ktu faculty of social sciences, arts and humanities 1922

44/2024
Research Journal
Studies about Languages
pp. 82–96
ISSN 1648-2824 (print)
ISSN 2029-7203 (online)
DOI 10.5755/j01.sal.1.44.35882

LINGUISTICS / KALBOTYRA

Lessons learned from an electronic botany dictionary compilation project

Received 12/2023 | Accepted 05/2024

HOW TO CITE: Sviķe, S. (2024). Lessons learned from an electronic botany dictionary compilation project. *Studies about Languages / Kalbų studijos, 44*, 82–96. https://doi.org/10.5755/j01.sal.1.44.35882

Lessons learned from an electronic botany dictionary compilation project

Elektroninio botanikos terminų žodyno parengimo projekto patirtis

SILGA SVIĶE, Ventspils University of Applied Sciences, Latvia

Abstract

The study offers insights into the design, working model, and solutions of the specialised electronic dictionary of botany terms in Latvian, Latin, English, German, and Russian, developed as a mobile application. The third version of this dictionary has been updated as a final one within the project 'Smart Complex of Information Systems of Specialised Biology Lexis for the Research and Preservation of Linguistic Diversity' and published in the Google Play Store and the Apple Store in January 2024.

The study describes various aspects of the conception of a dictionary of botany terms and offers a practical dictionary compiling model for the dictionary's mega-, macro-, and microstructure. Prior to commencing the project, a survey of the potential users was conducted to understand the preferences of the respondents and to take these preferences into consideration when developing the dictionary. New solutions to the aspects typical of a specialised dictionary of botany terms (labelling of taxonomic categories, visualization with interactive images, automatic retrieval of definitions from existing sources and their usability, etc.) were found. The study is summarised by discussing the main lessons that have been learned in the process of developing a small (~3000 entry words) specialised dictionary. The conclusions might be useful for lexicographers and translators interested in compiling dictionaries.

KEYWORDS: specialised lexicography, botany terminology, mobile application, specialised electronic dictionary.

Introduction

Various specialised dictionaries are widely used not only for translation, but also as a source of information; thus, there is a high demand for such terminological works comprising both translating and explanatory dictionaries (Sviķe, 2018, p. 239). The intention of compiling a new botanical dictionary is based on the necessity for the source where one of the working languages is Latvian, and that might be a useful translation tool not only for translators, but also for students and teachers of natural sciences, journalists, workers of biopharmacy, pharmacists, as well as nature enthusiasts and other interested persons. Considering the necessity for such terminological work and the previous studies of this field (Sviķe, 2015, 2016, 2017), the existing resources of various types and the collaboration opportunities between the researchers of the Faculty of Translation Studies (FTS) and the Faculty of Infor-

mation Technology (FIT) of Ventspils University of Applied Sciences (VUAS) were explored; available materials and the materials yet to be obtained were identified and collected. A first prototype of the botanical dictionary was already developed in 2017 (Sviķe & Šķirmante, 2019), then the prototype of the mobile application was updated, and the term list for the dictionary was expanded to 890 units in 2018, but in 2023, the final version of the dictionary was ready for publication with about 3000 terms in Latvian and their equivalents in Latin, English, German, and Russian. This study describes the implementation of the dictionary compilation project, the challenges faced, the solutions and results of the practical work in lexicography, and the lessons learned.

Through mostly using a descriptive method, the goal of the study is to describe the conception of the dictionary compilation project, the practical working model of the dictionary, the mega-, macro-, and microstructural units of the specialised dictionary as a mobile application *The New Botanical Dictionary. Latvian-Latin-English-German-Russian Terms*, which was developed and collected throughout the practical part of the project by outlining the problematic aspects, offering initial solutions, and considering the future perspectives of the dictionary. In this study, the dictionary is mostly described from a lexicographical perspective since a detailed terminology issue analysis requires a separate study.

There is no doubt that the importance of electronic lexicography has increased, especially over the last decade. Several studies and presentations at conferences and congresses, published monographs and articles (e.g., Bergenholz et al., 2009; Granger & Paquot, 2010; Fuertes-Olivera & Bergenholtz, 2011; Fuertes-Olivera & Tarp, 2014) as well as newly developed electronic online information resources, databases, and electronic dictionaries are indicative of that (e.g., QA repository of the State Language Agency[1], Electronic dictionary of Latvian surnames[2], Website for personal name rendering[3], Latvian Language Manual[4], e-PUPA[5], The Historical Dictionary of the Latvian Language[6]).[7] To make a comparison, the Institute for German Language in Germany (*Institut für Deutsche Sprache*) has developed the Online Bibliography of Electronic Lexicography (OBELEX)[8], which is a collection of studies on electronic lexicography and provides an online database of dictionaries (see more in Müller-Spitzer & Möhrs, 2010, pp. 439–444). However, according to the answers given by 117 respondents within the lexicographic survey conducted at the end of 2017 (Sviķe, 2018, pp. 233–235), the existing resources, databases, and electronic encyclopaedias with Latvian as one of the languages are not always useful when translating from Latvian into other languages or when translating into Latvian; this especially refers to translating specialised texts. The respondents also indicated that the information in the resources is outdated and not being renewed. This also applies to Latvian National Terminology Portal[9] (Latvian – *Latvijas Nacionālais terminoloģijas portāls*). One more disadvantage of the existing electronic resources is the fact that some electronic dictionaries can only be used for a fee, so they are not available to everyone (e.g., the translation dictionaries of *Letonika*[10]).

The further sections of this study provide an insight into the development and working model of the electronic botany dictionary compilation project, as well as discuss some practical lexicography solutions from the perspective of specialised lexicography.

[1] www.valodaskonsultacijas.lv

[2] www.uzvardi.lv

[3] www.personvarduatveide.lv

[4] www.valodasrokasgramata.lv

[5] www.epupa.valoda.lv

[6] www.tezaurs.lv/lvvv

[7] Generally, resources intended for searching information in the Latvian language when translating.

