

Jolanta Zabarskaitė

Lithuanian lexicography in the 21st century: quo vadis?

1. A few facts about the Lithuanian lexicography

The Lithuanian lexicographical tradition dates back to the 16th century. Authored by Konstantinas Sirvydas, a professor at Vilnius University, the first dictionary was published circa the year 1620. It was a trilingual Polish–Lithuanian–Latin dictionary entitled *Dictionarium Trium Linguarum* (Pakalka 1997).

As far as Lithuanian philology is concerned, the voluminous thesaurus called *The Dictionary of the Lithuanian Language* (further LKŽ) is of particular relevance; back in 1930, a national Dictionary Desk was set up to compile the dictionary. This desk has eventually evolved into the Institute of the Lithuanian Language, which operates Lithuania's only Centre of Lexicography. The Centre of Lexicography specialises in theoretical aspects of lexicology and lexicography, the lexicon and semantics of Lithuanian, monolingual and multilingual lexicography, electronic lexicography and bilingual dictionaries of lesser used languages. The Centre of Lexicography also deals with the prospects of future dictionaries and has established three new pillars (see Zabarskaitė 2009, 135-147).

2. The first pillar: a new live for classical lexicographic resources

The biggest work in Lithuanian linguistics of the 20th century is *The Dictionary of the Lithuanian Language* (LKŽ I-XX/1941-2002). The final 20th volume of this was released in 2002, completing the work of several generations of linguists. The Dictionary serves as a source of national identity and linguistic investigation of the development of written and spoken Lithuanian and targets the world community working in the humanities and society at large. The scope of the Dictionary is ca. 22,000 pages, 0.5 million lexicographic entries, and 11 million words. It gives access to archaic, dialectal and contemporary layers of the lexicon of Lithuanian. Illustrations of word meanings have been taken from various sources covering the period from 1547 to 2001: research and religious texts, fiction, folklore and dialects recorded in the last century, starting from 1902. The Dictionary is based on a card index consisting of 4.5 million items collected from almost 1,000 lexicographic sources (both hand-written and printed); dialectal words have been recorded in more than 500 Lithuanian settlements (more: Kažukauskaitė 2002, 29-33; Klaviņš 2002; Топоров 2004, 408-415; Забарскайте/Нактинене/Шяптитте 2005, 340-355; Zabarskaitė/Naktinienė 2007, 32-34).

IT specialists together with lexicographers had to face the issues of the sophisticated Dictionary structure (more: Zabarskaitė/Naktinienė 2004, 187-190; Zabarskaitė 2007, 261-272), the variety of characters and symbols, and other problems (for information on the technical solutions for creation of electronic version further refer to Zinkevičius 2004, 170-173; Zinkevičius 2007, 349-355).

Researches and doctoral students all over the world use the academic edition of *The Dictionary of the Lithuanian Language* (LKŽ) on the website www.lkz.lt. Yet rapidly evolving digital technology requires new solutions for the application and distribution of LKŽ (as well as other dictionaries) (see Zabarskaitė/Naktinienė 2010, 780-787).

2.1 Solution no. 1: multi-platform

The technological environment for dictionary applications (and desktop applications in general) has however changed quite a lot in the recent years. While just some time ago, one could get away with a Windows-only dictionary application sold as a CD, nowadays linguists, software developers and managers have to consider the colourful and diverse range of devices that their users hold in their hands. As there is no more a single winner in the operating system or mobile device market, publishers of dictionary applications must “hit” them all: Windows, Mac OS X, Linux based computers, iPhone mobile phones, iPod touch multimedia players, iPad tablets, Android mobile phones, Android tablets, etc. Currently, LKŽ supports three platforms: Windows, Mac OS X and Linux. Dictionary applications for iPhone, iPod, iPad and Android platforms are at the development stage.

2.2 Solution no. 2: software delivery

The approach to software delivery in Compact Discs (CDs) is highly outdated. The CD did not prove its reliability in household environments as it is prone to scratches. Also, newer laptops do not even have a CD-ROM that would be able to read the CD with the dictionary application. The LKŽ application is delivered to the users using two new methods:

- USB flash drives. While the USB flash drive adds some additional cost to the final price of the product (compared to a CD), it has increased longevity and almost all desktop PCs (both new and old) have a USB port. Delivering the software product on a USB flash drive is useful for those clients who like to have a “material thing” (a physical installation medium) with their purchase (as opposed to software deliveries over the Internet, where users do not get a physical installation medium).
- Delivery over the Internet. Nowadays, not all users can (or want to) go and buy a physical installation medium with the dictionary application. In the case of our dictionary, LKŽ, the product is in high demand among the Lithuanian diaspora in various countries (the US, the UK, and Australia, to name a few), and it would be both easier and cheaper if the dictionary users from foreign countries could avoid ordering physical media with the dictionary, and be able to buy and download the dictionary directly, in a matter of minutes, instead. Thus, we argue that a modern dictionary application must have a strategy on how it could be purchased and delivered over the Internet to the end-user.

