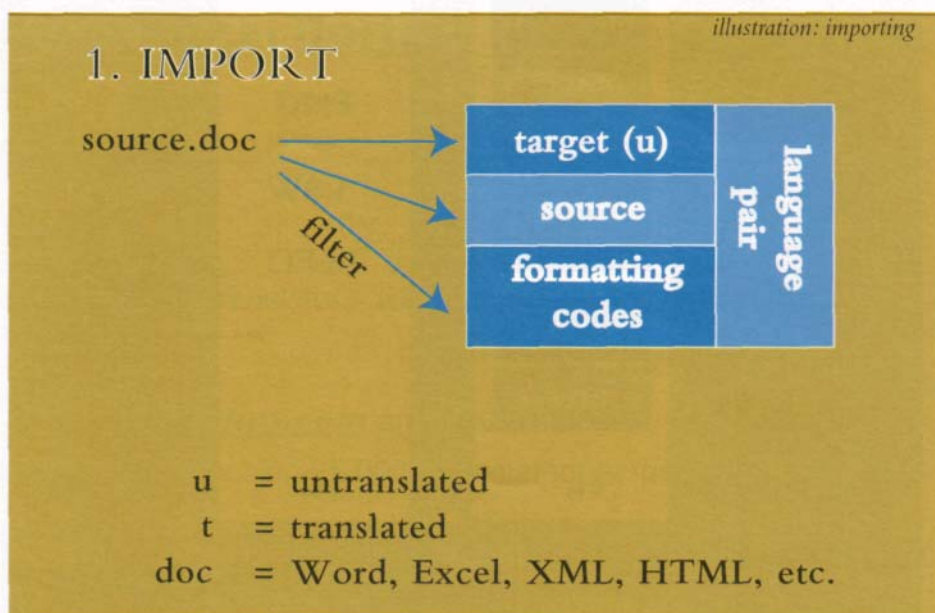
this end attention and discipline are needed, along with the necessary experience and time.

For the actual translation work, STAR offers its Transit translation tool in a number of versions: Transit Satellite Personal Edition (available free from www.star-transit.com), Transit Professional, and its derivative and somewhat 'stripped-down' Workstation version. The program can be extended with optional components such as spell checkers in various languages, a routine for preparing texts which have not been translated by Transit for inclusion in the translation database, and, for a freelance translator certainly not very affordable, filters for the more serious DTP packages.

Transit Professional is mostly used in our office for, consecutively: the importing, pre-translation and preparing for dispatch of translation assignments for our translators, the actual translation, the quality control, the exporting of translated documents and finally the maintenance of translation data and dictionaries by the project managers.

I. Importing

Below I will explain this process in more detail. Documents to be translated must be imported into Transit because the translation module of Transit is format-independent. With Translator's Workbench, for example, Microsoft Word is normally the interface in which the actual translation is carried out as a basis. 'Format independent' means that the formatting codes of the original file are converted into translator-readable *tags*,

illustration: importing

1. IMPORT

source.doc → target (u)

filter → source

→ formatting codes

language pair

u   = untranslated
t   = translated
doc = Word, Excel, XML, HTML, etc.

which are either visible during translation in Transit or can be hidden.

2. Pre-translation

The contents of a document are pre-translated alongside the importing process: the sentences or segments of the document are compared against the contents of the translation memory, a sort of text database containing bi- or multilingual segments which were translated earlier. Should there be total agreement between the segment to be translated and a segment already in the translation memory, the target-language segment is replaced by the previously-available translation. In other words, old material is simply recycled.

The quality and quantity of the pre-translation is entirely determined by the quality and quantity of the available translation memories, i.e. the reference material. Let us simply regard translation memories as

the archive on which the translation environment is based. Transit offers the user powerful possibilities to lay out this archive and to set it up: by language combination, by subject, by client. The archive, the refer-

In linguistic circles little is known about the black box, the nucleus of the products, the specific routines which the programs use—in short the engines.

ence material linked to the pre-translation process, can actually be limited with similar ease to just one single file if it is important to avoid conflicts within the version management of texts or products.

Being able to regulate the environment for the reference material with such focus is an advantage not to be underestimated, markedly raising the reliability and usability of pre-translations.

