Ieva Zuicena (ieva.zuicena@inbox.lv) Ilga Migla (ilga.migla@inbox.lv)

The Dictionary of Contemporary Latvian Language and its Lexicographical Process

1. Introduction

The Dictionary of Contemporary Latvian Language (DCLL) (Mūsdienu latviešu valodas vārdnīca/MLVV) is a new type of dictionary in the Latvian lexicography which is completely accessible on the Internet (www.tezaurs.lv/mlvv) since the year 2014. This dictionary is compiled based on different kind of sources: specifically made card index with selected excerpts from the newest literary works, newspapers, magazines etc., various websites, e.g. www.letonika.lv; www.news.lv; www.google.lv, also The Balanced Corpus of Modern Latvian (www.korpuss.lv)

Initially, the DCLL was expected to be as a dictionary which would contain the Latvian vocabulary used in the last 30 years, however, during the work, it was proven that it is not possible to draw strong lines concerning this question. After the restoration of independence of the Republic of Latvia in 1991 and the shift to market economy, huge reforms took place also in the vocabulary of the Latvian language. Those words which were not commonly used in Soviet times or were regarded as only to be characteristic for capitalists (*shareholder*, *unemployed*, *dividends*, *benefice*, etc.), now have been fully integrated into contemporary Latvian language, whereas, the passive layer of the language now has words, i.e. sovietisms, which are connected with soviet life (*piecgade* 'five-yearly', pionieris 'pioneer',*komjaunatne* 'Young Communist League'*, stahanovietis* 'stahanovite'*, mičurinietis* 'michurinite', etc.) If the DCLL stayed as a book, then, possibly, words like these would be crossed out from the entry list, due to the unlimited storage space of the Internet, the DCLL kept these sovietisms. The DCLL also includes some historicisms and archaisms which are pretty often found in publications about historical questions or in the works of the classic authors.

The DCLL has been created in the Latvian Language Institute of the University of Latvia under the guidance of Dr. philol. Ieva Zuicena. She works in the lexicography field already since 1976, starting her work with the first Latvian explanatory dictionary in eight-volumes: The Dictionary of Standard Latvian (1972-1996). Also several other DCLL authors (Mg. philol. Laimdota Oldere, Dr.philol. Ārija Ozola, Dr. philol. Imants Šmidebergs) became professional lexicographers by working at this dictionary. The newest authors (Dr. philol. Ilga Migla and Dr. philol. Anita Roze) are researchers/lexicographers of the Latvian language, they have obtained the knowledge of a lexicographer by working at the DCLL.

2. The lexicographical workflow of the DCLL

Work at the DCLL can be divided into following phases:

- Preparation – gathering of materials, making the design of the dictionary, working out the instructions for authors, the writing sample entries and pilot version of the dictionary (the writing 1-3 % text of the whole dictionary, then carrying out the necessary corrections).

- Writing and Editing – writing the text of the dictionary and editing.

- Computerization – putting the prepared text on the Internet.
- Corrections and Improvements – error correction, making the indicator of entry words and performing other improvements.
- Afterlife – regular (every 4 years) revision of the dictionary.

2.1 Phase of preparation

Work at the DCLL was started in 1997. As stated in the introduction, firstly, the DCLL was planned as a book, because the computerization was not so widespread in Latvia. Personal computers were very expensive and there were not many people who could afford them. Also the Internet, its abilities were not accessible to all our community.

At that time the Latvian language corpus was not yet made, hence, a new card index was formed for the DCLL`s needs, with excerpts from the newest belles-lettres and periodicals, slightly also from postscripts of spoken vocabulary.

The DCLL project was developed by the editor – Ieva Zuicena, taking the first Latvian scientific explanatory dictionary as the basis – *The Dictionary of Standard Latvian* („Latviešu literārās valodas vārdnīca"), for its framework and principles. Unlike *the Dictionary of Standard Latvian, the DCLL* incorporated more colloquial and slang words, the explanations were formed shorter, simpler, and more easily perceivable by the reader. After the editor Ieva Zuicena did fieldwork in the Oxford University Press, in the Dictionaries Department (1998), the DCLL instruction was reworked and the dictionary supplemented with several innovations, such as including the most widespread abbreviations in the dictionary, plant and animal families or species provided with Latin names, reflection of the word origin not with the traditional sign (<) but with a full sentence, commentary about the usage of the word. These innovations were adopted from *The New Oxford Dictionary of English* published by the Oxford University Press in 1998.