[8] www.owid.de/obelex/

[9] www.termini.gov.lv

[10] www.letonika.lv

Defining and Understanding the Basic Terms of the Study

Prior to the description of the conception and the working model, the basic terms and the meanings thereof should be clarified. At the beginning of conceptualizing the project, the idea was that the dictionary conception and the planned working model would result in a "specialised electronic lexicographical work" (Fuertes-Olivera & Tarp, 2014, p. 12) – a dictionary. The term '*specialised electronic lexicography*' is defined as "the branch of lexicography concerned with the theory and practice of specialised e-dictionaries in general" (Fuertes-Olivera & Tarp, 2014, p. 12). The term '*dictionary of terms*' throughout this study is understood as "a dictionary where separate words, collocations (with the term function), and basic terms are shown together with synonyms [..]" (VPSV, 2007, p. 401), as given in the Explanatory Dictionary of Linguistic Terms (Latvian – *Valodniecības pamatterminu skaidrojošā vārdnīca*); however, this dictionary does not provide any explanation of the terms '*electronic dictionary*' and '*electronic dictionary of terms*' in the Latvian language. The electronic Handbook of the Latvian Language (Latvian – *Latviešu valodas rokasgrāmata*) (Jansone & Vulāne, 2018) explains that "the data in an electronic dictionary is given in an electronic (digital) form, and they can be accessed through various digital devices." A similar definition and a collection of broader examples were provided by Nesi (2000, p. 839) in her study published in 2000:

> "The term *'electronic dictionary'* (or ED) can be used to refer to any reference material stored in electronic form that gives information about the spelling, meaning, or use of words. Thus, a spell-checker in a word-processing program, a device that scans and translates printed words, a glossary for on-line teaching materials, or an electronic version of a respected hard-copy dictionary are all EDs of a sort, characterised by the same system of storage and retrieval."

Therefore, '*an electronic dictionary of terms'* is a dictionary including words and collocations that are used with the term function; the data there is collected in an electronic form and can be accessed through digital devices. Taking into account the study fields and working languages provided by the FTS VUAS and the collaboration opportunities with the researchers of the Faculty of Information Technology (mainly to gain experience), the idea of an electronic translation dictionary was considered at the very beginning. In this study, an '*electronic dictionary*' is understood as a tool that has been compiled by humans and has been developed for human use (de Schryver, 2003, p. 145). As the developed dictionary of terms will be further analysed and described in the study, '*an electronic translation dictionary of terms*' here is understood as an electronic dictionary of terms providing equivalents in other languages and additional information about the items searched.

Jērāne (2017, pp. 74–85) has compared the aspects of dictionary structure, their conceptions and terminology in the works of many authors (e.g., Bergenholtz, Jakaitienė, Karpinska, Nielsen, Svensén, Tarp), and has distinguished between the functional structure (cross-reference structure), access structure, distribution structure, and theme-rheme structure, as well as the structural aspects (microstructure, macrostructure, frame structure). One must agree that the structure of a dictionary is mainly defined by its functions; therefore, the structural and functional aspects in a practical analysis of a dictionary cannot always be strictly separated (Jērāne, 2017, p. 82). Since the goal of the present study is to describe the conception of the dictionary and discuss the practical result of the working model, a '*dictionary structure*' in this study is understood as all the parts of a dictionary: i.e., the megastructure or the overall structure of the dictionary, the macrostructure (the overall term list and its organisation), and the microstructure, or the architecture of individual parts within the dictionary entry. Unlike printed dictionaries, an entry of an electronic dictionary may include not only the textual information, but also hyperlinks, graphic information and even videos; therefore, for the description of the developed dictionary, a broader traditional terminology which was initially attributed to printed dictionaries has been used: when looking at macrostructure elements, the term '*macrostructural units*' has also been used regarding electronic dictionaries, but the term '*microstructural units*' has been used when looking at microstructure elements that refer to a dictionary entry (Müller-Spitzer, 2013, pp. 367–381), where some of the elements are also the internal references or links (within the particular dictionary) and external references or links (references leading to an external information or information outside the dictionary). Structural levels of dictionary publication models have been used when characterizing dictionaries: the lexicography database level that has been described

from the point of view of the dictionary developers, and the data presentation level that has been characterised from the point of view of the dictionary user (see more Müller-Spitzer, 2013, pp. 367–381). When describing the presentation level of lexicographic data, common and basic terms of the IT field (view, menu, input field, etc.) have been applied.

As vernacular plant names are characterised by a broad synonymy, they belong to the specialised lexicon of the field of botany. Latin or scientific plant names are terms that form a part of the specialised lexicon. In this study, *'term'* is a word or combination of words denoting a specific concept of a particular thematic area and a special field (VPSV, 2007, p. 400). A significant part of the plant names (in Latvian, English, German, and Russian) are also terms, but some names cannot be called terms because they are synonymic names that are used widely or in some exceptional cases; for example, Latvian *margrietiņa* is a synonymic but not terminological name for Leucanthemum vulgare, as the Latvian terminological name is *parastā pīpene*. In a few exceptional cases, such popular names are also included in the dictionary. The dictionary is compiled mainly using an onomasiological approach (in the creation of the basic term list) and includes mostly terminological plant names, which are considered to be terms, and other botanical terms.

Major Aspects of the Electronic Dictionary Development

Conception of the Dictionary

The author of this article follows Fuertes-Olivera and Tarp's (2014, pp. 62–64) '*functional approach to dictionary development*': functions of the dictionary should be included as foundation stones in any lexicographic concept since they define what will be done with a final product – the dictionary, such as its content and form, the data and their selection, preparation and accessibility. The starting point of the functional approach to dictionary development is the situations, characteristics and needs of the user (which have been ascertained in this project through a survey of potential users), and then initializing the lexicographic process in terms of specific dictionary design and compilation (Fuertes-Olivera & Tarp, 2014, p. 57). However, for a detailed description of the abovementioned stages it is necessary to obtain the practical experience of compiling an electronic dictionary. During practical work, various planned and unplanned changes occur, and it should be noted that the path to a complete electronic dictionary may differ for different compilation projects.

Before the development of the first prototype (the first version of the dictionary), research on the necessity of such a dictionary was conducted (Balode, 2012, p. 40; Sviķe, 2017, pp. 22–30; Sviķe, 2018, pp. 228–241; TTC, 2007, pp. 42–45), the working group for the project was created, as well as the time schedule was elaborated and a prototype dictionary of 500 entries was developed. In the second dictionary updating project, in 2018, the term list was expanded, and better technical solutions were implemented. In the third version, or final project, the extended working group was formed, and the final concept of the dictionary was created, the time schedule was set up for the development of the dictionary and the required funding within a larger project was specified, estimated, and allocated. During the very first stage of the project preparation for the first dictionary prototype, issues regarding the determination of the potential target audience of the dictionary (users) and the functions of the dictionary were already solved. Furthermore, based on the previous considerations, questions regarding the structure of the dictionary were expanded and the solutions for the dictionary development were practically realised.

Working Model for the Development of the Dictionary

A lexicographer (who is also a linguist) and a terminologist as an expert in the particular field should definitely participate in the development of a dictionary (Fuertes-Olivera & Tarp, 2014, p. 192; Kaufmann, 1998, p. 206; Vachková, 2011, pp. 31–33), as dictionary compilation is mainly the result of a teamwork. Since the plan is to develop an electronic dictionary of terms, the working group should definitely include computer specialists (programmers or computer linguists) and field experts (Fuertes-Olivera & Tarp, 2014, p. 192; Vachková, 2011, pp. 90–102). For the complete electronic dictionary to reach its users, the working group should also involve experts in the fields of marketing and selling online dictionaries (Fuertes-Olivera & Tarp, 2014, p. 192), in case the dictionary is intended to be offered to its users for a fee. As this dictionary project is not a commercial one and is offered free of charge, experts in the fields of marketing and selling were not engaged.

