

Bridging the Gap – EuroTermBank Terminology Delivered to Users’ Environment

Tatiana Gornostay

Tilde, Riga, Latvia

tatiana.gornostay@tilde.lv

Signe Rirdance

signe.rirdance@gmail.com

Andrejs Vasiljevs

Tilde, Riga, Latvia

andrejs.vasiljevs@tilde.lv

Roberts Rozis

Tilde, Riga, Latvia

roberts.rozis@tilde.lv

Abstract

The paper presents EuroTermBank Terminology Add-in – an innovative multilingual terminology translation tool. This tool integrates the federated terminology of the currently over 100 EuroTermBank terminology resources in Microsoft Word environment and accesses over 2 million terminology entries in 27 European languages. The new tool provides a range of functions necessary and useful for target user groups improving productivity and effectiveness in the process of translation. User feedback shows definite interest in the new tool which is considered to be extremely useful and efficient for translation needs. Beta versions of EuroTermBank Terminology Add-in for Microsoft Word 2003 and 2007 can be freely downloaded from the EuroTermBank multilingual terminology portal.

1 Introduction

Typically, translators spend 30-60% of total translation time on terminology research. Therefore, it is of vital importance to ensure that they can use all the required terminology resources in the right format and in the right environment. Increasingly, terminology research is done using sources that are available on the Internet. Currently, translators spend a lot of time inefficiently, searching and processing information from multiple online sources, copy-pasting or changing the format to the one that they require in their work environment. Spending time on technical aspects instead of focusing on true terminology

research results in cost inefficiencies and reduced translation quality.

Faced with difficulties in accessing the terms they need and participating in collaborative activities for creation of new terms, many translators create their own terminology resources. They typically store these terms in a spreadsheet or other proprietary formats that are not efficiently connected to a multitude of translation environments that they might use. Moreover, these resources are not shared with other translators and potential users. This results in redundant work or even reduced translation quality and does not bring additional value to the creator of this custom terminology.

To increase efficiency and quality of translation, translators need an easy access to multiple terminology databases, facilities to enable collaborative efforts in creation of new terms, productivity tools to get necessary terms right from translation environment (Lengyel and Vasiljevs, 2008).

There have been several efforts to provide reasonable solutions to support translators accessing multilingual terminology resources. For example, Quest tool brings consolidated terminology content closer to its user and is used internally by translators in the DG for Translation of the European Commission (European Commission Directorate-General for Translation, 2008).

Other major consolidation effort of multilingual terminology is public multilingual terminology resources, such as EuroTermBank, database of federated terminology. It is targeted to a broad audience of translators, terminologists and other language workers.

EuroTermBank is a rich source of consolidated public terminology that provides access to content that is stored in its central database as well as a number of interlinked term bases in a federation model (Auksoariute et al., 2006). It

provides access to terminology resources in 27 European languages and has particularly extensive coverage of Latvian, Lithuanian, Estonian, Polish, Hungarian terminology. Federation model enables EuroTermBank to mediate content from interlinked portals, such as IATE, the richest source of EU terminology.

EuroTermBank web portal includes all necessary features to provide single access point to multilingual terminology resources – web-based browsing, searching and information access features. A survey of users was carried out in 5 countries with 174 participants (Blaudums et al. 2006). It showed that 69% respondents consider finding the required term to be very simple and 85% liked the appearance and user interface. At the same time there is still an extra effort required from the user to switch from translation environment to terminology web-page, specify the search query, select a result and go back to translation tool and type the term there.

New approaches should support the most widely used translation environments, such as SDL Trados, Deja Vu, Wordfast, MemoQ, as well as other applications that are commonly used in the translation process, such as Microsoft Word, PowerPoint, Excel as well as open-source applications.

The developed tool, presented in the paper, integrates terminology search features in Microsoft Word environment. By invoking web-service based queries it provides instant access from the text editing environment to the EuroTermBank web-based terminology data.

The structure of this paper is as follows: this introduction gives an overall description of the issue, the EuroTermBank Overview section briefly introduces EuroTermBank as a multilingual terminology resource which the newly developed tool gives access to, the subsequent sections define user scenarios and target user groups needs, focus on the tool design and functionality, and provide analysis of user feedback. The concluding section outlines future directions for research and practical applications.

