

An Integrated Solution: Applying PROMT Machine Translation Technology, Terminology Mining, and the TRADOS TWB Translation Memory to SAP Content Translation

Dr. Ulrich Boehme, Svetlana Svetova

Paper type: U
Dr. Ulrich Boehme
SAP AG
Neurottstrasse 16, 69190 Walldorf
Germany
ulrich.boehme@sap.com

Svetlana Svetova
PROMT
Vyborgskaya emb., 61, St. Petersburg,
Russia
svetova@promt.spb.su

Abstract

This paper describes the experiences of SAP and PROMT specialists with applying the PROMT English-Russian machine translation system, the PROMT Terminology Manager Tool for automatic terminology extraction, and the TRADOS TWB translation memory system to the automated process of translation of SAP content from English into Russian.

Keywords

English-Russian machine translation system, automatic terminology extraction, automated translation process.

Introduction

SAP AG is one of a few German companies that are investing in Russian translation and localization.

SAP started translating its products into Russian 10 years ago and is now able to provide the Russian-speaking countries in the CIS with 15 software products as well as documentation and training courses.

As a result of translation, SAP also provides its customers with extensive Russian terminology: about 35,000 entries in the SAP terminology database STERM and more than 50,000 entries in SAP's electronic dictionary, provided by PROMT.

The volume of documentation and training courses translated into Russian so far is about 8 million words.

If SAP's Russian market continues to grow steadily, we can expect to translate an even larger volume of knowledge content.

Knowledge content includes SAP's implementation guide, SAP release notes, selected SAP industry solutions, and mySAP.com documentation. These can be accessed via the SAP Help Portal.

It is very difficult at the moment to forecast the exact volume of words or lines not yet translated into Russian.

It is also clear that SAP Language Services' (SLS) budget for translation of the above-mentioned text types will increase dramatically. Therefore, SLS has been forced to

look for efficient technology-supported translation into Russian to reduce time and costs of future translation projects.

>From the very beginning, SLS was looking for an integrated solution that combined translation memory technology (TM), machine translation (MT) technology, automatic terminology extraction technology, and, if possible, (even simple) text control mechanisms for the source texts to be translated.

The first step towards reducing time and costs for translation was implementing translation memory technology, namely TRADOS TWB, in 1996.

However, the unexpected increase in entirely new texts (*No matches*) has forced SLS to look for a commercial MT system (English to Russian) that

- Runs on Pentium PC with MS NT/Windows and MS Office
- Has development potential and is maintained by the software developer
- Is easy to use

After successfully implementing TRADOS translation memory technology, SLS evaluated several English to Russian machine translation systems.

The requirements were as follows:

- Provide steady results from machine translation while processing large volumes of documents with acceptable quality
- Allow easy integration with TM systems
- Provide easy-to-use, effective tools to extract, process, and manage terminology, and also to store

extracted terminology in the user dictionaries for machine translation

- Reduce costs by 50%
- Improve productivity, translation, and terminology management dramatically (30%)

As a result of this evaluation, SLS chose the PROMT98 machine translation system (developed by the Russian company PROMT) for three main reasons: (i) acceptable, higher linguistic quality of translation into Russian, (ii) the special tool for integration with the TRADOS Translators' Workbench, developed by the PROMT company, and (iii) development of an additional tool for automatic terminology extraction.

SLS purchased the system and tested it with SAP content translation. The results were quite encouraging, especially after intensive updating of the PROMT user dictionary with SAP-specific terminology. However, adding all the terminology provided by SAP's central terminology database (SAPTerm) – about 35,000 Russian entries – to the PROMT user dictionaries revealed the need for additional terminology. SLS asked PROMT to add extended functions to the existing terminology extraction tool, which could be easily integrated with both the TM and MT systems already in use.

Today, because of close cooperation with PROMT, SLS has a fully integrated solution that efficiently combines the following translation tools:

- PROMT Terminology Manager (PROMT TerM)
- PROMT machine translation system
- PROMT P4T and TRADOS TWB

Since 2000, this integrated solution has been successfully deployed in the translation process in the PROMT Translation Center. This automatically translates SAP content in cooperation with SLS in Walldorf and Moscow.

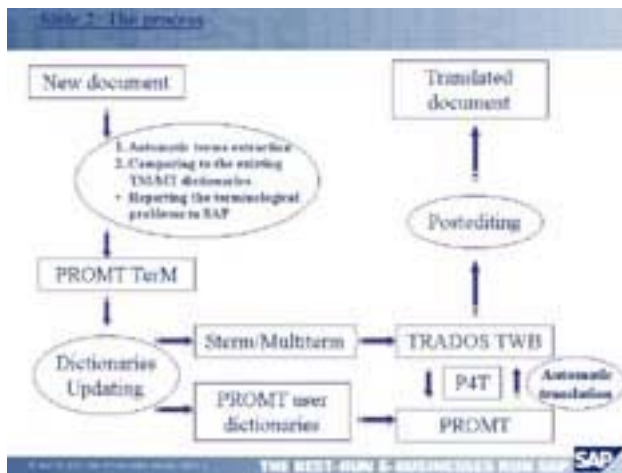
(Slide 1)



Process Description

This automated translation process involves the following stages.

