EPTAS A Client/Server Based Translation Support System

Dr. Klemens Waldhör
EP Electronic Publishing Partners GmbH
Fürther Str. 212
D-90429 Nürnberg
e-mail: klemens@epp.de
WWW: http://www.epp.de

Introduction

This paper presents the translation support system EPTAS. EPTAS combines the power of different translation approaches like machine translation or translation memories, terminology management within one system. EPTAS can be integrated into different systems using definable interfaces and supports centralised as well as de-centralised working environments within Intranets as well within the Internet. This paper explains the architecture behind the system and gives a typical example how a translator works with EPTAS.

General considerations

The demand for translations is increasing year by year. This concerns the public as well as the private sector. Nevertheless all customers require that this increased demand should be done faster and better in order to meet competition. In order to solve these contradictory requirements the translator is forced to use various translation support tools which should improve the speed as well as the quality of the translation process; this is done by using different applications like:
- electronic dictionaries which improve the retrieval of relevant terminology
- translation memories for reusing existing translations
- machine translation
- phrase translators for automatically replacing repeating phrases and terminology
- terminology management to ensure consistent translations over different projects and customers

A term widely and likely to be used in this area is the notion of **"integrated translation environment".** This term covers a broad range of applications starting from the integration into the word processing environment to a company wide usage of tools and re-usage of data including modelling the workflow of documents through the organisation and an automatic billing system. The introduction and efficient use of those tools and environments is not possible without a certain amount of time and money, maybe besides the usage of electronic dictionaries. It is questionable if for small companies the usage of such tools is worthwhile and how fast the Return-On-Investment can be reached. One has also to recognise that many of the above mentioned features are recommended from an academic point of view but no system really exists supporting all features.

EPTAS - The architecture

EPTAS is designed as an integrated translation environment. This tackles the following points:
- Integration of different translation and support tools
- Data storage and retrieval: fast access to multilingual databases
- Data protection and security: special concepts for a restricted access to the system, data compression and data encryption
- Administration: user and user group support, controlled system access
- Workflow: sequence of definable translation steps; integration of external translators by e-mail
- Statistics: accounting tools for a transparent cost control

System design and components

EPTAS was designed with an "open architecture" in mind. Open architecture means a system which subsumes different types of translation technologies and tools (as translation memories, phrase translators, terminology systems or machine translation - even from different vendors) within one shell and the capability to be extended by integrating additional components. In order to fulfil different requirements with regard to performance and operating systems EPTAS is available as a stand-alone-system and as a client-server-application. EPTAS is implemented in such way that it is not very

difficult to add new types of applications (e.g. including a new text filter or a concordance tool). EPTAS is available for Windows 95, 98, NT Workstation and NT Server. A version for UNIX exists too. A future version will support the tight integration between tools running under UNIX and tools running under Windows.

The following variants are available:
- **stand-alone-system:** All tools and applications reside on one PC. This is the typical environment for a single translator or small translation bureau.
- **e-mail based system:** The main translation tools are running on a server machine. The user communicates through a client interface with the server. The text to be translated is sent by e-mail to the server together with the necessary parameters controlling the translation process. The server receives the mail, inspects the incoming translation message (called a translation order), executes the necessary applications and returns the result to the user by e-mail. The client starts the post translation editor and the translator may now inspect and change the translation. When exiting the translation editor the translation is automatically converted back to its original text processing format. The advantage of this approach is that it requires a simple Internet access to be able to translate from all over the world without loosing the advantage of a centralised database but staying independent from the location of the central organisation. This variant is attractive for use in bigger (transnational) organisations or for translators who work for companies preferring to keep their data centralised. A less important disadvantage is the degradation of performance due to the fact that no direct connection is kept to the central server. In order to keep the files transferred small the files are compressed. In addition the files can be encrypted to avoid hostile attacks during the way through the Internet.
- **TCP/IP based system:** The basic approach is similar to the e-mail approach. The main difference is that here a direct connection is realised between the client and server. As a result the system is faster. In order to get real advantages to the e-mail approach this requires the existence of a fast network system. This may a the solution of choice within an Intranet.