In order not to lose one's way among this flood of set-up possibilities and parameters, all specific settings per project or subproject can be stored in a type of project definition which can be reproduced at any time, for example on rounding-off a project after receiving the various translations, or when launching a new project based on the same or a similar set-up. These project parameters would include, among other things: the client name, corresponding folders for the importing and exporting of projects, the file format of the texts to be translated (standard filters available are Microsoft Word in all versions, RTF-Help, HTML 4, Excel, PowerPoint, SGML, ASCII, ANSI and XML; optionally available: FrameMaker, PageMaker, QuarkXPress), the folders containing the

Being able to regulate the environment for the reference material with such focus is an advantage not to be underestimated, markedly raising the reliability and usability of pre-translations.

reference material to be used, the dictionaries to be linked to this project, and the source and target language or languages of the project.

3. Preparing for dispatch

During the pre-translation process, various language pairs are generated according to the number of target languages desired for the translation project, during which identical segments from previous projects are completely pre-translated. During pre-translation it can also be indicated whether segments or sentences must be sought which match by less than 100 per cent and if so, what percentage must constitute the minimum quality of the agreement (*fuzzy match*) and how many segments each time must be regarded as reference material. These adjustments determine the so-called 'extract' which is laid down for the archive components prior to the pre-translation,

2. PRETRANSLATION
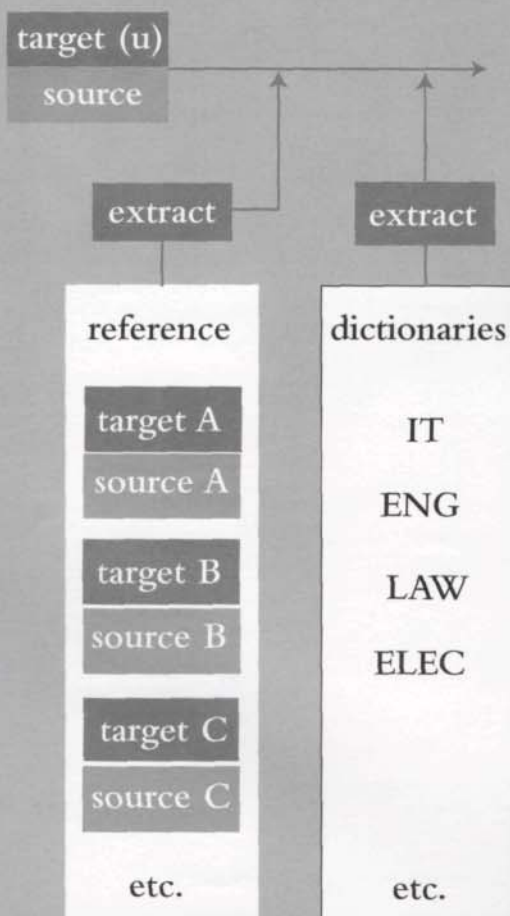
target (u)

source

extract   extract

reference	dictionaries
target A	IT
source A	ENG
target B	LAW
source B	ELEC
target C	
source C	
etc.	etc.

illustration: pre-translation

3. SEND

| pxf SPANISH |
| pxf FINNISH |
| pxf ENGLISH |

4. TRANSLATE

| target (u) |
| source |
| extract |
| dictionaries |

illustration: send & translate

along with the abovementioned language pairs. Optionally, a statistical tab-separated text file can be generated with the pre-translation, to be processed for order-administrative purposes by a spread-sheet application.

The project manager now prepares the material for dispatch to the translators. This step is also carried out in Transit, using the function *Send Project*. Here the project

Once a segment has been translated and marked as such, it is immediately available in turn to be recycled within the running translation.

manager indicates the desired target languages and whether the dictionary linked to the project must be sent or not. With a simple routine he also generates a file per target language containing the project particulars such as file format and language combination, the required language pair, the translation memory extract and the dictionary, and this file is immediately compressed.

The file is sent as an e-mail attachment to the translators or transferred to the FTP server. Via the receive option in Transit (*Receive Project*) the translator installs this file and all necessary items with one single click. Thereafter the *human translation* process can begin.