2.3 Solution no. 3: advanced tools

Even in the age of automatic machine translation one of the main tools for the day-to-day human translator remains the monolingual and the bilingual dictionary application.

The dictionary application is a highly important and useful tool for the amateur translator too, who just wants to find out how to say “pancakes and coffee, please” in French. Additionally, the monolingual and bilingual dictionaries can be used as applications for various languages and have a considerable demand in the general user market (more Zabarskaitė/Naktinienė/Valiukas 2013).

Software available for the Lithuanian language is limited in its functionality. Advanced tools, such as treebanks, WordNet, and so on, are yet to be developed for the Lithuanian language, or are still in the pipeline (Vaišnienė/Zabarskaitė 2012). Even though machine translation tools have been developed recently, it is only just now that the first steps are being taken to design the most advanced of resources – the so-called general applications. Obviously, texts that have a structure as complicated as LKŽ has can be used for the purposes of developing this kind of research. And so, dictionaries outlive their authors.

Conclusion: classical dictionaries should be adapted to latest technology.

3. The second pillar: innovative technology and dictionaries

The 21st century is the century of digital technology and Internet. In contrast to printed dictionaries, online dictionaries and thesauruses offer a complete range of possibilities to use sound and video materials. The array of available tools is practically endless: from images, historical pictures and photographs of works of art to videos, movie clips and animation. All of this not only helps learn a language, but also broadens the understanding of language and culture. If such tools were used to create multilingual dictionaries, or even dictionaries that cover several languages, such multimedia dictionaries could reflect two cultures or more and carry out a two-way function of helping learn a language better and opening the door to a culture a little bit wider. A multimedia Lithuanian-English dialectal dictionary (Lietuvių tarmės 2000-2011) has been developed in Lithuania:

The Lithuanian language, still rich in dialects, is interesting for linguistic studies world-wide as one of the most ancient Indo-European languages, and therefore it must be safeguarded and made accessible for scientific research everywhere. The first attempt in this direction was made by creating the Multimedia Dictionary of Lithuanian Dialects. This interactive CD-ROM demonstrates multimedia dictionary possibilities: a reader can see a picture of an object, to listen to pronunciation of a word in different dialects and to see how it is written in the Lithuanian language. Therefore it can be used without any previous knowledge of a language. (Telksnys/Kligienė 2001, 57-65)

This dictionary won the gold medal at the world EXPO 2000 in Hanover, Germany for utilising multimedia in lexicography and for the dissemination of ethnic culture.

Dictionaries are more than just a tool to learn and use a language. Every dictionary is a cultural text, one that reflects the unique character and values of the culture it represents.

Another very important thing is that multimedia is indispensable when sign language dictionaries are prepared.

Conclusion: In future, lexicography will most probably use an increasing amount of multimedia options both when it comes to designing new dictionaries, and adapting classical dictionaries to the digital media.

4. The third pillar: lexicographical infrastructures

It is becoming apparent that corpuses, dictionaries, catalogues, digests, and databases have to be integrated to create uniform infrastructures that would allow performing searches against a variety of parameters and combine different linguistic resources through common searches. The most important search unit within such infrastructures is a word. Hence, the future in lexicography belongs not to isolated dictionaries, but also to various infrastructures of dictionaries and other infrastructures of linguistic resources, which should cover both classical and modern dictionaries. Currently, Lithuania is implementing a programme called “The Lithuanian Language in Information Society”. The framework of this programme will include the digitalization of many Lithuanian language resources, improvement or development of language technologies or solutions (such as tools for machine translation, syntactic/semantic analysis, etc.) for the Lithuanian language. The Lexicography Centre of the Institute of the Lithuanian Language coordinates the implementation of “Developing and Providing Access to IRT Solutions and Content Geared Towards Preservation of the Lithuanian Language in Public Space”, a project within the aforementioned programme. This project will see the development of a Lithuanian language information system, which will provide the public with free access to digital Lithuanian thesauruses and dictionaries covering the languages of our closest neighbours: Lithuanian-Latvian-Lithuanian, German-Lithuanian and Lithuanian-Polish-Lithuanian bilingual dictionaries. The system will come with many functions that will allow its users to easily and conveniently find both translations of words they are after, and synonyms and antonyms for those words, among other things. In future, this infrastructure will be augmented with other lexicographical resources for one or several languages, it will be integrated into international alliances like META-NET, etc. In future, we will be able to see more clouds of multi-lingual linguistic resources, dictionaries among them, which will most likely be joined together to make a digital sky. That would greatly facilitate communication on the European level and would promote the dispersion of its cultural diversity.