The working group for the development of the dictionary model described in the article consists of the project manager, representatives of the academic staff – a lecturer (and a translator), a proofreader, a leading researcher (a terminologist and lexicographer), as well as last-year students of FTS VUAS bachelor's and master's study programmes (participating in the project for the scientific internship required in the study program). A bachelor and master student of computer sciences from the FTS VUAS (who developed a bachelor's thesis on mobile applications) and a FIT lecturer (a researcher and a doctoral student) were involved in the project as well. Botanists and agronomy specialists were engaged as external experts: a doctor of biology, a practitioner in the field of botany, who works at the National Botanic Garden of Latvia, as well as a doctor of agronomy, a researcher of organism names and a terminologist. An artist, a master of arts (also a doctoral student of art), was involved in the project to draw the pictures for the visual part of the dictionary. As the planned result of the project is a multilingual dictionary, also external expert consultations and lexicographic works (multilingual dictionaries, encyclopaedias, and other specialised literature) of recognised authors in foreign languages were used to select data (equivalents and additional information) in each language. By creating a working group, an interdisciplinary cross-sectoral partnership was ensured for the project, as well as compliance with the regulations outlined in the project proposal for the Latvian Academy of Sciences regarding the involvement of employees of all levels in the relevant project: the team of project developers should include students of master studies and/or doctoral studies, early career researchers and students. An important aspect is that the work performed by the students should always be verified and supervised by the project manager. By involving staff of all levels, impact for inter-related sectors and users is also ensured, as not only the academic staff is involved.

Users of the Dictionary

Any dictionary is developed for a particular target audience or the potential users of the dictionary as expected by the creators thereof. Defining a target group is one of the basic requirements prior to planning a dictionary (Barz et al., 2005, p. 15; Fuertes-Olivera & Tarp, 2014, pp. 45–57); therefore the compilers should conduct a lexicographic survey, perform a data analysis, and summarize the desires of the potential users (Bergenholtz & Tarp, 1995, p. 77). Questionnaires, interviews or similar methods are some of the ways to understand the desires of the potential users, alongside with such methods as observation and deduction (Fuertes-Olivera & Tarp, 2014, pp. 53–54). Questionnaires were the method used in this project, and based on the results acquired and by taking into consideration the suggestions and desires expressed by the respondents, the dictionary prototype development project was implemented practically (see Sviķe, 2018).

The expected target audience includes students and teachers of translation and natural sciences, pupils of the higher grades, professional and unprofessional translators, gardeners, florists, pharmacy employees, flower arrangers, journalists and nature enthusiasts who use the terminology of this field and who would need the translations and explanations of botany terms. Since the modern educational system is characterised by mobility (pupils and students going on exchange programs) and learning specialised subjects like geography and biology through CLIL (Content and Language Integrated Learning), such a dictionary would also be useful for pupils to translate and learn plant names and other botany terms. The methodical materials and lesson samples for secondary school pupils available on the homepage of the National Centre for Education of the Republic of Latvia (Latvian – *Valsts izglītības satura centrs*) – *Biology in German and English* (Latvian – *Bioloģija vācu un angļu valodā*)[11] – show that the CLIL methods are being used in schools. Therefore, pupils were also included in the group of respondents for the lexicographic survey conducted prior to developing the dictionary. The group of potential dictionary users is rather wide:

1. students of translation and professional translators;
2. students of the natural sciences and natural science professionals;
3. professionals and novices in gardening and other fields linked to botany;
4. secondary school teachers and students;
5. laypersons interested in botany topics.

[11] https://www.visc.gov.lv/lv/metodiskie-materiali-1

The expected users of the dictionary have very good knowledge of the Latvian language, as the basic language of the dictionary is Latvian (the interface is also available in English, German, and Russian) and good knowledge of another language contrasted in the dictionary.

The Dictionary Usage Situations

The dictionary is mainly intended for translation in various situations. It also includes the functions of an explanatory dictionary, as it contains definitions and additional information. One must agree to Vachková's (2011, p. 29) view, as she quotes the studies of other lexicographers (e.g., Engelberg, Lemnitzer, Mugdan) on the fact that in the lexicographic practice the boundaries for distinguishing the so-called active (*'designed to help with encoding tasks, such as the production of a tex*t' (Hartmann & James, 2002, p. 3)) and passive dictionaries (*'designed to help with decoding tasks'* (Hartmann & James, 2002, p. 106)), such as the comprehension of a text, are not clear enough, and the functions of the dictionaries overlap. It is the electronic environment where translation in both directions (source to target and target to source) can be easily done, therefore the idea is to develop a dictionary that includes the functions of both translation directions, i.e. it would be available for translating from Latvian into another language and from another language into Latvian.

Taking into consideration that a substantial part of specialised botany lexis is composed of plant names, the international scientific designations of which are Latin names, the new botanical dictionary should definitely be a multilingual one. The contrasted languages (Latvian, Latin, English, Russian, and German) were specified after having evaluated the resources and possibilities of the FTS VUAS. These are also the most frequently used languages indicated by students in the lexicographic survey (Sviķe, 2022). English, Russian, German are the languages to be acquired in the bachelor's and master's study programmes of the FTS VUAS. Back in 2007, it was estimated that these languages would be among the dominating languages in term collections to be compiled in the future (TTC, 2007, p. 29), as well as they are three of the most frequently mentioned working languages in the lexicographic survey (Sviķe, 2018, p. 232).

The listed users of the dictionary could use the dictionary for reading, writing and translating texts in the field of botany (in communicative situations) and/or to obtain knowledge in botany (i.e. in a cognitive situation). The dictionary includes three types of educational games to learn the terms. To sum up the abovementioned, the dictionary is intended to be mostly used for translational or educational purposes, and as a reference material.

Dictionary as a Mobile Application

Since the number of mobile devices is continuously globally growing, and in order to ensure a quick and easy access to the dictionary, the realization of the dictionary in the form of a mobile application is one of the most successful options, since most people have mobile devices (e.g., tablets or smart phones) and they are available most of the time. Still, the chosen option of the electronic dictionary, a mobile application, has several advantages, as well as certain disadvantages (rather small display sizes of mobile devices; the need to estimate the expected size of the final application size and the space in a device, etc.). The choice of the dictionary as a mobile application is also determined by the fact that more and more mobile applications are being used to collect data on various public initiatives, such as the Latvian Fund for Nature's mobile application *Dabas dati* for data on nature observation. Considering the disadvantages and the possible situations of usage, the group of potential users could be quite specific (for instance, there might not be many users of older generations and users who prefer dictionaries in the traditional printed form due to individual preferences).

