



# THE ALIGNMENT TOOL

## How to exploit previously translated texts



Pierluigi Melotto

A text alignment tool enables translators and terminologists to exploit a precious resource: previously translated texts. In this article I will briefly explain the purpose, the main features and how an alignment tool works, thus providing an overview which should render this software attractive as a time and cost saving solution. The author wishes to thank the Public Relation department of Trados GmbH for their contribution.

### What it is

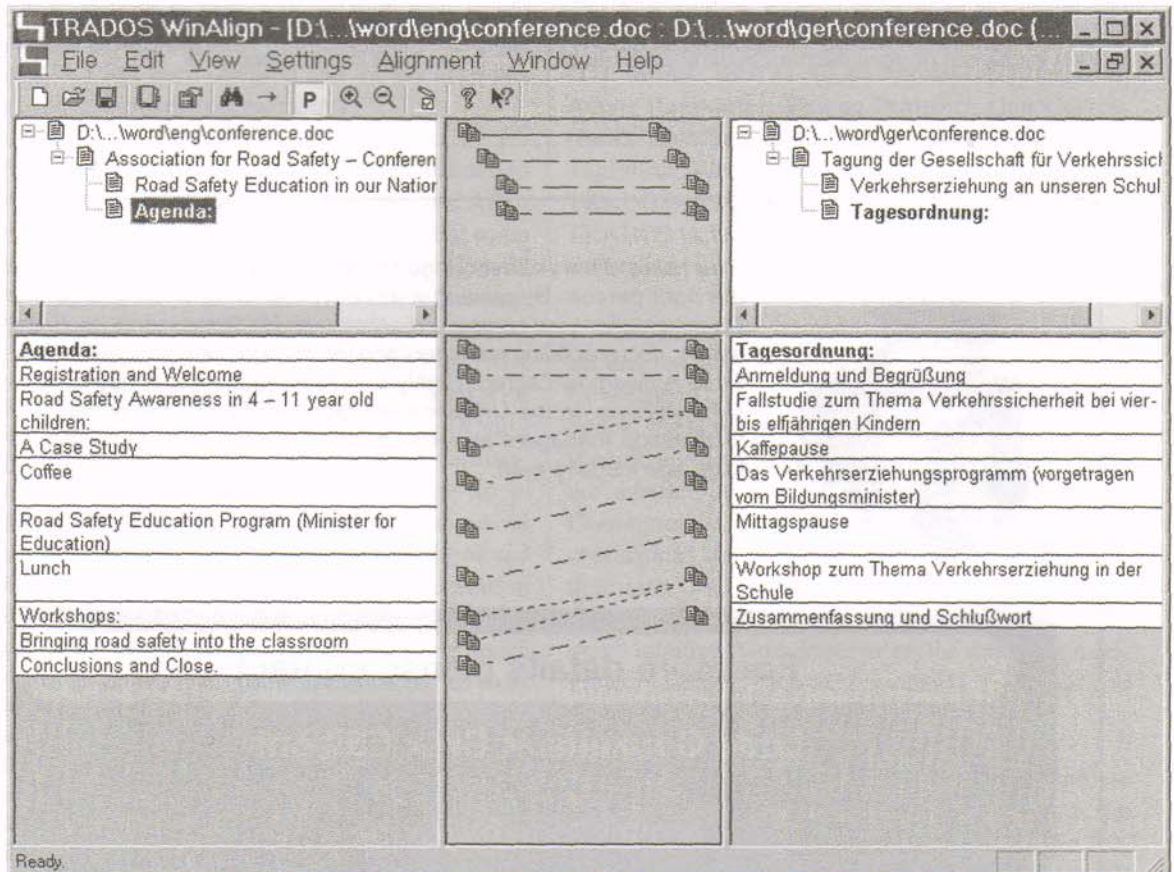
To some readers the difference between Machine Translation and Computer-Aided Translation is a common knowledge. Basically the former provides an automatic translation of parsed source texts, the latter allows a machine-aided human translation. Although research continues, semantic disambiguation of input texts remains a major hurdle for MT. CAT is considered a promising solution for high quality translations of large texts and the use of this technology is rapidly growing. In order to generate a translation in the target language CAT systems process the source text in batch mode,

consulting terminology databases and translation memories.

