eLexicography, collaboration, bilingual corpora - all in one!

Vera Kuzmina

ABBYY

127273, Otradnaya 2B, Moscow, Russia

E-mail: vera_k@abbyy.com

Abstract

While working on a dictionary or glossary, a lexicographer faces the problem of choosing the right resource and convenient tools for finding usage examples, checking the meaning of words, and their translation into other languages. According to Atkins, B.T.S. & Rundell, M., it is a standard for the lexicographer today to use different types of corpora and language data material for these purposes. Each of them is stored in different databases with several access points, sometimes the query language is very sophisticated and lexicographers need additional training to start using these sources. In my paper I would like to draw your attention to the online dictionary and corpora website called ABBYY Lingvo.pro. The idea behind the website was to provide a single access point for the professional user and to meet his multiple needs, including lookup in various dictionary sources, searches in bilingual and monolingual corpora, collaboration with dictionary users, easy search, and other nice features for dictionary makers.

Keywords: bilingual corpora; single access point for content; term extraction tool; collaborative lexicography

Software demonstration

While working on a dictionary or glossary, a lexicographer faces the problem of choosing the right resource and convenient tools for finding usage examples, checking the meaning of words, and their translation into other languages. According to Atkins, B.T.S. & Rundell, M., it is a standard for the lexicographer today to use different types of corpora and language data material for these purposes. Each of them is stored in different databases with several access points, sometimes the query language is very sophisticated and lexicographers need additional training to start using these sources. In my paper I would like to draw your attention to the online dictionary and corpora website called ABBYY Lingvo.pro. The idea behind the website was to provide a single access point for the professional user and to meet his multiple needs, including lookup in various dictionary sources, searches in bilingual and monolingual corpora, collaboration with dictionary users, easy search, and other nice features for dictionary makers.

In general, Lingvo.pro is an online dictionary (see figure 1) which aggregates different types of content and convenient lookup and search tools, including high-quality morphology support. On the other hand, it also accumulates a large amount of bilingual parallel texts for different language pairs, for instance English, Russian, German, French, with usage examples showing words and phrases as they occur in real-life texts. This means that the user will automatically see appropriate bilingual parallel examples for each sense of the word and its translation. For instance, when looking up the word "table" in the sense "spreadsheet", the corpus will provide sentence examples and translations only for this sense (e.g. "Tabelle" in German, see figure 2).