In 2001, after the re-worked dictionary project, it was written approximately 2% of the DCLL`s text. When this amount of text is written for a dictionary, then calculations can be done in order to see how much time it will be needed for the complete dictionary and how much finances will be required. Due to the insufficient

financing, the DCLL project could not employ as many workers as it would require; hence, the writing of the text happened very slowly. Understanding that the compilation of the dictionary would take more time than expected, a decision was made to publish the dictionary on the Internet by parts, as soon as a new part was written and edited, because the community was in dire need of a new, sufficiently broad explanatory dictionary of the Latvian language. It can be added that the first three volumes of *The Dictionary of Standard Latvian*, published from 1972 – 1975 were particularly aged.

2.2 **Writing and Editing**

The writing and editing of the DCLL`s text took 14 years, from 2001, till 2014. In the beginning, the dictionary had little funding from the Latvian Council of Science, from 2006, till 2009, the DCLL received support from the National Research Programme "Letonika", but from 2010 – 2014, the DCLL project was included in the National Research Programme "National Identity".

When the body matter of the DCLL was started, the Latvian language corpus was not yet made, thus, the example material was chosen from the excerpts of card index or found on websites which only had appeared recently. This work was done by the DCLL`s authors themselves, because the only assistant was occupied with the arrangement and supplementation of card catalogue. It must be told that for the authors it took quite a lot of time to find suitable examples and the best collocations. Illustrative examples in the DCLL usually are given as groups of words or in the form of reduced sentences. Stylistically marked, outdated, etc. word illustrations sometimes do need a broader context; in such cases, the DCLL offers an example behind the circle (○) in the form of citation, denoting the author in the brackets.

Step by step various corpora of the Latvian language were developed. They were formed in a different research institution - Artificial Intelligence Laboratory /Institute of Mathematics and Computer Science at the University of Latvia. The Balanced Corpus of Modern Latvian currently comprises 4, 5 million word uses from various sources, in the future it is expected to be increased to at least 100 million word uses (Levāne – Petrova K. 2012). The website www.korpuss.lv offers more other specialized (unbalanced) corpora, e.g. the Latvian Internet language corpus, transcripts from the Latvian government (Saeima) meetings, corpus of the ancient Latvian texts. Unfortunately, all these corpora are too small for such a bulky dictionary which is the DCLL, thus, they are more useful as a supplementary aid to ascertain the most typical usage. To illustrate the rare words, examples must be searched in other Internet sources.

The written DCLL text is handed in to the editor for editing. Depending on the quality level of the written text, the editing is required 2 or 3 times. When the entire text of the dictionary is written, another editing of the entire text is necessary, coordinating

and unifying the explanations of lexicosemantic groups of words, eliminating inconsequences and other deficiencies. This work has not been done for the DCLL.

2.2 Computerization

The text of DCLL was written without using a special dictionary writing system. Such system has been installed for another Latvian e-dictionary: corpus – based electronic historical dictionary of Latvian, which will cover the texts from 16 and 17 centuries.We look at the corpora manual, it is small and relatively scanty.

Since the Latvian University Language Institute does not have their own computer specialists, the DCLL`s input on the Internet is made by the employees of the Artificial Intelligence Laboratory of the Latvian University. The text of the dictionary is written in DOC format; it is modified to HTML format, after the text is cut into entries and indexed according to the first keyword.

This way in 2003 with the financial support of the State Language Agency (now known as the Latvian Language Agency), the first part of the DCLL was placed on the Internet with entries from letters A to D. Gradually publishing every written part, in 2014 the entire dictionary was placed on the Internet.

2.4 Corrections and Improvements

These activities relate on the time period from 2014 till 2017.

In this time period, it is anticipated to perform several improvements of technical nature (for example, implementation of word indicator, reproduction of the pronunciation of glossary words, not graphically but with the help of a voice synthesizer, explanation supplementation with sound recordings, depictions and video materials) so that the DCLL user could make use of all the opportunities the modern technologies have to offer.