Some *details on the client-server communication*

The translation server (TS) runs in a Windows NT and in several UNIX environments. The following tasks are performed by the TS: It receives "orders" from the user (configured with a special application TSI or called from within text processing system), inspects the content of the order with regard to input document names, tools to activate and so on, converts the input texts using text format filters into the internal EPTAS document format **(TTE format)** based on Unicode and HTML 3.2 and then applies a series of different translation tools to this document. In the last step the document is returned to the user. The main advantage of the system is that it allows the simultaneous usage of different translation methods, like TM and MT in parallel. The user may either check the translation results with a special translation editor (TSE) or directly reintegrate the results into his text processing system. The translation tools are available either as stand alone applications or as ActiveX server components. In addition EPTAS allows the incorporation of translation tools from other vendors.

The following **translation tools** are available:
- a sophisticated database environment based on a SQL database and an EP developed database system
- a translation memory system consisting of an alignment tool, an importer into the database a fine tuned retrieval tool and various TM maintenance tools
- a phrase based translation tool
- a translation editor for post processing the translated texts
- a terminology editor
- special MT filters which allow the easy integration of MT systems of different vendors, e.g. LOGOS
- Various support tools for converting from and into the central database(s)
- merging data and maintaining TMs is implemented

The following **text converters** have been realised:
- WinWord converter
- RTF converter
- HTML converter
- Word Perfect converter
- ASCII/ANSI (plain text) converter

- MT filter
- Additional converters will be available soon (e.g. SGML converter)

EPTAS has also been integrated into EPs SGML publishing environment **Sirius,** where it supports the storage of documents in multiple languages within one database and giving access to features like version control and so on.

A typical EPTAS session

In the first step the translator loads the document into his text processing system, e.g. WinWord. Alternatively he can start the system directly from his desktop (see picture 1). Each button represents a specific operation like translate a document, choose target language, align documents, import aligned documents into database, set the translation environment (see picture 2), start terminology editor or start translation editor.
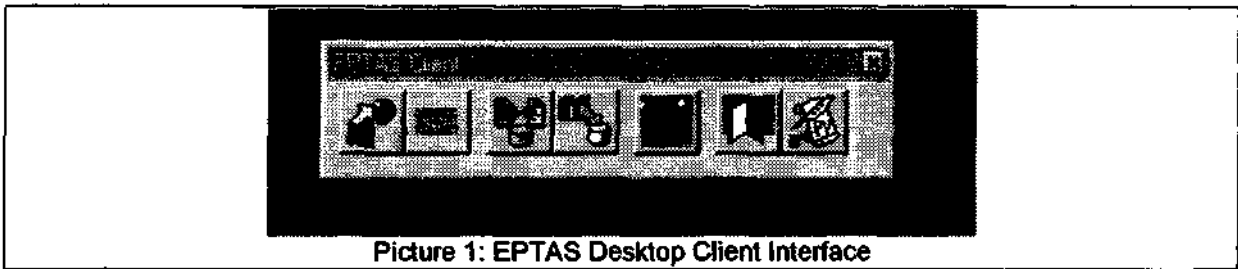