The Lithuanian digital resource infrastructure facilitates the learning and teaching of language(s), as well as linguistic field-work. However, with a plethora of different infrastructures in place, the matter of training to use them in a creative way gains a certain degree of relevance. Lithuania is currently playing host to seminars for teachers, aiming to introduce them to the latest digital resources and the possibilities to use them.

Conclusion: Lexicographical infrastructures and infrastructures of other linguistic resources are a much better tool that helps satisfy the needs of different users and enhance the possibilities to utilise various dictionary data.

5. References

- Kažukauskaitė, O. (2002): Le grand dictionnaire d'une petite nation, une histoire de cent ans. In: *Cahiers lituaniens* 3, 29-33.
- Kļaviņš, P. (2002): Lietuvju valodas vārdnīca (I-XX sējums). In: *Brīvā Latvija* (Rīga), 27 jūlijs-2 augusts.

- Lietuvių kalbos žodynas* (I-XX/1941-2002): elektroninis variantas / Naktinienė, Gertrūda et al., Vilnius: Lietuvių kalbos institutas, 2005, 2008. Internet: www.lkz.lt. [The dictionary of the Lithuanian language (Vol. 1-20/1941-2002): electronic release [online], 2005, renewed version, 2008.]
- Lietuvių tarmės* 1-3 (2000-2011): elektroninis išteklius /Lithuanian dialects: electronic release. Vilnius: Lietuvių kalbos institutas. Internet: www.mch.mii.lt/tarmes/Tarmes/.
- Pakalka, K. (ed.) (1997): *Senasis Konstantino Sirvydo Žodynas*. Vilnius: Mokslo ir enciklopedijų leidybos institutas. [The old dictionary of Konstantinas Sirvydas. Facsimile and comments.]
- Telksnys, L./Kligiene, N. (2001): Facing challenges of new media on the crossroads of East-West and past-future (Lithuanian case study). In: *Cultural Heritage and Technologies in the Third Millennium*. Milan: Politecnico di Milano, 57-65.
- Vaišnienė, D./Zabarskaitė, J. (2012): *The Lithuanian language in the Digital Age/Lietuvių kalba skaitmeniniame amžiuje*. (= White Paper Series/Baltųjų knygų serija). Berlin: Springer.
- Zabarskaitė, J. (2007): Prameny a databáze Institutu pro litevský jazyk. In: *Europeica – Slavica – Baltica*. (= Publikace Slovanské knihovny 56). Praha: Národní knihovna České republiky, Slovanská knihovna, 261-272.
- Zabarskaitė, J. (2009): The Institute of the Lithuanian Language and Centres of Baltic Studies. In: *The Baltic languages and the Nordic countries*. Vilnius: Lietuvių kalbos institutas, 135-147.
- Zabarskaitė, J./Naktinienė, G. (2004): On the linguistic databases of the Institute of the Lithuanian language. In: *Human Language Technologies: the Baltic perspective: the first Baltic conference*. Riga, 187-190.
- Zabarskaitė, J./Naktinienė, G. (2007): Слоўнікі ў інтэрнеце: дыялог мовазнаўцы і носьбіта мовы. In: *Слово и словарь. Vocabulum et Vocabularium*, Гродно: ГрГУ, 32-34.
- Zabarskaitė, J./Naktinienė, G. (2010): The dictionary of Lithuanian (LKŽ) and its Future in Databases and Electronic Version. In: *Proceedings of XVI Euralex International Congress*. Ljouwert: Fryske Akademy – Afuk, 780-787.
- Zabarskaitė, J./Naktinienė, G./Valiukas, L. (2013): Contemporary dictionary application: tips and tricks of getting it right. In: *TRALOGY II: futures in technologies for translation*. Paris. Internet: www.tralogy.eu.
- Zinkevičius, V. (2004): Creating the electronic version of the dictionary of Lithuanian. In: *Human Language Technologies: the Baltic perspective: the first Baltic conference*. Riga, 170-173.
- Zinkevičius, V. (2007): The digitization of the dictionary of the Lithuanian language. In: *The Third Baltic Conference on Human Language Technologies*. Kaunas, 349-355.
- Забарскайте, Й./Нактинене, Г./Шпятыте, Р. (2005): Современность и перспективы большого “Словаря литовского языка”. In: *Исторический путь литовской письменности*. Вильнюс: Институт литовского языка, 340-355.
- Топоров, В.Н. (2004): К выходу в свет большого “Словаря литовского языка”. In: *Балто–славянские исследования XVI*. Москва, 408-415.

