System Demonstration

TRADOS TRANSLATOR'S WORKBENCH

Mark Berry
MCB Systems

1. System Builders and Contacts

Developer

TRADOS GmbH
Hackländerstrasse 17
D-70187 Stuttgart
Germany

Tel. +49 (711) 168 77-0
Fax +49 (711) 168 77-50
Email info@trados.com
Web http://www.trados.com

North American Sales and Support

MCB Systems
325 West Washington Street Suite 6
San Diego, CA 92103-1946
USA

Tel. (619) 298-8300
Fax (619) 298-8399
Email sales@mcbsys.com
Web http://www.mcbsys.com

2. System Category

Commercial translation memory system.

3. System Characteristics

The TRADOS Translator's Workbench can be used either in interactive or batch mode. The interactive mode works in conjunction with either Microsoft Word for Windows or WordPerfect for Windows, whereas the batch mode works directly on documents stored in Rich Text Format (RTF). Sentences pairs are stored in translation memory in RTF format.

The following formats are supported by running a pre-process Word macro: Windows resource (RC), SGML, Ventura, Interleaf, DCF (Bookmaster), troff (Unix), HTML, and Quark. An optional filter program can pre- and post-process FrameMaker files, with a similar filter under development for Interleaf.

4. Resources

Both the terminology database and the translation memory are empty upon purchase. The user builds the terminology database manually or by importing existing terminology data. Translation memory is built automatically as the user translates; it is also possible to use an alignment program (TAlign) to build translation memory from existing translations.

5. Hardware and Software

The software runs on an IBM-compatible PC under Windows 3.1, Windows 95, or Windows NT. A Pentium with at least 16MB of RAM and a 17-inch monitor is recommended.

6. Functionality Description

The Workbench includes two components: a terminology database program, MultiTerm '95 Plus, and the Workbench itself, which is responsible for translation memory access and maintenance.