2 EuroTermBank Overview

EuroTermBank² is a centralized online terminology database for languages of new EU member countries interlinked to other terminology resources (Auksoariute et al., 2006; Vasiljevs and Rirdance, 2007). Currently, EuroTermBank pro-

vides federated access to 5 interlinked external term banks, the major of them being IATE, the interinstitutional terminology database of the EU (Rummel and Ball, 2001). EuroTermBank serves basic terminology needs of a user by providing a single access point to distributed term resources and implementing query schemes suitable for particular usage scenarios.

The specific functions of the EuroTermBank portal include user authentication, term search, data editing, administration, user feedback, and communication facilities with external databases as well as data import and export (Auksoariute et al., 2006).

API based access for third party software enables software manufactures to provide their users with direct access to the content of EuroTermBank internal database. This is especially useful in the translation usage scenario since such a solution will deliver well-targeted content from EuroTermBank to productivity environments used routinely by translators and other language workers.

EuroTermBank Add-in is a tool to better exploit potential of this major online source of multilingual terminology resource infrastructure.

3 Usage Scenarios

Target clients of the new tool can be segmented as follows:

- translation service providers:
 - freelance translators (and other individual users: editors, technical writers, etc.),
 - translation agencies,
 - localization service providers;
- translation service consumers (using outsourced and / or in-house services):
 - commercial companies with products / services in global markets,
 - international organizations,
 - EU institutions,
 - national government institutions;
- minor client groups:
 - various organizations: term banks, libraries, publishing houses, universities, providers of web-based CAT tools etc.,
 - various individuals: students, academia, “general reference” users, etc..

² www.eurotermbank.com

Nevertheless, freelance translators and in-house translators are foreseen to be major target user groups for the new tool.

A typical portrait of a freelance translator is as follows:

- produces translations for various clients;
- works from a home office;
- may use a CAT tool like Trados or Wordfast, but not necessarily;
- usually works from 1-2-3 source languages into 1-2 target languages;
- often specializes in a certain subject area;
- price elasticity is high - has to finance IT purchases from own budget;
- IT literacy varies from very high to moderate and low.

A typical portrait of an in-house translator is as follows:

- produces translations for a certain employer which is often a translation agency / localization service provider / government institution / commercial company;
- may use a CAT tool like Trados or Wordfast, but not necessarily;
- usually works from 1-2-3 source languages into 1-2 target languages;
- often specializes in a certain subject area;
- is less price-sensitive, as IT purchases are done by employer;
- IT literacy varies from very high to moderate and, in rare cases, low.

3.1 User Needs Survey

To determine target user groups' expectations regarding the new tool and use of EuroTermBank resources beyond the portal, a survey of user needs was carried out. A questionnaire was developed and sent via e-mail to 80 translators. Among questions were inquiries about the platform, software, Internet resources and search engines used in translation process as well as inquiries about workflow organization and terminology management.

The idea was also introduced during the Estonian Terminology Society Board Meeting and Tallinn University Terminology Conference, asking experts to describe an "ideal" terminology tool.

3.2 User Needs Analysis

In fact, the survey results show that translators and terminologists have difficulty envisioning computer software tools, and they prefer using existing tools they have got used to already. The quantitative results of the survey are as follows:

- 80% respondents use Microsoft Word in their everyday workflow, 25% use SDL Trados in conjunction with Microsoft Word, the rest use free tools like OpenOffice or Globalsight;
- 80% respondents do not manage terminology, the rest use MultiTerm or SDL Trados;
- 40% respondents process files in Microsoft Word format, 25% – in html, 15% – in pdf and 20% in a mix of other formats (e.g. Excel or PowerPoint formats).

Furthermore, about 90% respondents use Google for terminology research. Nevertheless, the survey results show users' interest and necessity for additional terminology tools especially for Microsoft Office (with high priority for supporting Microsoft Word, lower – for Microsoft PowerPoint and OpenOffice Writer, and the lowest – for Microsoft Excel and e-mail writing tools). Therefore, the biggest potential for meeting users' needs is in supporting Microsoft Office. Besides, Microsoft Word integrates with SDL Trados and thus bridges the gap to the user of CAT tools. Both Microsoft Word 2003 and 2007 should be supported since the first is still popular among many users and the second offers broader functionality.

The goal is to provide an access to EuroTermBank content with a single keyboard shortcut, even without opening a browser window.

4 Design and Functionality

Conceptual design of the developed tool comprises the following components (Figure 1):

- EuroTermBank engine that provides users with the following functions (see the description below):
 - search,
 - segment analysis,
 - entry request;
- internal EuroTermBank terminology database;
- external EuroTermBank terminology databases.