Picture 1: EPTAS Desktop Client Interface

The translator actives the EPTAS client and checks the presettings of the system, defines source and target language, which translation memory and phrase databases to use and if available which machine translation system to activate.
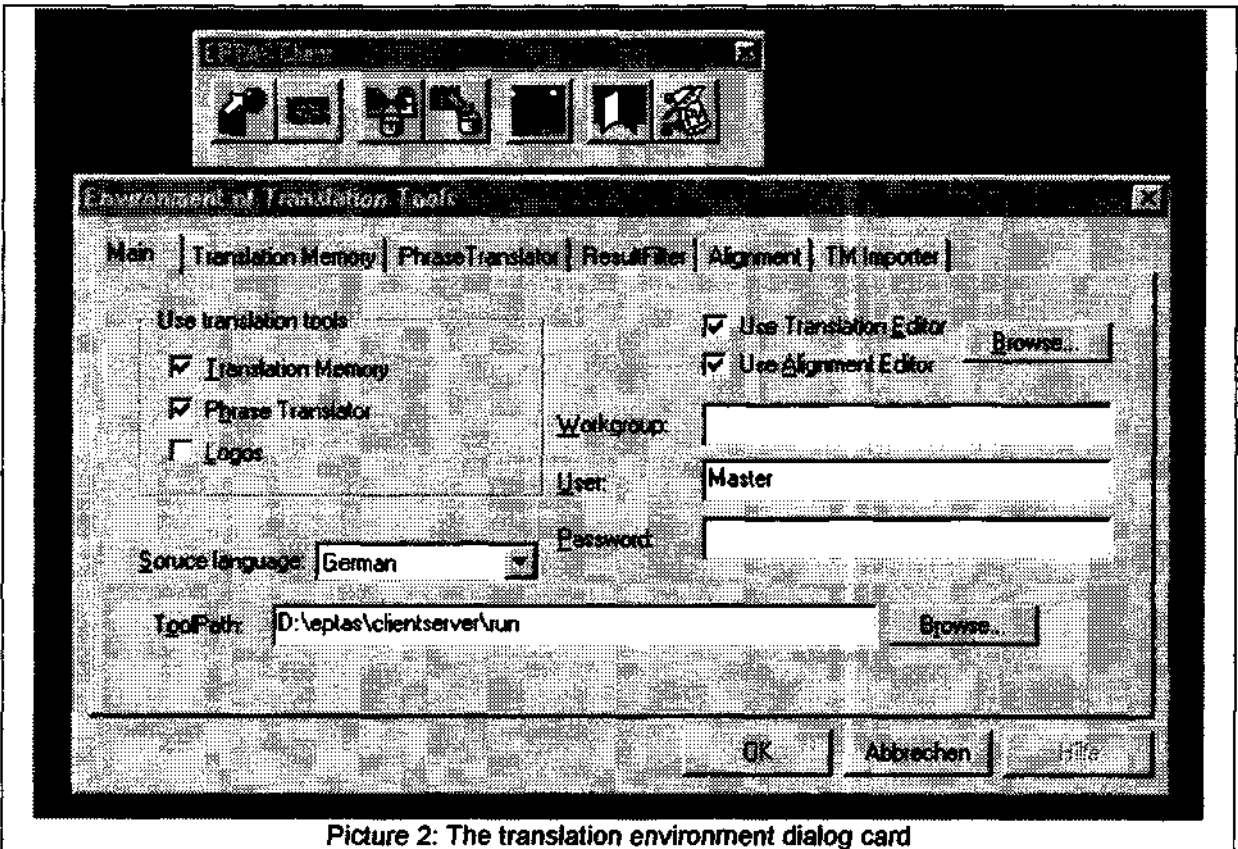
Picture 2: The translation environment dialog card

Picture 3 shows the translation memory settings. Its main purpose is to define the databases to be matched against the text to be translated. The advanced button is used to define the amount of