6.1. MultiTerm '95 Plus

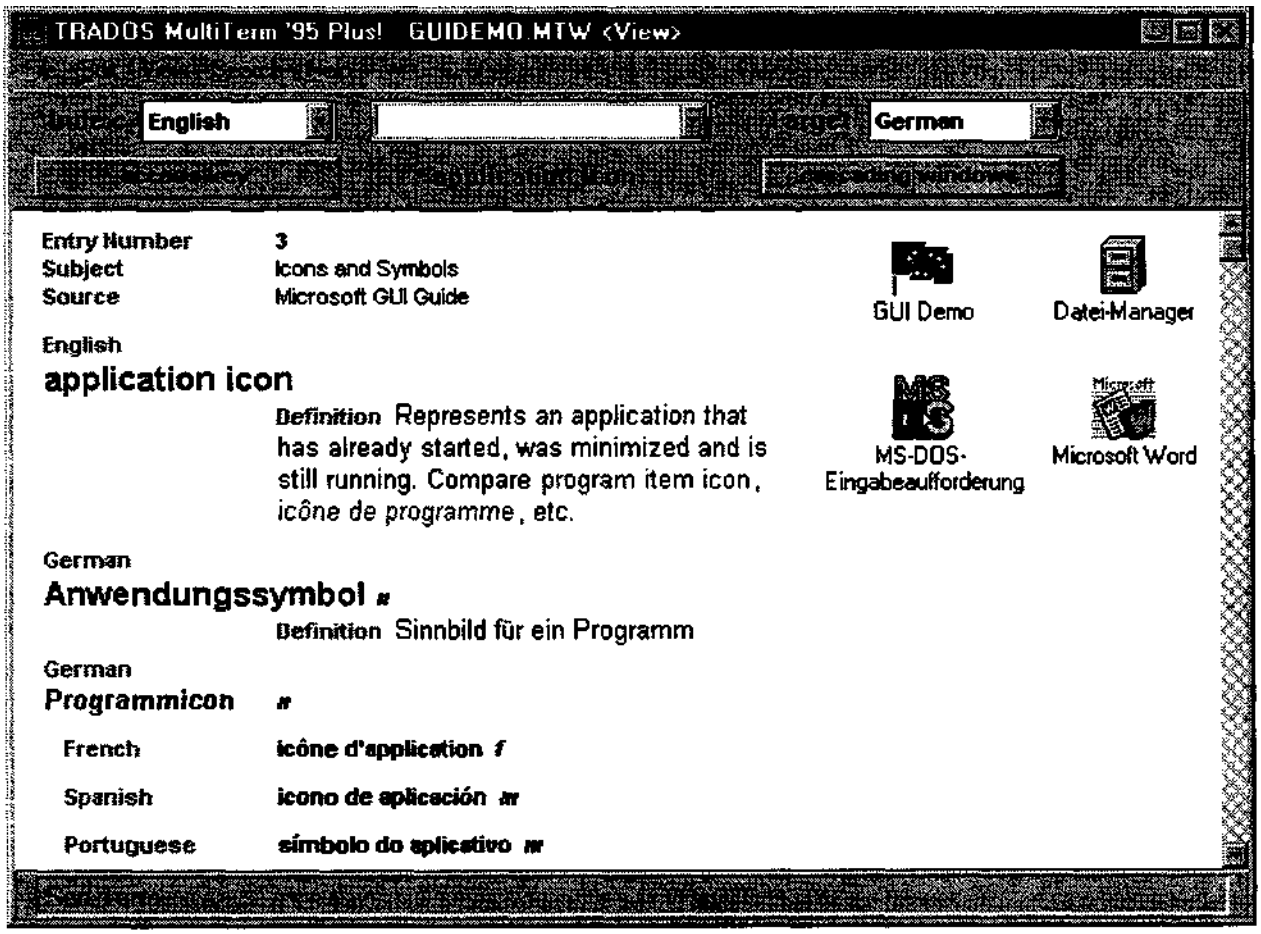

Figure 1. MultiTerm '95 Plus

MultiTerm is a program designed specifically for building and maintaining terminology databases. Each record in a MultiTerm database is called an "entry." The entries are "free-format": you decide how much or how little information to include about each concept. A simple entry might include two terms and perhaps a definition. A complex entry could include terms in five or ten languages, plus synonyms,

cross-references, examples, and identification of the sources for the information. There are no fixed-lengtti fields as in conventional databases.

Terms in MultiTerm can be one or more words. Terms are always indexed in all languages. You can instantly switch from, for instance, English-Spanish to French-Spanish to German-English.

MultiTerm lets you classify terms with user-defined "attributes." For instance, in a glossary of computer terms, you might classify some as PC terms, and some as mainframe terms. Or you might note that IBM prefers one translation for a term, while Digital and HP prefer another. These attributes can be used later as search and export criteria.

To search for an entry in MultiTerm, you can type the first few letters of the term, or you can use a "wildcard" character to display a hit list of matching terms (for example, * box* to display all terms containing the word "box"). MultiTerm also supports fuzzy searching: if you precede the search term with a pound sign (#), MultiTerm will display a hit list of all terms that are similar to the term you enter. This allows finding terms when you are unsure of the spelling or word order (for example, #interior ministry would find "Ministry of the Interior").

MultiTerm has been specifically adapted for network use. Multiple users on a network can all read from and write to the same database, so they are all using the latest data at all times. "Entry-level locking" prevents two users from trying to change the same entry simultaneously. The network administrator can set access rights giving different privileges to different classes of users.

Additional MultiTerm features include the ability to define fonts and colors for display, support of color graphics in entries, logical operators in filter and export criteria, and sophisticated entry classification via pick list fields.

MultiTerm is available as a stand-alone terminology management program as well as being included with the TRADOS Translator's Workbench.

6.2. *TRADOS Translator's Workbench for Windows*

The Workbench is a translation memory program designed for making human translators more efficient in their work. It is not a machine translation program, which means it does not try to formulate sentences on its own. However, it is possible to pre-load Workbench translation memory with sentences translated by a true MT system, in which case the Workbench becomes a post-editing environment with the additional benefit of remembering previous edits.