The above usage scenario helps lexicographers avoid multiple steps, starting from the lookup of the word in different dictionaries and making queries to a corpus. In Lingvo.pro, all these steps are merged into one quick lookup. The lexicographer may extract terms with the help of an online Terms Extraction tool and check the terms and their translations directly in this list. This list of terms can be used as a draft for the new dictionary or as an additional source to check possible translations while working on word senses. This simplicity is the main advantage for both professional lexicographers and those who only start their careers in dictionary creation. At the same time, there are also much more sophisticated tools available on Lingvo.pro.
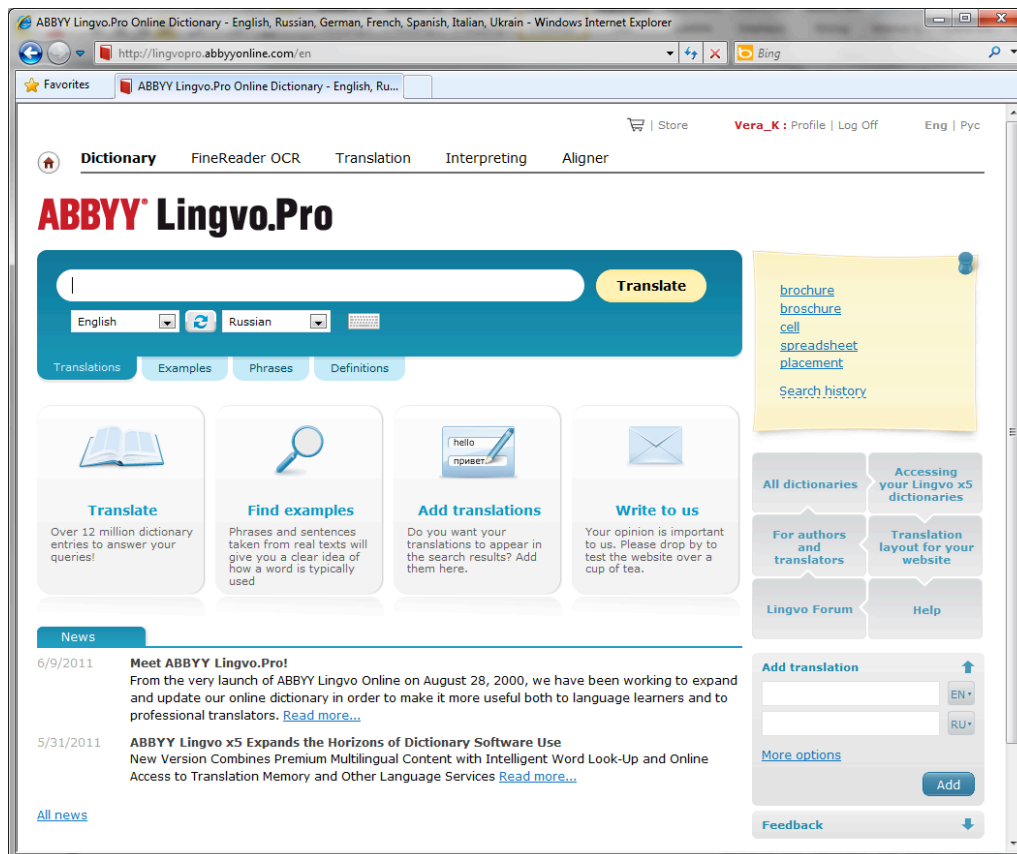
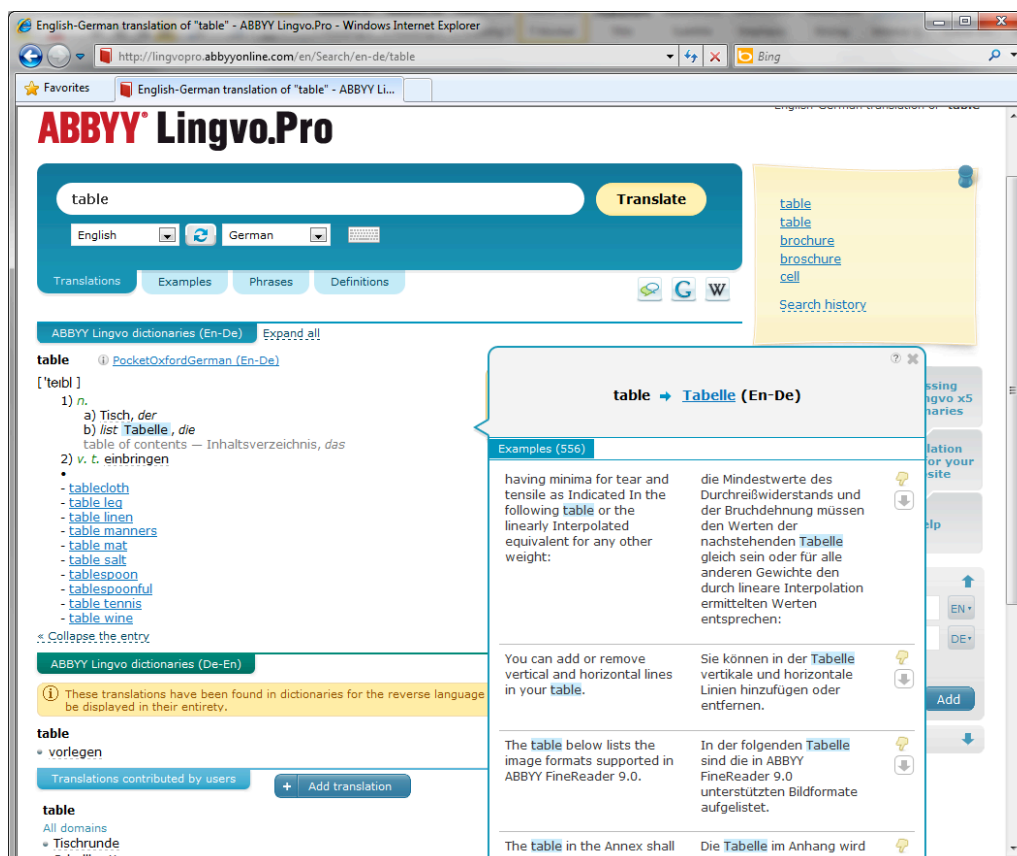Figure 1. ABBYY Lingvo.pro General view

Figure 2. ABBYY Lingvo.pro Translation example

One of the useful tools is the advanced search tool, a professional, yet easy to use, corpus query system. Firstly, there is a bilingual search feature (see figure 3) with the possibility to find a word in one language and its translations into another language. Users can make queries for words in different senses and search for words and phrases in different positions. Then, there is a "do not" option, a match word order option, ability to find inflected forms, case-sensitive searches, and many other things which may make the life of dictionary maker easier. The user interface of the website and the corpus query system within are user-friendly and easy to use.