information to be produced by the TM. Selection Criteria define restrictions on the TM entries to be retrieved, e.g. that customer specific entries should be used only. Replacement Classes define general concepts like numbers, proper names and so on. The bonus field allows to set specific matching points for a given database, e.g. that matches within the customer database should receive a higher matching quality than other databases. The number of matches considered can be restricted using the Best Matches field.
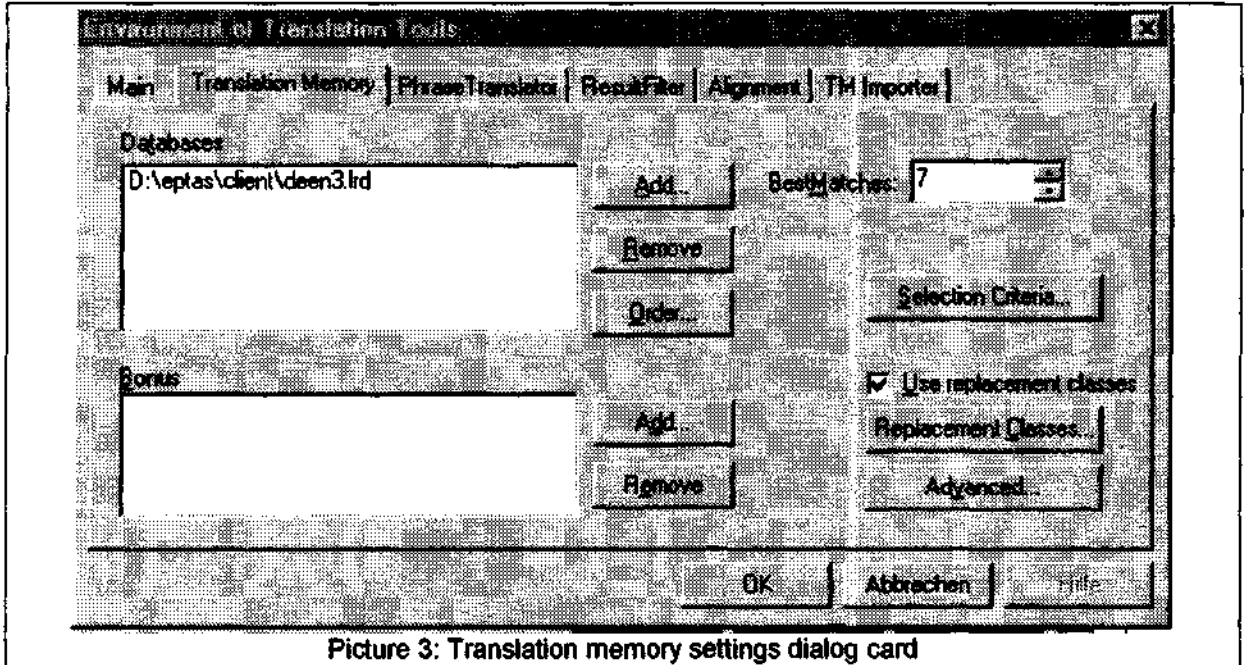
Picture 3: Translation memory settings dialog card

The order of the translations found by the translation tools can be defined by using the **Result Filter** card. Picture 4 shows that first the user translation should be presented if one exists, next the translation memory result if available, than the phrase translator and last machine translation.
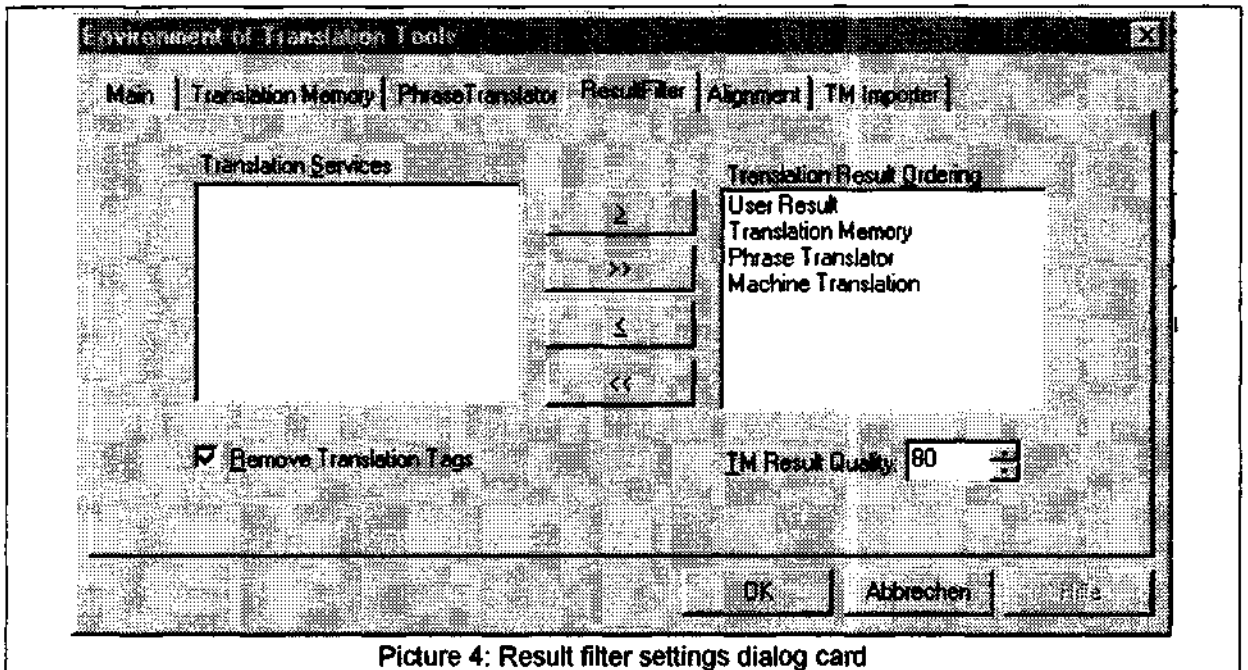
Picture 4: Result filter settings dialog card