4. The translator at work

The text is translated segment by segment by the translator during which TermStar–the dictionary integrated in Transit–allows available words to be 'pasted' into the translation with a simple key combination, while the translator can make use of translation memory extracts with a threshold which can be set for the minimum desired quality of the *fuzzy matches*. Once a segment has been translated and marked as such, it is immediately available in turn to be recycled within the running translation.

Remarks can be linked to each segment by the translator, for example in case of doubt about the accuracy of a translated technical or specialist term. The project manager

goes through this list during the quality control, during which the segments requiring his or her attention are immediately shown.

5. Terminology management

During the translation process the translator can very easily add terminology to the project dictionary. The Transit dictionary component TermStar, mentioned only incidentally here, is extremely well integrated and offers very powerful functions for efficient terminology management. Terminology management is an integral component of computer-aided translation and it is of great importance that a translation tool has seamless access to the translation module's dictionary function. TermStar can be set-up to be as simple or as complex as you wish: from a summary two- or multi-language word list to a complex terminology databank with clickable cross-references, import fields for synonyms, context and reliability of the term, source, grammar and even an illustration, for example of a mechanical component. I have worked with electronic dictionaries since 1985 and have got to know a large number of such programs covering various media. I believe the functionality of TermStar cannot be surpassed right now.

The translator translates the project, proofs it, carries out a spell check and checks the formatting codes. The translation is now ready to be returned to the translation agency. The translator can use the *Send Project* function and the transmission possibilities offered by the Internet.
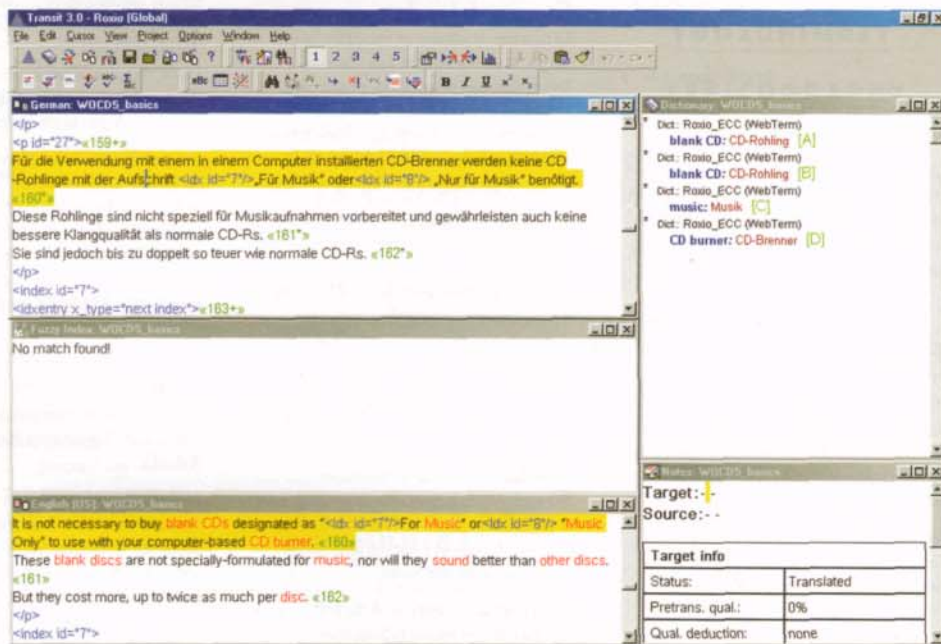
illustration: transit screen