(Slide 2)



1. Preparation for automated translation with the PROMT Terminology Manager (PROMT TerM)

(Slide 3)



As it is well known, the quality of machine translation output is primarily influenced by the comprehensiveness of its dictionaries.

Therefore, checking and updating of pre-translation is essential to ensure acceptable MT quality and reduce the time required for post-editing.

During the first stage of automated translation, the dictionary developer responsible for SAP TM and MT dictionaries has the following tasks:

- Process the texts (to be translated in TRADOS TWB and PROMT) with PROMT TerM to:
 - Extract terminology candidates
 - Compare them with the existing SAP-specific terminological dictionary for TRADOS TWB (Multiterm) and PROMT user dictionaries (at the moment there are 30 user dictionaries for the PROMT English-Russian system containing over

50,000 entries from the different SAP R/3 applications)

- Report any terms missing from the dictionary to the translators and terminologists responsible at SAP
- Together with the application translator:
 - Define translations for the extracted terminological candidates, if they are not in the dictionaries
 - Report terminological problems to SAP
 - Update the dictionaries before automated translation

To carry out this task, the dictionary developer responsible uses the new automatic terminology extraction tool – PROMT Terminology Manager (PROMT TerM). This tool leads to:

- Reduced manual terminology creation
- Increased consistency of terminology
- Management of newly created terminology (by integration with the PROMT machine translation and TRADOS TWB translation memory systems)

Processing the selected set of files in PROMT TerM gives the dictionary developer a list with the following information:

- Terminology candidates (extracted statistically): words or word combinations
- Their translations from the chosen PROMT machine translation user dictionaries
- Or, if no translation exists in the dictionaries, the result of machine translation
- Their translation from the Multiterm terminological dictionary for TRADOS TWB
- TM segments where terminology candidates were found for the chosen language pair

This enables the dictionary developer not only to get a list of terms missing from the dictionary and prepare the dictionaries, but also to compare and manage the terminology stored in the terminology dictionaries for the TM and MT systems.

2. Automated translation in TRADOS TWB and PROMT MT integrated with the PROMT for TRADOS (P4T) program

The next stage is automated translation using the customized TM and MT systems.

First, the text is translated in TRADOS TWB, then the unknown segments (*No matches*) are sent to the PROMT system for machine translation.

To facilitate this procedure, PROMT developed a special integration tool (PROMT for TRADOS – P4T).

This easy-to-use tool enables users to carry out the following in one environment:

- Send the files to the chosen TM
- Get the results of TM translation and check them before machine translation (optional)

- Automatically have all the unknown segments translated in PROMT with the attached dictionaries (updated during the first stage) and inserted into the TM for future processing

The application translator then gets the text to be edited in TRADOS TWB format.

Here are some interesting points about customizing the PROMT machine translation system:

- Before the translation project started, Russian terminologists in SLS entered over 30,000 entries in 30 PROMT user dictionaries in different applications (such as Logistics and Material Management) in six months
- Large amounts of terminology have been extracted and entered in the user dictionaries during the translation process with the help of the PROMT Terminology Manager (over 10,000 entries)
- PROMT user dictionaries can be easily converted into PROMT Electronic Dictionary format, providing an additional resource
- All machine translation problems and errors are reported to the PROMT developers so they can correct existing algorithms and create new ones

Post-editing

At this stage

- Application translators should:
 - Edit the results of machine translation
 - Store the translation in TM format
 - Provide reports on improvements to the dictionaries
 - Dictionary developers should:
 - Update the dictionaries using the translators' reports
 - Update the dictionary database. This has the following elements:
- (Slide 4)

SAP-specific dictionary database

SAP user dictionaries for the PROMT machine translation system form the basis for:

- SAP electronic dictionary (the PROMT system component used as an additional lexicon by application translators)
- SAP templates (sets of user dictionaries and user settings chosen in the PROMT machine translation system while translating the different applications)
- PROMT TerM processing (to compare with the lists of terminology candidates)

Results of applying this technology and future activities

Results:

- Increased translation speed (30%)
- Reduced time and costs (40%) and improved productivity
- More consistent and accurate terminology stored in the user dictionaries – the common translation resource while using the machine translation system
- More consistent terminology in SAPTerm (for example, reduced ambiguity)
- 30 SAP-specific user dictionaries containing about 50,000 entries
- Dramatically improved quality of machine translation in the given domains due to close cooperation between SLS experts and PROMT developers
- SAP electronic dictionary – a new PROMT Translation Office component that displays all the terminology in the SAP user dictionary in a pop-up mode. It can be the common resource for translators and customers
- Consistent style in documentation

Goals for future development

These include providing solutions for SAP customers in Russia that meet the high demand for information about new SAP products and technology. These solutions will be based on current dictionaries and automated translation technology.

The customers can use PROMT tools, SAP dictionaries, and special PROMT Internet solutions to access SAPNet. SAP's Web site has a great deal of information that customers can access quickly and in their chosen language, thanks to the translation tools and dictionaries.