Structure of the Dictionary

From the perspective of the user, the structure of the dictionary is essential to perform its functions. Dictionary criticism is a widespread practice in the evaluation of dictionaries, and to some extent it is due to the fact that the desires and habits of potential users are not taken into account. The research of user desires and habits is also complicated because often the real user of the dictionary is the so called 'unknown' person (Hartmann, 2001, p. 80) and the compilers of the dictionary cannot always predict all actual users of each dictionary. One of the desires expressed by the potential users in the survey conducted prior to the development of the dictionary

(Sviķe, 2018, pp. 230–238) was an easy and convenient usage and a dictionary with a simple structure; therefore, these are the suggestions that have been taken into account when developing the mobile application prototype of the new botanical dictionary.

There are three structural levels in the dictionary: mega-, macro-, and microstructure (Hartmann, 2001; Svensén, 2009). The megastructure represents all components of the dictionary, including the macrostructure and the outside matter (Hartmann & James, 2002, p. 93). The further subsections of the article contain a description of the structure of the dictionary (its mega-, macro-, and microstructure), organisation of the data and preparation of the dictionary from the compilers' perspective, and access to the dictionary data from the user's viewpoint.

The Megastructure

The megastructure of the dictionary is the lexicographic database level of the mobile application. At the presentation level, it has the following sections (mobile application menus), as shown in **Fig. 1**:

1 *About the Dictionary* – the view providing a description of the three versions (first, second, and the final ones) of the dictionary, the persons involved in the project and their functions.

2 *Entry Structure and searching features* – this section includes a search video showing the three search types in the dictionary – *Simple Search, Search by Filters* and *Semantic Search*. The *Simple Search* offers the possibility to type in the term and find the searched item. A list of specific plant names by certain characteristics, such as yellow-flowered, annual, and 0.5 metres tall can be found in the *Search by Filters*. The *Semantic Search* shows relationships at the level of taxonomic categories, e. g. genus, species and cultivars, a description of how to understand the structure of a section (explanation of the microstructure), search features explaining how to perform a simple search, how to search for terms alphabetically, how to use and understand the semantic search function and description of the option *Search by Filters*.

3 *Labels Used* – the section showing a table of abbreviations used in the dictionary and the terms they stand for in all the dictionary languages.

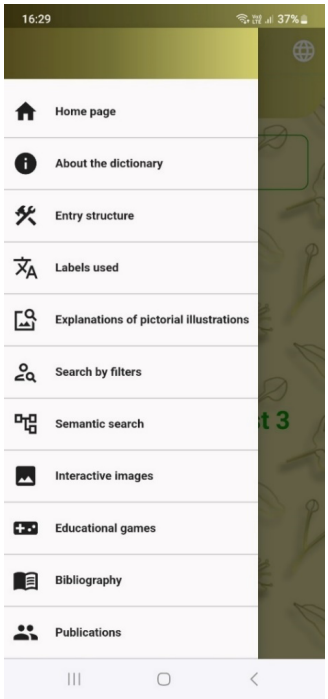

Fig. 1 The megastructure of the dictionary

4 *Explanations of Pictorial Illustrations* – this section contains an explanation of the illustrations used in the dictionary in words.

5 *Interactive Images* – this section contains the following parts: *Plant Structure, Parts of Flower, Parts of Root, Seed Structure, Simple Leaf Parts, Compound Leaf Parts, Inflorescences* – the views showing pictures of a plant, flower, root, seed, leaf or inflorescence where one can interactively translate the term of the chosen part of the plant in all dictionary languages.

6 *Educational Games* – includes three different games to learn the terms: *Questions of the month* (a plant species name is chosen for each month and multiple-choice questions are created), *Put together* (the source and target language are chosen and plant names are given to match them correctly), *Write correctly* (source and target language are chosen and the correct plant name has to be written in the target language). Students were involved in the development of the games content. However, it should be noted that students' work needs to be carefully checked and corrected if necessary.

7 *Bibliography* – the section showing all of the sources used while developing the dictionary.

8 *Publications* – the section containing links to the articles on botany terminology developed during the implementation of the dictionary project.

As already mentioned, the creation of the megastructural units of the dictionary was mainly determined by the wishes and suggestions of the respondents within the lexicographic survey. Additionally, the structure of the existing electronic translating dictionaries that include Latvian was investigated to find the best solution.

The Macrostructure

The macrostructure of the dictionary is two-dimensional, encompassing both alphabetical and thematic organization: alphabetical list of terms in each dictionary language and thematic in the *Interactive Images* section, where it is possible to search for terms thematically, i.e. by looking at an image and selecting the relevant term that explains the image. An updated, revised and improved list of plant names (500 units) created in Sviķe's promotional thesis (2016) is used as the basic plant names list which represents a set of representative plant samples. By involving specialists in the field of botany, the plant names list for the second updated version of the dictionary prototype has been expanded to 890 Latvian plant names and other botanical terms and the list of external selection (selection of plant names in Latvian as the basic language of the dictionary) has been updated. The internal selection (it means selection of additional information to be included in the entries) – selection of term equivalents in other languages and other additional information (grammatical notes, definitions, descriptions) has been described in more details in the section on microstructure of the dictionary (for more on external and internal selection of entries, see Bergenholtz, 2001, pp. 12–14). The term list for the final version of the dictionary consists of approximately 3000 units in Latvian with their equivalents in Latin, English, German, and Russian. The *Home Page* offers a simple search (see **Fig. 1**) within a dictionary database. In the simple search view (*Home Page*), terms can be viewed alphabetically by selecting the pictorial illustration *ABC* and the corresponding letter of the alphabet. The macrostructure of the dictionary also has *Search by Filters* and *Semantic Search* views when opening a mobile application (see **Fig. 1**). A thematic macrostructure is provided in the section *Interactive Images*, where it is possible to search for terms thematically by image.

The Microstructure

Prior to commencing the development project of the dictionary, several conception ideas for the entry microstructure were considered, including all microstructural units, which are shown to the user in the menu fields at the presentation level. Since this is the first attempt to compile a dictionary where researchers and students of two university faculties collaborate, the initial idea was a simple microstructure of a multilingual dictionary, which would generally incorporate only term equivalents in all dictionary languages. However, during the implementation of the project, a challenge was accepted to seek solutions that meet the requirements of a modern electronic dictionary, e.g. by offering the user access to additional data when needed and by providing links to additional data stored in external sources (Fuertes-Olivera, 2014, pp. 92–98). To a large extent, the choice of microstructural units was defined by the respondents' replies in the lexicographic survey: the dictionary should be multilingual with pictures, and be linked with the largest existing dictionaries or databases; it should combine the functions of translating and explanatory dictionaries (Sviķe, 2018, pp. 336–339), thus adapting the units arranged in the microstructure of the dictionary to the desires of the potential users. The microstructural units of the dictionary at the presentation level are described further.