All the text corpora accessible on Lingvo.pro are morphologically and syntactically tagged so that even disambiguation becomes possible. For dictionary makers, such precise and sophisticated approach to word senses means that their everyday language investigations become simpler and the results achieved are significantly better. For example, an author may list the meanings of the word according to their usage frequency. With Lingvo.pro, a lexicographer may see the frequency of a word or sense in real texts. This is a very important and unique feature of bilingual text collections. Frequency is automatically shown next to each retrieved word sense. This makes it really easy for dictionary makers to investigate usage frequencies, which are now only a couple of mouse clicks away.
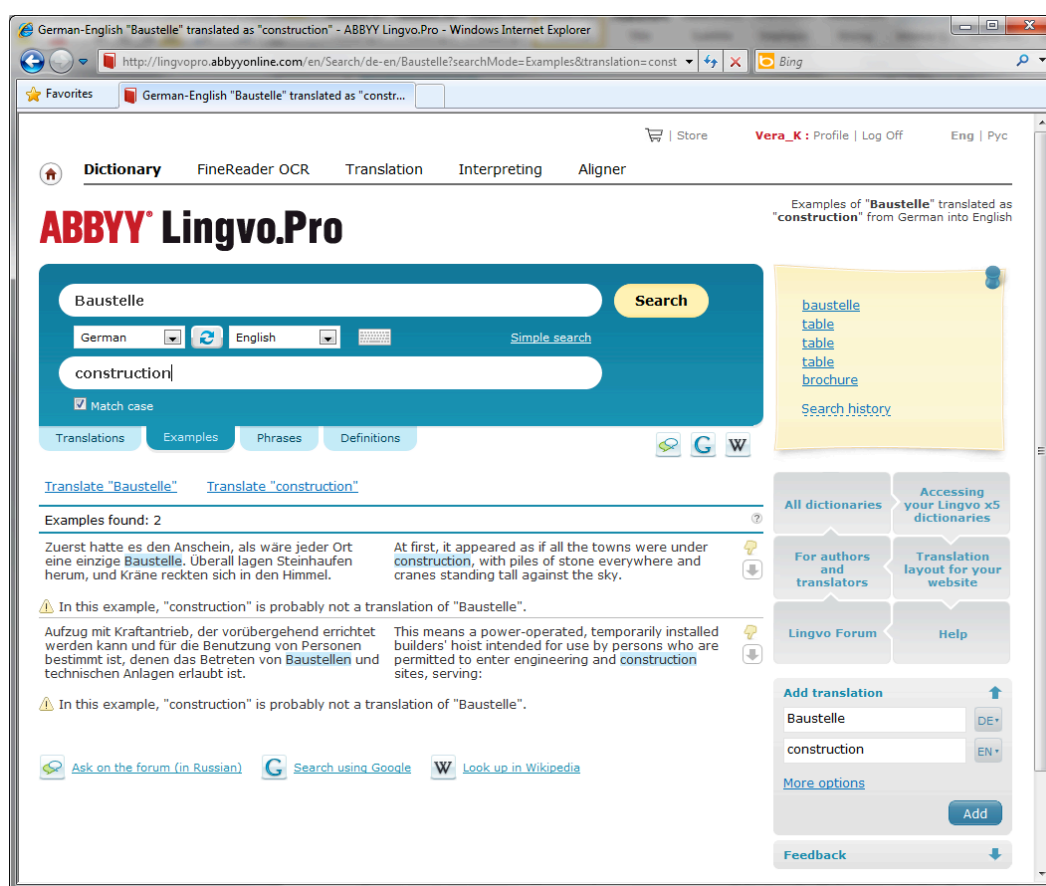

Figure 3. ABBYY Lingvo.pro Bilingual search

Another important feature is the ability to create glossaries and add new words into glossaries and dictionaries available on the website (see figure 4). Members of the community can work together, adding new words, senses, and translations. Convenient