Now he starts the translation process. Depending on the EPTAS system available the document is transferred to the server and the translation tools applied to the text. When finished the translated document is returned to the client and the translation editor activated (see picture 5). The translation editor displays the source and translation text in tabular form subdivided into sentences. For each

sentence the translator may change the translation by hand or choose and change alternative translations if more than one translation is available, e.g. from MT.
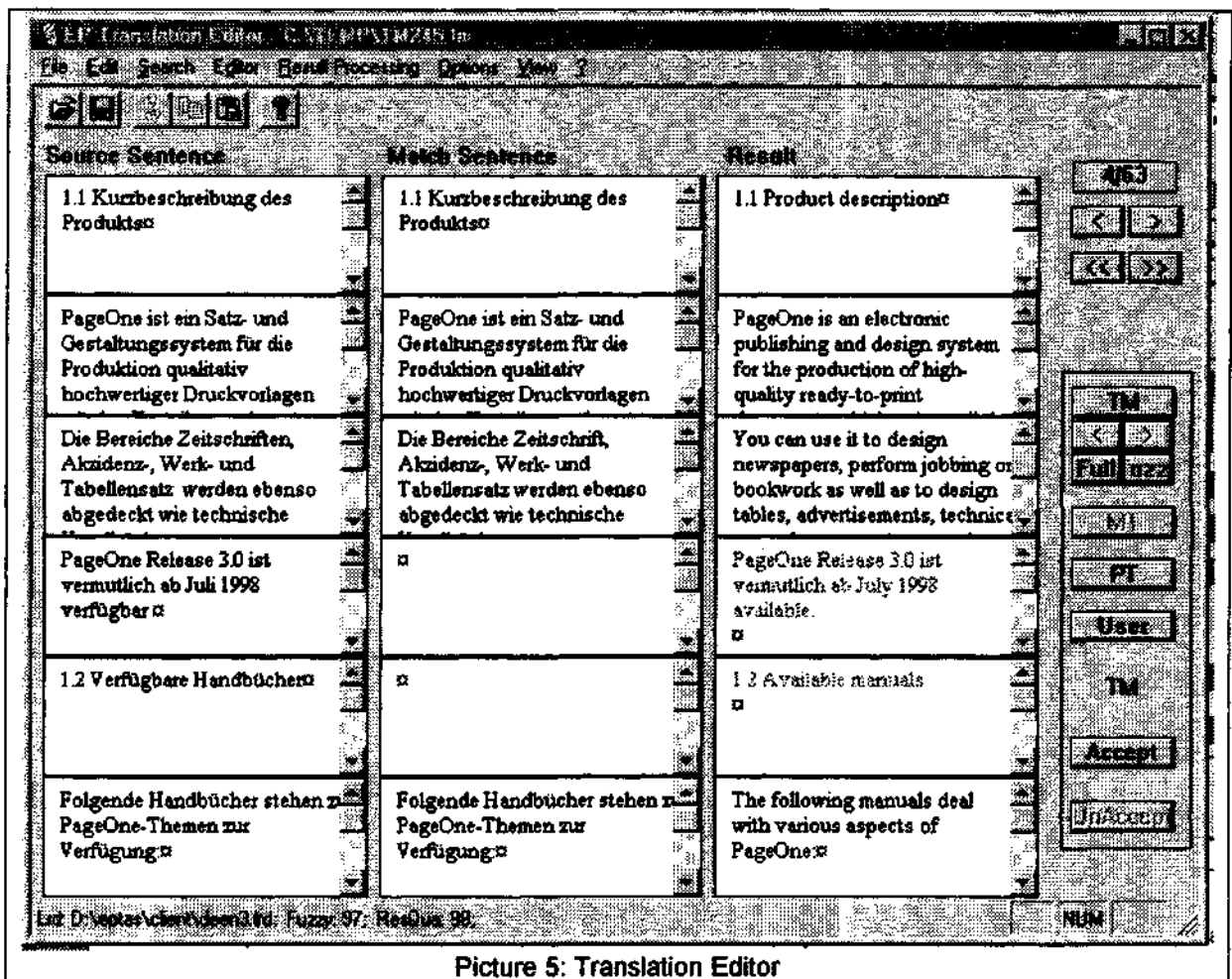
Picture 5: Translation Editor