Thereafter a post-editing of the target sentences can be carried out by human translators. In-house translators of translation agencies work together on the same translation project in a networked environment sharing the same translation memory. Freelance translators work on translation projects at home with their own translation memory system. Initially translation memories are empty.

They can be created by translating the text with the CAT system or by means of a program called "Alignment Tool". Translation memories are of value and their ownership is becoming a strategic issue. On this point I suggest reading Who Owns Translation Memory? at <http://www.mcbsys.com/html/news/10/tmowner.html>. Investments in these systems are justified also by the possibility to exchange translation memory data among different translation tools (see The Importance of TMX for Translation Tool Buyers at:

<http://www.languagepartners.com/reference->



center/whitepapers/110nwp/tmx-importance.htm).

As a terminologist I appreciate alignment tools. In 1997 I have updated one of the most important bilingual technical dictionaries published in Italy. I created a list of 6.000 German/Italian entries, covering nearly 70 sublanguages, like aerospace engineering, automotive industry, biotechnology, chemistry, etc. To do this I worked hard manually with MS Word having a tight schedule in front of me. But later, when I started working with alignment tools, I understood how helpful they might have been for this job. In the future dictionary projects in which I will be involved I will surely benefit from the potential offered by this software. In addition, TRADOS 5 now includes ExtraTerm, a new revolutionary terminology extraction tool. ExtraTerm dramatically cuts the effort involved by automatically extracting a list of candidate terms to be easily edited and validated. Context sentences can be included if required. The output from ExtraTerm can be easily imported into MultiTerm.

### Setting Up the Tool

Alignment tools are fed with previously translated files as input and provide translation memory files as output. This allows to reuse old translations for current and future translation projects. In order to describe the functionalities of one of these tools I have run an interesting tutorial of the program WinAlign, contained in a CD-ROM provided by Trados GmbH. Alignment files created with WinAlign are exported in translation memories which can be used with Trados Translator's Workbench. The first step is the creation of an Alignment Project. Before starting an alignment the user is required to define various project settings, some of which are listed below.

- Source and target language. It should be noted that WinAlign is a UNICODE-based program capable to support all languages supported by Windows NT. A description of the Unicode Standard can be read at <http://www.unicode.org/unicode/uni2book/uc20ch1.html>.

- Segmentation rules. An abbreviation list can be defined by the user. In this way WinAlign is prevented from interpreting abbreviation dots as full stops during the alignment phase.

- File name pairs. The graphical interface allows to select the files and carry out the file name alignment in a user-friendly environment, made up by source file frame, link frame and target file frame.

- Alignment options.

- Granularity. The default text alignment is by sentences. Paragraph-based alignment is suggested for Asian languages.

- Tuning. Here the significance of tags and numbers encountered during the alignment process can be increased or lowered. Also other tuning options are available, such as the Formatting Significance.

- Structure Recognition. The alignment results can be optimized by selecting the options Styles, Paragraph Numbering, Font Sizes, and Tags.

### Let's get aligned

WinAlign allows to align automatically all files of a project or to align file pairs individually. The results can be examined and edited in the alignment editor of WinAlign. This interactive interface is made up by the outline area, in which the structure of the alignment results is displayed, and the segment area, in which the segment alignment results are displayed. Each area is subdivided in source, link and target frame. It should be highlighted that each alignment unit retains the formatting information of the aligned texts. This feature is very useful in order to verify the correctness of the alignment. Misalignments can be easily corrected by connecting or disconnecting the various source and target segments. Thereafter the segment alignment results are confirmed. WinAlign offers also the option of re-aligning the file pairs, thereby increasing the reliability of the results. At this point an alignment file is generated and imported in a translation memory. ■