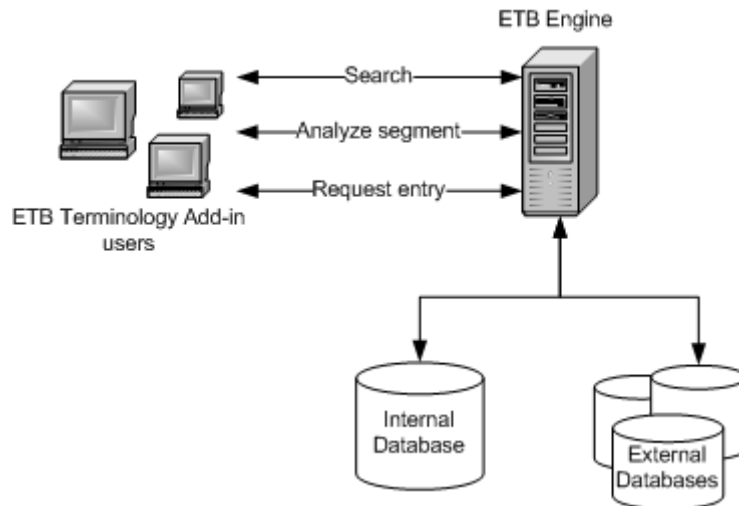


Figure 1. Design of EuroTermBank Terminology Add-in.

The EuroTermBank Terminology Add-in meets the following requirements:

- easy download, quick setup, low usage of computer resources;
- integration into Word Research pane and compact / clear arrangement of terms in it;
- intuitive use of the tool and no hidden or complicated features, keyboard shortcuts.

The EuroTermBank Terminology Add-in integrates in Microsoft Word as follows³:

- a button “EuroTermBank” in Microsoft Word in the Review Ribbon, Terminology Group (Figure 2);
- a custom Microsoft Research pane “EuroTermBank Terminology” (Figure 3);
- contextual menu when right-clicking a text segment (Figure 4).

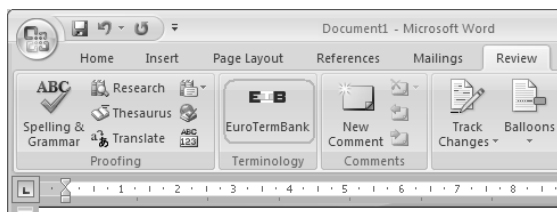


Figure 2. EuroTermBank button.

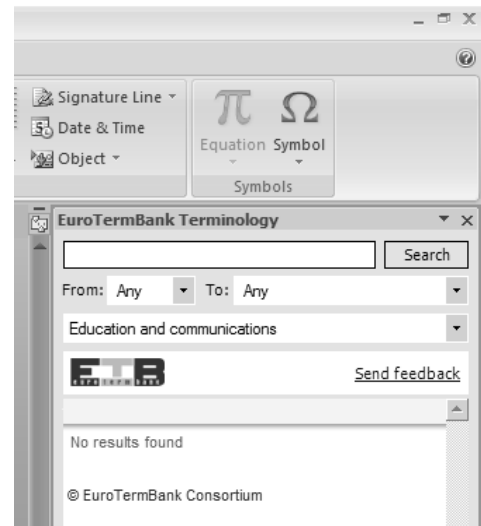
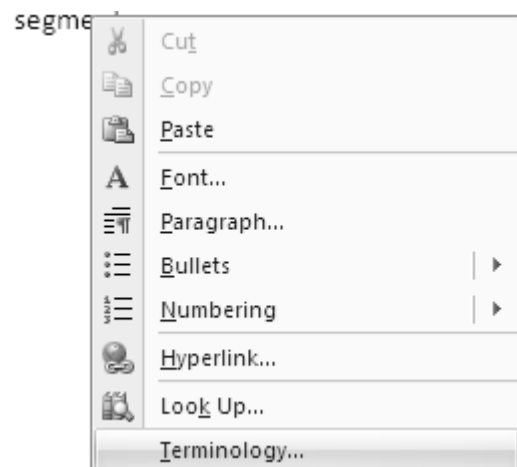


Figure 3. EuroTermBank pane.



³ For demonstration purposes, add-in for Microsoft Word 2007 is described in this paper.

Figure 4. Terminology contextual menu.

The EuroTermBank Terminology Add-in provides the following functionality:

- provides targeted search results in Word Research pane;
- filters terminology by domain;
- filters terminology by language;
- automatically detects source language;
- identifies terms in a segment / sentence and researches the EuroTermBank internal and external resources for the identified terms;

- provides the option of reaching full search results by opening the EuroTermBank portal in a web browser;
- provides user feedback function.