5. RETURN

6. EXPORT

```
      ┌──────────────┐  filters
      │ target (u)   │
    ┌─┤──────────────│           → target SPANISH.doc
    │ │ source       │
    │ └──────────────┘
  } │ pxf SPANISH
    │ ┌──────────────┐
    │ │ target (u)   │
  } ├─┤──────────────│           → target FINNISH.doc
    │ │ source       │
  } │ └──────────────┘
    │ pxf FINNISH
    │ ┌──────────────┐
    │ │ target (u)   │
    └─┤──────────────│           → target ENGLISH.doc
      │ source       │
      └──────────────┘
      pxf ENGLISH
                          ┌────────────┐
                          │ formatting │
                          │   codes    │
                          └────────────┘
```
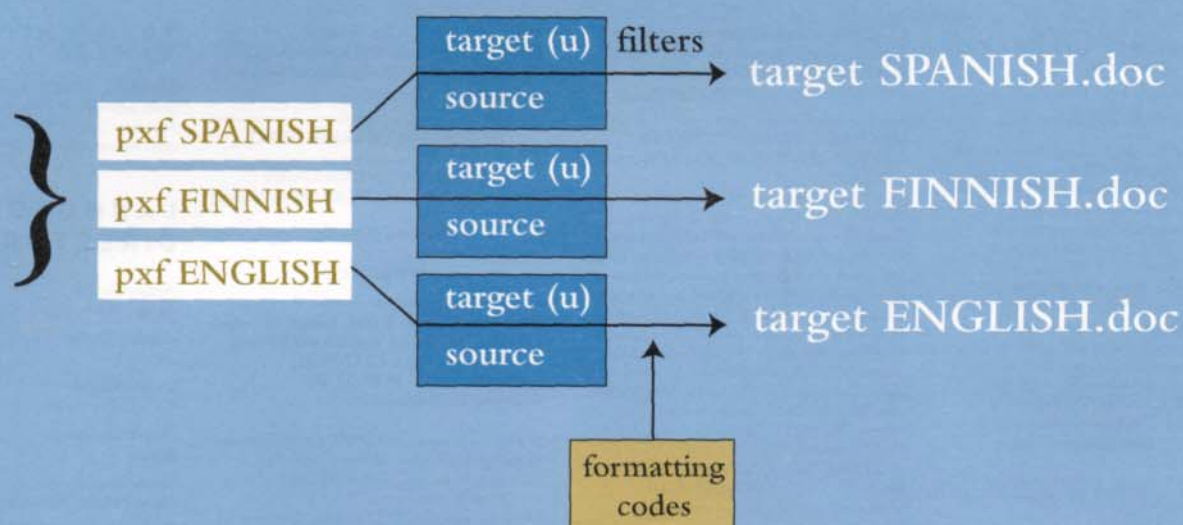
illustration: quality control & export

6. Quality control and exporting the translation

The translated project is copied to the specific project by the project manager via *Receive External Project*, after which the quality control cycle is completed and the translation is exported, together with the other language combinations, all in one single step. During the export the translated text is converted back to its original format. After a final visual check the translation is ready to be delivered to the client. In this article I have reproduced the *workflow* in broad outline as it is currently incorporated as a structural component in our quality route and operational management. It will be obvious that the software described here performs excellent service in a complicated translation environment such as that of a translation agency, and that in a lighter version it is extremely suitable for the actual production of the translations. The integration of these environments continues to be improved, also making steadily greater use of the Internet by implementing Internet based workflow and terminology instruments like WebTerm, ProActive, and GRIPS.

By the time this article is published, developments will have overtaken the facts once more. In March or April of 2002, STAR releases Transit XV, successor to Transit 3.0 discussed in this article. During the presentation of this latest STAR offspring at the Tekom show in Wiesbaden in early December last year, I learnt about the most important improvements. The new version

> **I have worked with electronic dictionaries since 1985 and have got to know a large number of such programs covering various media. I believe the functionality of TermStar cannot be surpassed right now.**

has a powerful concordance function which offers not just the ability to seek by keywords in translation memories, but which also allows the user to process found segments, a very crucial feature for the improvement of reference material. From Transit XV it's possible to retrieve so-called *internal repetitions* ; this refers to segments which do not (yet) occur in any translation memory, and which are repeated a number of times in the new text to be translated. Furthermore, Transit can now notify a translation agency via the Internet on a translator's progress with a project, a function which will be very useful for project managers concerned with status reporting on larger projects.

English translation: Terence Kennedy

Rolf F. Wennekes is a linguist and littérateur, and is a sworn translator and founder/director of protext TRANSLATIONS B.V. in the Netherlands. He studied German, general literary theory and philosophy in Bonn and Leiden. He can be contacted at rolfw@protext.nl.