By clicking on the TM button TM matches are shown sorted by their matching quality. Clicking on PT presents the results of the phrase translator. MT displays the result of the machine translation and User the translation entered by the user. The actual chosen translation type is mentioned by the text below the User button, in the case of picture 5 it is TM. Different translation types are shown in different colours too (e.g. blue for MT). Additional information is shown in the status bar of the editor, e.g. the matching database and match quality. The Accept button marks the translation which should be inserted into the final translated document.

As the translation tools insert mark-ups for the translation quality the translator may choose only to inspect sentences which do contain a translation memory quality below 100 %. Thus he jumps only to sentences with a smaller translation quality. When finished he quits the translation editor. If the document was launched within a text processing system the client now automatically pops up the text processing system and converts the translated document back to the text processing format. He can do now the final polishing. If the translation is validated the translator may choose to integrate the new translated document into the translation memory. This is of importance if the document will be revised in the future.

Who benefits from EPTAS

EPTAS is designed in order to support the translator at the one hand and the organisation behind the translator on the other hand. This concerns the following points:

• validated translation quality through re-usage of data and existing translations
• enhanced productivity through shorter production times and data throughput
• faster translation times

- one translation tool for all languages
- as flexible as the organisation, e.g. supports remote work of translators
- enhanced working satisfaction for the translator by concentrating on translation relevant activities
- clear efficiency enhancements through avoiding unnecessary work and repeating activities
- quick return on investment

Conclusion

EPTAS is a modular and configurable system which is suitable for all kinds of translation situations. This applies for the single translator who can create his custom made system of translation modules by choosing the optimal subset of tools he needs. But the system also supports bigger organisational units which keep a stock of translated texts to handle the translation process faster, cheaper and more efficiently. It is an ideal tool through its TCP/IP or Internet /Intranet orientation in a client-server-environment. The big advantage with regard to similar systems is the fact that EPTAS can access central databases all over the world. The translation service stores all relevant data at the server and gains and keeps total control about the translation. Projects may be executed customer specific - despite the central database through its supporting an enhanced access right system to the databases. The unintentional use or modification of data of other customers can therefore be prevented. Storing new entries into the database can be controlled and thus suppresses the proliferation of the database by repeatedly including different translations for the same sentences. The central servers allows the easy migration to higher performance by simple adding more processor power or disk storage to the system. Alternatively additional servers can be integrated. The central server is replaced or enhanced and all translators or users integrated into the system take advantage from that.