3. Time span of the different phases

Phase	Duration
Preparation	1997 - 2001
Writing and Editing	2001 - 2014
Computerization	2003 - 2014
Corrections and Improvements	2014 - 2017

References

1) Andronova E., Trumpa A., Vanags P., Latviešu valodas vēsturiskās vārdnīcas (16.–17. gs.) projekts: problēmas un risinājumi. In: Apvienotais Pasaules latviešu zinātnieku III kongress un Letonikas IV kongress „Zinātne, sabiedrība un nacionālā identitāte". Valodniecības raksti. Rīga: LU Latviešu valodas institūts, 2012, 196.–209.

http://www.lulavi.lv/media/upload/tiny/files/PLZK_Letonika%20Valodniecibas%20raksti-Saturs.pdf

2) Ceplītis L. Pirmā latviešu valodas skaidrojošā vārdnīca. In: Latviešu valodas kultūras jautājumi. 6. laid. Rīga: Liesma, 1970, 5.–6.

3) Jērāne S. Darbības vārdu semantikas skaidrojums latviešu literārās valodas vārdnīcās. Maģistra darbs LU. Rīga: LU, 2013.

4) Levāne - Petrova K., Līdzsvarots mūsdienu latviešu valodas tekstu korpuss un tā tekstu atlases kritēriji.In: Baltistica VIII Priedas, 2012, 89.-98.

5) Stengrevica, M. Latviešu valodas aktuālās problēmas. In: Valodas prakse: vērojumi un ieteikumi, Nr. 1, 2005, 6.-10.

6) Stengrevica, M. Latviešu literārās valodas vārdnīca (problēmas un risinājumi). In: Vārds un tā pētīšanas aspekti. Liepāja: 1997, 155. - 161.

7) Stengrevica, M. Daži vārdi par Latviešu literārās valodas vārdnīcu (sakarā ar pēdējā sējuma iznākšanu pēcvārda vietā). In: Linguistica Lettica 2. Rīga: LU Latviešu valodas institūts, 1998, 115. - 120.

8) Vārdnīcu izstrāde Latvijā:1991–2010. Pētījums. J. Baldunčika vad. Atb. red. A. Lauzis. Rīga: LVA, 2012. 248. http://www.valoda.lv/Petijumi/Petijumi/mid_509

9) Veisbergs, A. Elektroniskās vārdnīcas un datorizētu vārdnīcu resursi. In: *Linguistica Lettica 12*. Rīga: LU Latviešu valodas institūts, 2003, 163. - 175.

10) Veisbergs A. Historical Comparison of the Iconic Dictionaries of the Three Baltic Nations. In: Proceedings of the 15th Euralex International Congress. Oslo: University of Oslo. 2012. 245-249.

11) Veisbergs A. Semantic aspects of reversal of a set of bilingual dictionaries. In: Dictionary Visions, Research and Practice. Ed. by Henrik Gottlieb and Jens Erik Mogeson. John Benjamins 2007. 71-82.

12) Zuicena I., Ilustratīvā materiāla atveide „Mūsdienu latviešu valodas vārdnīcā". In: Baltistica. VIII Priedas. (Redaktorius Bonifacas Stundžia) Vilnius: Vilniaus universitetas. 2012, 181–188. http://www.baltistica.lt/index.php/baltistica/article/viewFile/2120/2026

13) Zuicena, I. Laimdots Ceplītis un "Latviešu literārās valodas vārdnīca". In: Linguistica Lettica *20*. Rīga: LU Latviešu valodas institūts, 2012, 209. - 218.

14) Zuicena I. Latviešu valodas attīstības procesu atspoguļojums „Mūsdienu latviešu valodas vārdnīcā". In: Apvienotais Pasaules latviešu zinātnieku III kongress un Letonikas IV kongress „Zinātne, sabiedrība un nacionālā identitāte". Valodniecības raksti. Rīga: LU Latviešu valodas institūts, 2012, 236.–246,

http://www.lulavi.lv/media/upload/tiny/files/PLZK_Letonika%20Valodniecibas%20raksti-Saturs.pdf

15) Zuicena I., Grūzītis N. Informācijas tehnoloģiju izmantošana leksikogrāfijā un "Mūsdienu latviešu valodas vārdnīcas" izstrādē. In: Letonika . Piektais kongress. Plenārsēžu materiāli. Latvijas Zinātņu akadēmija. Rīga, 2013, 129.-134.

16) Zuicenu I., Migla I. Ieskats Kārļa Mīlenbaha "Latviešu valodas vārdnīcā". In: Linguistica Lettica 12. Rīga, 2004,121.-128.