Fig. 2 The microstructure of the dictionary

The microstructural units of the main view are shown in **Fig. 2**, which displays the search results of Abelia, a scientific plant genus. In the dictionary, the languages are marked with flags. Additional explanations within the entry can be used and they are marked with the following codes: DE: Germany, OE: Austria, CH: Switzerland, GB: Great Britain, US: United States. Country codes were selected due to the fact that these are shorter variants of additional descriptions. For example, the equivalent for the scientific or Latin name *Hosta*, more commonly used in the UK, is '*plantain lily*', while in the US it is '*hosta*'. Language flags with equivalents are shown as a block in the term-search view of the main menu; therefore, searching is much quicker and easier. One of the most important tasks

of a lexicographer is to ensure a quick and easy access to relevant data (Fuertes-Olivera, 2014, p. 82). The languages in the new botanical dictionary are given in the following order after the scientific or Latin plant name: there is a term in Latvian, followed by the English, German, and Russian equivalent. This layout is consistent in simple searches, making it easier for the user to navigate and quickly find the translation. The conservative view of botanists is to highlight the Latin names of plants in another font (similar to the plant species database of the Internet Encyclopaedia Nature of Latvia (Latvian – *Latvijas daba*)[12]) or to give the Latin name as the first term, which is considered to be the *tertium comparationis* or meta-language mediating between the source and the target languages (VPSV, 2007, p. 402).

Regarding the microstructural units, some special issues have been outlined and solved. One of the issues is related to the selection of equivalents. As plant names are characterised by broad synonymy (Ēdelmane, 1991, p. 156; Gledhill, 2008, p. 1; Jessen & Schulze, 2008, p. 8; Spies, 1982, p. 222), organisation of synonymous names is one of the key issues in the compilation of a botanical dictionary. During the development of the first two prototypes of the dictionary, a solution for the inclusion and presentation of synonyms that would be acceptable to all participants of the working group was not found. In the final version of the dictionary, it was decided to include a limited number of synonyms, i.e., local language plant names, but not more than three in English and up to two in the other languages (Latvian, German, and Russian). The synonyms are labelled with *(pop.)*, meaning – also well-known or popular. A separate study should be devoted to the description of the completion of the list of plant names and the choice of equivalent criteria. According to a predefined work instruction, equivalents and the sources of their retrieval (lexicographic resources) were compiled into a lexicographic data table. The following principles were set for the selection of equivalents: choosing the most appropriate equivalent following the latest and most authoritative editions, checking their use in three data sources used (e.g., databases, encyclopaedias, dictionaries), and validation by experts. It must be admitted that this process is one of the most significant, time-consuming, and complex; it will continue within the stages of improving and completing the dictionary.

After pressing the search button, the application finds the specific term in the database and returns its equivalents in Latvian, English, German, and Russian. The definition of the term that the lexicographic database contains is returned and presented to the user if it is retrieved from the www.tezaurs.lv database and added to the dictionary (**Fig. 2** – *Tēzaurs*). One of the challenges and tasks of the dictionary development project was to seek solutions to link to and retrieve information from any existing databases, e.g., www.tezaurs.lv, an open access dictionary. The technological solutions for data retrieval are described in Sviķe & Šķirmante (2019).

There are recognisable advantages to such a retrieval of definitions, as it can be performed quickly with almost half of the term list's words in Latvian. Still, there are many disadvantages, as, mostly, definitions in this resource have been collected from various sources, and partially information is inaccurate; moreover, the definitions do not follow a common pattern. Certain objections were received from botanists and terminologists, e.g., indicating that the definition contains an incorrect representation of the Latin (scientific plant name has been recorded using quotation marks), and that the Latin name in the particular definitions is unnecessary, as it has already been included in the equivalent section of the entry after the language flag. It would be preferable for a dictionary to contain definitions that are created following a common pattern and intended specifically for the user of the dictionary. including the minimal information necessary for the user. The results of the practical work testify to Fuertes-Olivera & Tarp's (2014, p. 61) conclusion that subject-field knowledge combined with the needs of target users should not be replaced by blind faith in data extracted from other electronic dictionaries. These automatically retrieved definitions were corrected and adapted to the particular dictionary, but the link to www.tezaurs.lv was left so that the user can quickly navigate to this dictionary and assess further information, synonyms, and examples given in the consolidated dictionary site www.tezaurs.lv, where one can also access the language corpus data.

[12] www.latvijasdaba.lv

In addition to the previously described microstructural units, the user is offered another external reference – a link to the picture of the searched term and its description on www.wikipedia.com (generally, in English) in the microstructural unit *Information* (see Fig. 2). If an internet connection is available on the user's mobile device, the abovementioned websites can be viewed online. By offering such external links, the mobile application prototype creates a combined system for the retrieval of the necessary information from the local database (which can be used offline) and from external sources (which can be used online). This solution allows presenting less data on the screen after the first data request, which is crucial if the user works on a device with a small display (mobile phones, tablets).

The searched term is given in a colour (see colour squares in Fig. 2) that indicates the taxonomic category of the specific term, giving additional information to the user. The main view shows the grammatical categories of the searched term: the gender and the number for the terms in Latvian, Russian, and German. Usually dictionaries of terms do not indicate grammatical categories (VPSV, 2007, p. 401); mostly, they do not provide additional explanations and definitions (TTC, 2007, p. 7). However, grammatical notes (gender and number) were added to the new dictionary, which is a substantial complementary information when translating. Incorporation of such microstructural units in a dictionary of terms should be considered as useful information which, as mentioned above, is usually not incorporated in the dictionaries and databases of terms: e.g., see the Internet Encyclopaedia Nature of Latvia (Latvian – *Latvijas daba*[13]) or printed dictionaries of terms, such as the Dictionary of Agronomic Terms (Latvian – *Agronomijas terminu vārdnīca*) (Purviņa, 1973). However, it was observed that compilers of dictionaries of botanical terms added such information to terms in earlier dictionaries, e.g. N. N. Davydov in the multilingual translating botanical dictionary *Ботанический словарь. Русско-английско-немецко-французско-латинский: Ботаника, словарь* (see Давыдов, 2013), which was first published in 1962 with the entry structure visible in the 2013 reprint edition. It is likely that the addition of such information to translating dictionaries is necessary, in particular for electronic dictionaries. This allows users to receive comprehensive information without the need to consult multiple sources.