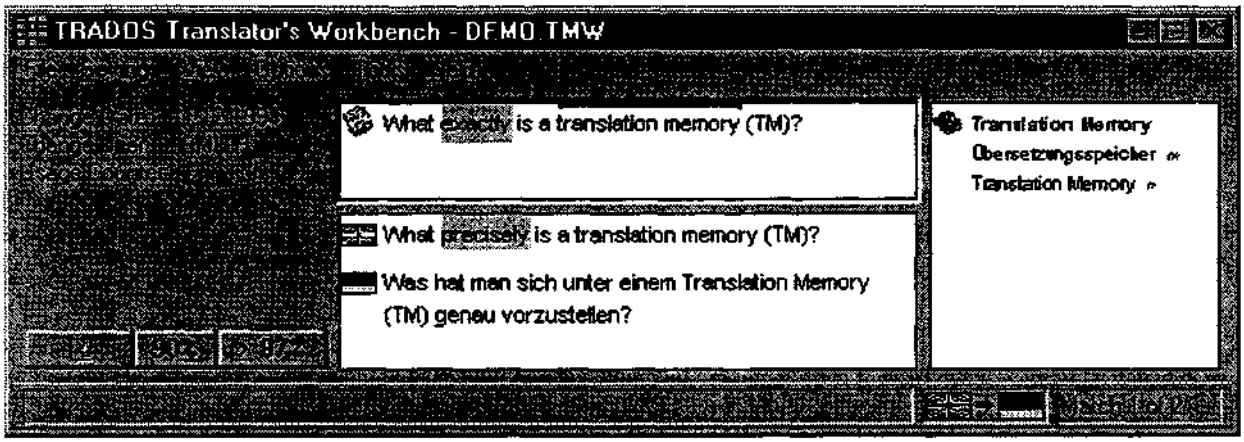

Figure 2. TRADOS Translator's Workbench

To use the Workbench in interactive mode, the user starts the Workbench, MultiTerm, and Word or WordPerfect. Once a document is opened in the word processor, the user places the cursor where he or she would like to start translating, and clicks a button that is tied to a word processor macro. This macro passes the text to the Workbench.

Workbench segments the text and displays the source sentence both in the Workbench window and in the word processor. Workbench then looks up the sentence in translation memory using a fuzzy search, and if a matching sentence is found (within user-defined tolerances), the previous translation is displayed in the Workbench window and in the word processor, and differences between the old and new source sentences are highlighted in different colors. If no matching sentence is found, an empty editing area is displayed in the word processor. Additionally, Workbench uses a fuzzy search in MultiTerm to find any known terms in the new source sentence; these are identified with a red bar, and their translations are displayed in a Workbench window.

Using standard word processor editing commands, the user either adjusts the old translation to make it a correct translation for the new source, or simply types in the new translation. Translations of terms found in MultiTerm can be pasted with a keystroke. A concordance feature allows the user to select any string and immediately display a list of all sentences in translation memory containing that string. The sentences are displayed with their translations, thus providing for quick terminology research of terms not (yet) included in the terminology database. Once a sentence is complete, the user clicks another button to move to the next sentence; the sentence just translated is immediately stored in translation memory, and the process repeats.

In batch mode, the Workbench directly reads RTF files, replacing sentences with previous translations. The user can specify whether only 100% matches should be replaced, or whether matches lower than 100% should be included, and whether term translations from MultiTerm should be inserted. The result is a document with some sentences already translated that can then be passed on and finished by someone who only has a word processor. A provision exists for updating translation memory with such work done "off-line." The color-coding of differences and the concordance feature are not available to the translator in this scenario.

Due to its reliance on neural network fuzzy matching, the Workbench is largely language-independent. The following list of available source and target languages is used primarily for identifying languages and their code pages and not for any difference in functionality. Users can set up a translation memory for translating from any of these languages into any other language or languages (multi-lingual translation memories are supported).

Brazilian Portuguese	French	Polish
Catalan	French Canadian	Portuguese
Czech	German	Russian
Danish	Greek	Spanish
Dutch	Hungarian	Spanish (Mexican)
English (UK)	Italian	Swedish
English (US)	Norwegian	Welsh
Finnish		

Recently, the target language list has been expanded to include Other European and Other Asian. Workbench only indexes the source language, and since Asian is only supported as a target language, it is not included in the fuzzy index.

7. System Internals

The Workbench uses neural-network database technology and fuzzy logic for storing and retrieving translation memory units. The Workbench also access data in the MultiTerm terminology database via fuzzy searches. The advantages of this approach are:

- language independence
- quick access, even to large translation memories
- possible to access translation memory based on sentence fragments (concordance feature)
- fuzzy terminology recognition finds multi-word terms even if the words do not appear together in the sentence

The Workbench uses Microsoft Word or WordPerfect as the "front-end" user interface. This provides for a familiar environment and automatic preservation of word processor formatting.

Dynamic Data Exchange (DDE) is used to communicate between the Workbench and MultiTerm and between the Workbench and the word processor.

Users on a network can share and simultaneously update the same translation memory. Since translation memory is updated immediately in interactive mode, users have access to their colleagues' work as soon as it the sentence is translated.