It should be mentioned that the function of identifying terms in a segment or sentence (Figure 5) and then searching the EuroTermBank resources for them is highly appreciated by end users (see Section 5 for user feedback). The tool identifies terms and shows them hyperlinked in the topmost part of the pane. Moreover, the user can change the language and domain settings, and the tool updates the relevant links in specified languages or domains.

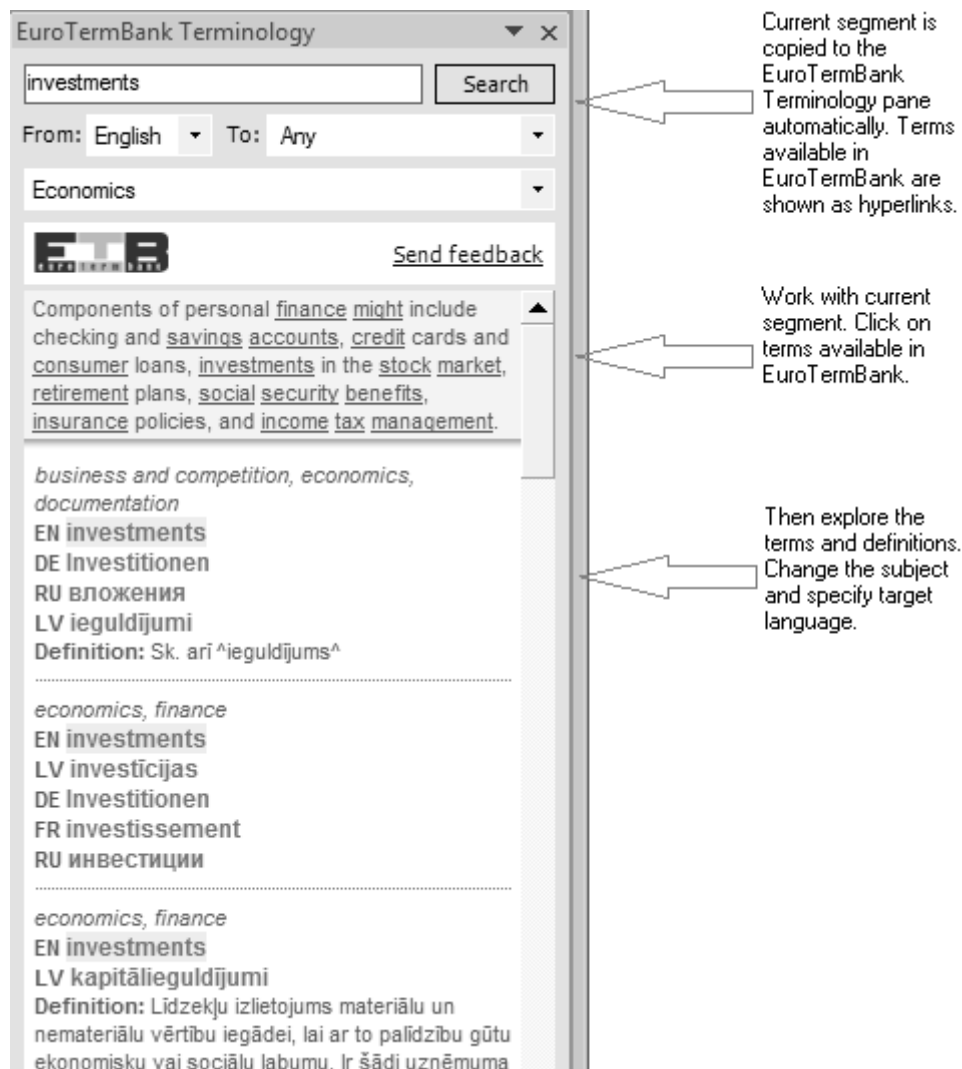


Figure 5. Identifying terms in a sentence.

Beta versions of EuroTermBank Terminology Add-in for Microsoft Word 2003 and

2007 can be freely downloaded from the Eu-

roTermBank multilingual terminology portal⁴.

5 User Feedback

The developed tool was tested and evaluated by end users before its release (internal beta testing) and after it (external testing). The following two subsections provide the description of user feedback.