The dictionary application provides a solution to access data with the option to search for partial term, e.g. by entering *lap*, which is a part of the word in the entry field of the main view, the dictionary shows all terms containing the abovementioned partial word: *lapa, lapegle, lapsaste, šaurlapu ugunspuķe, kauslapa*, etc. When making a selection, only the data of the selected partial term is retrieved, and the result is displayed in the main view. The data is retrieved from the database through queries, which, if correctly constructed, give the possibility to select the data already from the database. In this case, the query has been structured in such a way that the selected data would contain the part *lap*. These search options are crucial when translating, as the translator can search for the term both in its singular and plural form. Thus, not only the grammatical form is important for the translator, but also the overall possibility to get a set of related terms. Besides, the answers in the lexicographic survey highlighted this aspect as a particularly necessary (Sviķe, 2018, pp. 234–235).

As the dictionary is small, a copy function has also been introduced so that the user can easily copy the searched and retrieved term and paste it into a Google search to get more information. The alphabetical search option allows retrieving a term list for each letter of the alphabet in each language. To ensure that the data is up-to-date and accurate, the user can report errors by sending an email directly from the application (see Fig. 2 bottom – *Send a message*). To keep the dictionary up to date, which is planned for once a quarter, a special additional application was developed (see Fig. 3) that allows the dictionary compiler to correct the data in the dictionary database and update it automatically in the Google Play application. Fig. 3 shows that corrections can be made to each of the microstructural units (the boxes). Data can be added (Latvian – *pievienot*), corrected (Latvian – *izlabot*), or deleted (Latvian – *dzēst*), and when the application is finally refreshed, the data is updated in Google Play. The user of the dictionary will be informed of the updated version when using the dictionary online.

[13] www.latvijasdaba.lv

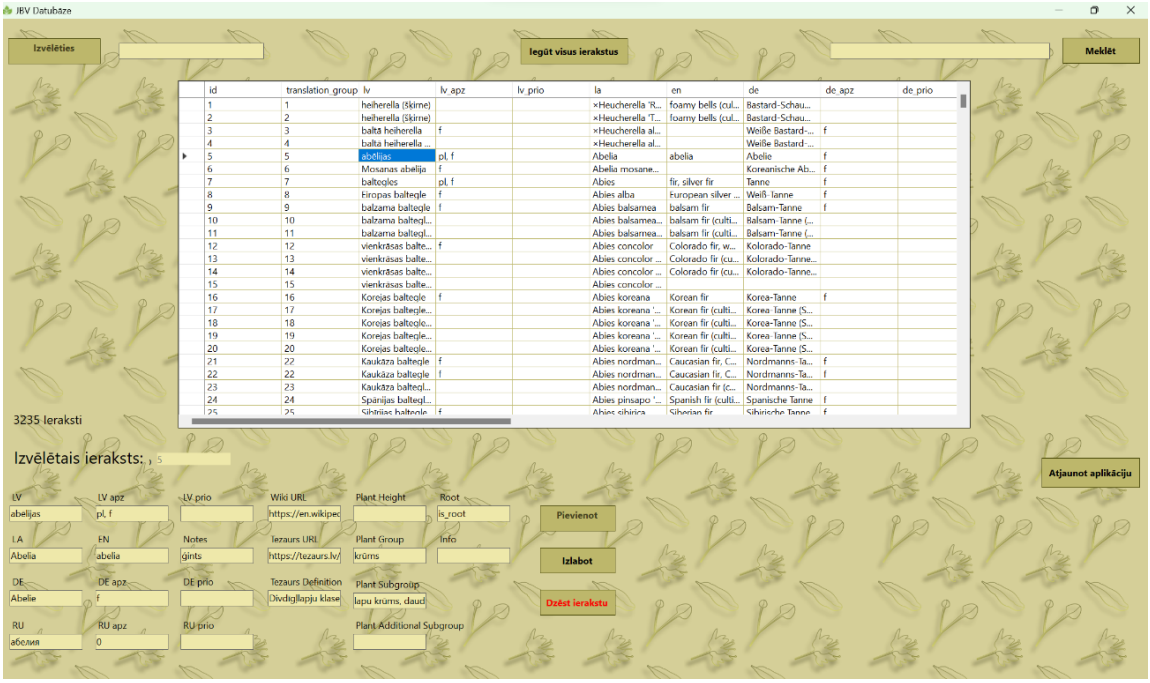

Fig. 3 Dictionary database update application

Conclusions

There is not much research conducted on the process of compiling dictionaries in Latvian lexicography. For example, Roze (1982, pp. 107–111) has written about the development of printed dictionaries. This study summarises lessons learned from this particular electronic dictionary compilation project that could be useful for other dictionary compilers, especially if the electronic dictionary is a mobile application. The dictionary development project has the following phases to be implemented:

1. Research on the need for a dictionary;
2. Research on user preferences and requirements, thus involving potential users in the development of the planned dictionary;
3. Study on raising funds for a dictionary compilation project. The funding is also connected to the time needed for the dictionary project, which must be carefully planned. It should be noted that in a project involving several participants, a deadline overrun is almost unavoidable. Deadlines must be planned with time to spare.
4. Establishment of a dictionary working group, including lexicographers, terminologists as experts in the field (recommended with foreign language knowledge), programmers, translators, as well as students and artists if necessary;
5. Detailed development of the dictionary concept, defining the target audience and the dictionary content (e. g. dictionary languages, microstructural units) tailored to it;
6. Resources to be used for term selection and compilation of the term list, collecting their equivalents and defining the dictionary macrostructure (arrangement of entry words in the dictionary);
7. Editorial proofreading of the whole dictionary and prototype testing before publishing the dictionary;
8. The involvement of students in the dictionary project in the first testing of the dictionary and the incorporation of necessary corrections after the testing;
9. Preparing the dictionary publishing process: environment, payment requirements, access. Publishing dictionaries on the Google Play and the Apple Store requires a specific developer account and creating this account is a multi-step process requiring specific expertise, and it is a paid service;

10 Solutions for the future maintenance of the published dictionary for a wider audience – updating, correcting, adding new data.

These aspects should also be taken into account when developing a small specialised dictionary, as described in this study. In particular, the cooperation of the working group is essential. A dictionary project manager and leading lexicographer oversees all phases and is aware of problems at all levels. For an electronic dictionary, it is important that the lexicographer and the programmers understand each other and closely cooperate to find the best solution for each aspect. The same applies to finding compromises and solutions in the negotiations between the terminologists as field experts and lexicographers who are also linguists on term issues.

The article describes the working model of the development of a dictionary project in one educational and scientific institution, provides an insight into the developed lexicographic conception, depicts the functional applicability of the dictionary, describes the mega-, macro-, and microstructure of the dictionary, describes its linkage with the existing free access dictionaries and websites, as well as outlines the issues, future solutions, and aspects to be investigated in electronic lexicography. One of the greatest benefits of the project is the acquisition of practical lexicographic experience for staff of all levels: students, academic and research staff.