collaboration tools are available: users can propose brand-new words to be included into the dictionary and discuss the new proposals with other users or with the editors and lexicographers involved in the project.
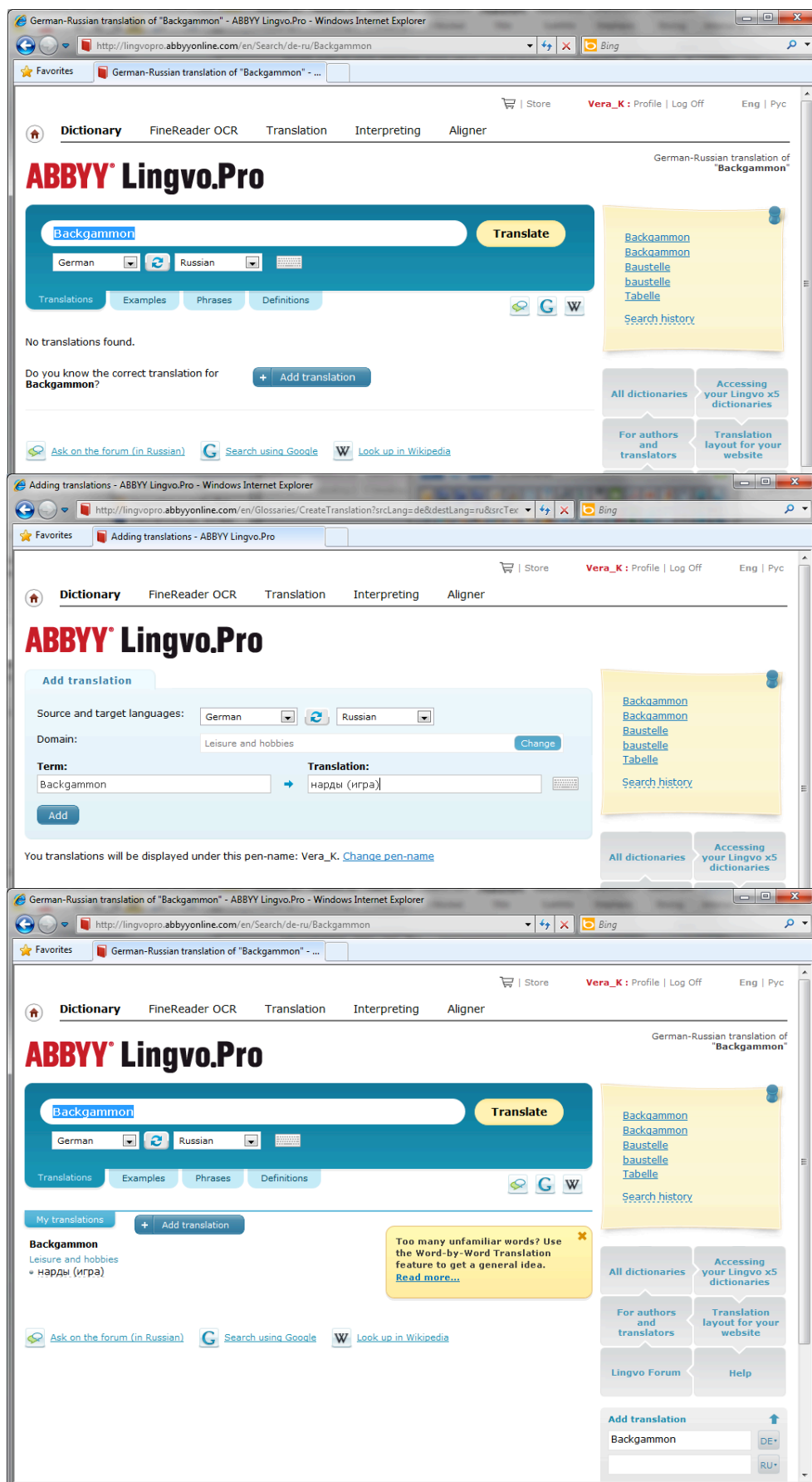
Figure 4. ABBYY Lingvo.pro Adding new terms

In conclusion, I would like to point out that dictionary creators today are sometimes overloaded with different sources and software tools which enable access to these sources. There is demand for one simple solution which will meet all the lexicographic needs at once and will be easily accessible from any computer even outside the office, because most dictionary creators are teleworkers. There is also a growing trend to make complicated things simpler, which is exactly the purpose of Lingvo.pro.

References

Atkins, B.T.S. & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

Hockey, S.M. (2000). Dictionaries and lexical databases. *In Electronic Texts in the Humanities: Principles and Practice.* Oxford & New York: Oxford University Press, pp. 146-171.

http://www.abbyy.com/

http://www.lingvo.com/

Kuzmina V., Rylova A. (2009) *The ABBYY Lingvo electronic dictionary and the ABBYY Lingvo Content dictionary writing system as lexicographic tools.* In S. Granger, M. Paquot (eds*.) eLexicography in the 21st century: New challenges, new applications, Proceedings of eLex 2009, Louvain-la-Neuve, 22-24 October 2009*. Louvain-la-Neuve: Presses Universitaires de Louvain.