5.1 Internal Beta Testing

The developed tool was tested by localization specialists during two month before release. 20 full-time and about 50 part-time professional translators were involved in beta testing. To receive feedback from terminologists, they were asked to fill in a questionnaire. Altogether 80 out of 148 sent questionnaires were filled in.

General results of the internal beta testing show that:

- 30% respondents (more than a half) consider the developed tool to be extremely useful and efficient for their translation needs;
- at the same time 40% respondents define the tool as good software which could be of help for their work;
- 70% respondents point out that an access to the terms is simple and fast (contextual menu, in particular);
- almost all respondents appreciate language coverage.

The main suggestions are as follows:

- 10% respondents suggest that instructions about how to use the tool should be built in as an internal function⁵;
- 20% respondents find that they would like to have localized web page and instructions;
- 89% respondents point out that the amount of displayed search results is too large and same results appear to be from different sources.

5.2 External Feedback

After the release of beta version we also got users' feedback. Among advantages users em-

phasize user-friendly interface and segment / sentence analysis function.

Users point out the following functionality to be developed:

- built-in help;
- Macintosh support;
- short keys change;
- users' settings saving (e.g. translation direction).

Users also suggest adding new languages, for example, Persian.

Received requests and suggestions will be taken into account in future work.

6 Future Work

Future work is oriented towards further developments and enhancements of translation environments. In particular, it is crucial to provide translators with fully-functional model of terminology research based not only on dictionaries and databases, but context solutions as well.

One of the acknowledged research directions focuses on corpus-based technologies and comparable corpora, in particular. The FP7 TTC project (Terminology Extraction, Translation Tools and Comparable Corpora)⁶, started recently, aims at leveraging machine translation tools, CAT tools and multilingual content management tools by automatically generating bilingual terminologies from comparable corpora in several European languages (English, French, German, Spanish and Latvian), as well as in Chinese and Russian.

The major TTC project objective is to develop a web-based platform for handling comparable corpora, while relying on existing technologies (UIMA⁷ and EuroTermBank, in particular) and standards (TMF, TBX). The platform will allow users to create thematic corpora given some clues (such as terms or documents on a specific domain), to extract monolingual terminologies from such corpora, to create a comparable corpus in a target language from a corpus in a source language, to align bilingual terminologies, to choose the tools to apply for terminology extraction, to expand a given corpus and to export monolingual or bilingual terminologies in order to use them easily in automatic and semi-automatic

⁴http://www.eurotermbank.com/products_download.aspx

⁵ During testing users followed the instructions on the EuroTermBank portal: http://www.eurotermbank.com/products_howitworks.aspx

⁶ The project has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 248005.

⁷ Unstructured Information Management Architecture

translation tools (machine translation systems and CAT tools).

A fine-grained user requirement analysis is being carried out currently to meet users' needs. The new solution will help translators to overcome difficulties they face and deliver terminology they require into their working environments.

7 Conclusion

So far and with few exceptions, the richness of terminology resources on the Internet has not directly translated into the expected increased productivity and quality levels of translation work. The area of providing tools for integration of terminological data in translation environments (including machine translation systems and CAT tools) is relatively new.

A new layer of tools and technologies that integrate translation environments with terminological resources on the Internet is required to significantly enhance the current productivity of human translation.

Acknowledgement

Work presented in this paper was partly funded by EU EUREKA Programme and European Social Fund. We would like to thank Raivis Skadiņš, Valters Šics, Andis Lagzdīņš from Tilde for their valuable contribution in development of EuroTermBank Terminology Add-in.

References

- Auksoriute, Albina [et al.]. 2006. *Towards Consolidation of European Terminology Resources : Experience and Recommendations from EuroTermBank Project*, Riga.
- Blaudums, Ainars [et al.]. 2006. *EuroTermBank Revised user needs consolidation : Requirements specification report*, EurotermBank Consortium.
- European Commission Directorate-General for Translation, 2008. *Translation tools and workflow*. Luxembourg, Office for Official Publications of the European Communities.
- Lengyel, Istvan and Andrejs Vasiljevs. 2008. *How to get the right terms to the right people – terminology sharing and integration in translation environments*. TCWorld Conference, Wiesbaden, November 2008.
- Rummel, D., S. Ball. 2001. *The IATE Project – Towards a Single Terminology Database for the EU*. The 23rd International Conference on Translation and the Computer, ASLIB 2001, London.
- Vasiljevs, Andrejs and Signe Rirdance. 2007. *Consolidation and unification of dispersed multilingual terminology data*. International Conference Recent Advances in Natural Language Processing, Borovets, Bulgaria.