Many lessons were learned during the dictionary development process and many additional functions were developed in the dictionary itself, which were not intended prior to the commencement of the project, e.g. visual section, grammatical notes (gender, number) for the terms in Latvian, Russian, and German, the interface in English, Russian, and German, adding photographs (to be viewed offline), external references (or hyperlinks) to the sources used, and internal references (or hyperlinks) inside the dictionary: from terms to interactive pictures (in the section *Interactive Images*) and vice versa, developing and adding the educational games to learn terms, developing and updating the function for semantic search of terms (reflecting the link between the terms indicating the taxonomic category in the dictionary). Besides the work already completed, technical and content-oriented additions to the electronic dictionary should be constantly made, since “dictionaries are not static but dynamic works” (Fuertes-Olivera & Tarp, 2014, p. 243).

It should be noted that the structural and functional testing of the mobile application prototypes is essential for the future improvement of the final version of the dictionary, so that the developed dictionary could be issued as an application for the wider public. The experience acquired may be useful in compiling other small-volume dictionaries, e.g., in developing an electronic dictionary (as a mobile application) of zoology, chemistry, or physics terms. Such dictionaries would be particularly convenient and useful for students and pupils studying specialised subjects at school or university. The development of electronic dictionaries never stops and never ends – it is an everlasting work in progress. The author therefore invites readers to get to know and evaluate this specialised dictionary, which has been available for download on the Google Play and the Apple Store since the beginning of 2024.

Acknowledgement

This study was funded by the Latvian Council of Science, project ‘Smart Complex of Information Systems of Specialised Biology Lexis for the Research and Preservation of Linguistic diversity’, No. lzp2020/10179. The author thanks Karina Šķirmante, the leading programmer, Ruslans Smolonskis and Niklāvs Bergs, programmers, for the technical solutions in the development of the dictionary. The students of the Faculty of Translation Studies and of the Faculty of Information Technologies also deserve thanks for participating in the project (their names are listed in the dictionary described in this article). Thanks to Natalja Malašonoka for improving the English language of the manuscript.

Conflict of Interest

The authors declare no conflict of interest regarding the publication of this article.

References

1 Balode, I. (2012). Vācu-latviešu un latviešu-vācu leksikogrāfija (1991-2010). In A. Lauzis (Ed.), Vārdnīcu izstrāde Latvijā 1991-2010. (pp 16-61). Rīga: Latviešu valodas aģentūra.

2 Barz, I., Bergenholtz, H., & Korhonen, J. (2005). Schreiben, Verstehen, Übersetzen, Lernen. Zu ein- und zweisprachigen Wörterbüchern mit Deutsch. Frankfurt am Main: Peter Lang.

3 Bergenholtz, H., & Tarp, S. (1995). Manual of specialised lexicography: the preparation of specialised dictionaries. Amsterdam and Philadelphia: John Benjamin. https://doi.org/10.1075/btl.12

4 Bergenholtz, H. (2001). Lexikografie ist Selektion ist Selektion ist Selektion.... In J. Korhonen (Ed.), Von der mono- zur bilingualen Lexikografie für das Deutsche. (pp. 11-30). Frankfurt am Main: Peter Lang.

5 Bergenholtz, H., Nielsen, S. & Tarp, S. (Eds.) (2009). Lexicography at a Crossroads: Dictionaries and Encyclopaedias Today, Lexicographical Tools Tomorrow (Linguistic Insights). 1st ed. Bern: Peter Lang

6 Ēdelmane, I. (1991). Latvisko augu nosaukumu bagātība un zinātniskā terminoloģija. In E. Melbārzde, D. Guļevska (Eds.), Latviešu valodas kultūras jautājumi. 26. (pp. 156-161). Rīga: Avots.

7 Fuertes-Olivera, P. A., & Bergenholtz, H. (Eds.) (2011). E-Lexicography: The Internet, Digital Initiatives and Lexicography. London, New York: Continuum.

8 Fuertes-Olivera, P. A., & Tarp, S. (2014). Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography. Berlin and Boston: De Gryuter. https://doi.org/10.1515/9783110349023

9 Gledhill, D. (2008). The Names of Plants. 4th Ed. Cambridge et al.: Cambridge University Press. https://doi.org/10.1017/CBO9780511550898

10 Granger, S., & Paquot, M. (Eds.) (2010). eLexicography in the 21st century. New challenges, new applications. Proceedings of eLex2009. Louvain-la-Neuve: Presses universitaires de Louvain.

11 Hartmann, R. R. K. (2001). Teaching and Researching Lexicography. Harlow: Longman-Pearson Education.

12 Hartmann, R. R. K. & James, G. (2002). Dictionary of Lexicography. Taylor & Francis e-Library https://doi.org/10.4324/9780203017685

13 Jessen, H. & Schulze, H. (2008). Botanisches Wörterbuch für Gärtner und Floristen mit über 2000 Namen. 24., unveränd. Aufl. Hannover: Verlag M. & H. Schaper

14 Jansone, I., & Vulāne, A. (2018). Latviešu valodas rokasgrāmata. Retrieved December 2023 from www.valodasrokasgramata.lv

15 Jērāne, S. (2017). Vārdnīcu struktūras koncepcijas: pamatprincipi un varianti. In A. Kalnača & I. Lokmane (Eds.) Res Latvienses. IV. Semantika. Sintakse. Valodas kultūra. Inta Freimane. Scriptum Festivum. (pp. 74-85). Retrieved December 2023 from https://www.lu.lv/apgads/elektroniskie-izdevumi/zurnali-un-periodiskie-izdevumi/res-latvienses/res-latvienses-4/ https://doi.org/10.22364/RL.4.7

16 Kaufmann, U. (1998). Wörterbücher mit und ohne Zusammenarbeit zwischen Lexikographen, Sprach- und Fachexperten. In J. E. Mogensen, V. H. Pedersen, & A. Zettersten (Eds.), Symposium on Lexicography VIII. Lexicographica Series Maior. Bd. 90. (pp. 199-207). Tübingen: Niemeyer. https://doi.org/10.1515/9783110929836-015

17 Müller-Spitzer, C. & Möhrs, C. (2010). The “Online Bibliography of Electronic Lexicography”. In S. Granger, M. Paquot (Eds.), eLexicography in the 21st century. New challenges, new applications. Proceedings of eLex2009. (pp. 439-444). Louvain-la-Neuve: Presses universitaires de Louvain.

18 Müller-Spitzer, C. (2013). Textual structures in electronic dictionaries compared with printed dictionaries: A short general survey. In R. H. Gouws et al. (Eds.). Dictionaries: An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with a Focus on Electronic and Computational Lexicography. (pp. 367-381). Berlin & Boston: De Gruyter Mouton. https://doi.org/10.1515/9783110238136.367

19 Nesi, H. (2000). Electronic Dictionaries in Second Language Vocabulary Comprehension and Acquisition: the State of the Art. In U. Heid, S. Evert, E. Lehmann & C. Rohrer (Eds.). Proceedings of the Ninth International Congress, EURALEX 2000, Stuttgart, Germany, August 8th-12th. (pp. 839-847). Stuttgart: Institut für Maschinelle Sprachverarbeitung.

20 Purviņa, R. (1973). Terminoloģija 9. Agronomijas terminu vārdnīca / Словарь терминов агрономии [Dictionary of Agronomy Terms]. Rīga: Zinātne.

21 Roze, L. (1982). Pasaule vārdnīcas skatījumā. Rīga: Zinātne.

22 Schryver de, G.-M. (2003). Lexicographers' Dreams in the Electronic-Dictionary Age. International Journal of Lexicography, 16(2), 143-199. https://doi.org/10.1093/ijl/16.2.143

23 Svensén, B. (2009). A Handbook of Lexicography: The Theory and Practice of Dictionary-Making. Cambridge, New York: Cambridge University Press.

24 Spies, G. (1982). Tier- und Pflanzennamen in der allgemeinsprachlichen Lexikographie. In E. Agricola, J. Schildt & D. Viehweger (Eds.), Wortschatzforschung heute: altuelle Probleme der Lexikologie und Lexikographie. (pp. 221-235). Leipzig: Enzyklopädie.

25 Sviķe, S. (2015). Augu nosaukumu atlase vidēja apjoma vispārīgajai vārdnīcai. In J. Baldunčiks & A. Veisbergs (Eds.) Vārdnīcas un valoda. Valsts valodas komisijas raksti. 7. (pp. 100-121). Rīga: Zinātne.

26 Sviķe, S. (2016). Speciālā leksika vispārīgajās divvalodu tulkojošajās vārdnīcās: augu nosaukumi / Special lexis in General Bilingual Dictionaries: Plant Names. Promocijas darbs / Promotional Paper. Ventspils: Ventspils University College.

27 Sviķe, S. (2017). A New Dictionary of Botanical Terms: an Insight into the Compilation Project of the Dictionary. In J. Titko (Ed.) Emerging Trends in Economics, Culture and Humanities (etECH2017). Conference Proceedings. (pp. 22-30). Rīga: The University College of Economics and Culture.

28 Sviķe, S. (2018). A New Dictionary of Botanical Terms: Data Analysis of a Lexicographic Survey. Economics World. David Publishing Company Journal. May-June 2018, 6 (3). (pp. 228-241). Retrieved December 2023 from https://davidpublisher.com/Public/uploads/Contribute/5abe0a970bfe5.pdf

29 Sviķe, S., & Šķirmante, K. (2019). Practice of Smart LSP Lexicography: The Case of a New Botanical Dictionary with Latvian as a Basic Language. In I. Kosem et. al. (Eds.) Electronic lexicography in the 21st century. Proceedings of the eLex 2019 Conference, 1-3 October 2019, Sintra, Portugal. (pp. 1-17). Lexical Computing CZ.

30 Sviķe, S. (2022). Survey analysis of dictionary-using skills and habits among translation students. In. A. Klosa-Kückelhaus et al. (Eds.) Proceedings of XX EURALEX Congress: Dictionaries and Society (pp. 494-508). Mannheim: IDS-Verlag. https://doi.org/10.14618/phpy-6r66

31 TTC (2007). Tulkošanas un terminoloģijas centrs. Latviešu valodas terminoloģijas resursu kvalitātes un pieejamības apzināšana dažādās zinātnes un praktiskās darbības nozarēs. Rīga: Tulkošanas un terminoloģijas centrs.

32 Vachková, M. (2011). Das große akademische Wörterbuch Deutsch-Tschechisch. Ein erster Werkstattbericht. Frankfurt am Main et. al.: Peter Lang.

33 VPSV (2007). Valodniecības pamatterminu skaidrojošā vārdnīca [Explanatory Dictionary of Basic Linguistic Terms]. Rīga: LU Latviešu valodas institūts.

34 Давыдов, Н. Н. (2013). Ботанический словарь. Русско-английско-немецко-французско-латинский: Ботаника, словарь. Москва: Книга по Требованию.

Santrauka

Silga Sviķe

Elektroninio botanikos terminų žodyno parengimo projekto patirtis

Straipsnis suteikia galimybę susipažinti su specialiųjų botanikos terminų žodyno kūrimu. Tai daugiakalbis (latvių, lotynų, anglų, vokiečių, rusų) vertimo žodynas, sukurtas kaip mobilioji programėlė. Tyrime aprašyta žodyno koncepcija, žodyno kūrimo modelis, projekto įgyvendinimo sunkumai bei sprendimai. Trečioji arba galutinė straipsnyje aprašyto žodyno versija buvo parengta įgyvendinant projektą „Viedais bioloģijas speciālās leksikas informācijas sistēmu komplekss lingvistiskās daudzveidības saglabāšanai un pētniecībai“ („Išmanusis specialiosios biologijos leksikos informacinių sistemų kompleksas kalbinei įvairovei išsaugoti ir tirti“), Nr. lzp- 2020/1-0179, ir 2024 m. sausio mėn. paskelbta *Google Play Store* ir *Apple Store*.

Tyrime aprašyti įvairūs terminologijos žodyno kūrimo koncepcijos ir praktinio darbo eigos aspektai. Aprašyti sunkumai, su kuriais susidūrė žodyno sudarytojai, bei žodyno mega-, makro- ir mikrostruktūroje įgyvendinti sprendimai. Prieš pradedant projektą buvo atlikta apklausa, siekiant išsiaiškinti potencialių vartotojų pageidavimus. Į juos buvo atsižvelgta rengiant žodyną, pavyzdžiui, rasti nauji sprendimai dėl botanikos terminų žodynui būdingų aspektų – taksonominių kategorijų žymėjimo, taksonų (pvz., genties, rūšies, veislės) vaizdavimo, automatinės apibrėžimų paieškos. Straipsnio pabaigoje apibendrintos pagrindinės išvados, į kurias reikėtų atsižvelgti rengiant nedidelį (apie 3 000 terminų) terminų žodyną. Jos galėtų būti naudingos leksikografams, terminologams ir vertėjams rengiant specializuotos leksikos žodynus.

About the Author

SILGA SVIĶE

Assistant Professor, Researcher at Ventspils University of Applied Sciences, Latvia

Research interests
Lexicography, terminology and translation

Address
Inženieru 101a, Ventspils, LV-3601, Latvia.

E-mail
silga.svike@venta.lv

Orcid ID
0000-0001-8374-3237

This article is an Open Access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 (CC BY 4.0) License (http://creativecommons.org/licenses/by/4.0/)