

Entry Word Selection for a Specialised Dictionary: Plant Names

Silga Sviķe
Ventspils University of Applied Sciences
silga.svike@gmail.com

Abstract

In the era of printed dictionaries, the issue of entry word selection was very topical because printed dictionaries could not include all the entry words that the compilers wanted to include. Henning John Bergenholtz (2001: 12–14) analysed 252 dictionary reviews and concluded that 22% of all the remarks were related to the selection of entry words. An analysis of six printed modern Latvian bilingual dictionaries at a 50-year publication interval also shows that they did not contain essential units of the special lexis of botany – plant names that are often used in daily routine. (Sviķe 2016) It seems that modern electronic lexicography and terminography could solve this problem of entry word selection since dictionaries published electronically might include an unlimited number of words. But is this really the case? The process of updating general and specialised dictionaries is also poorly described in existing studies but should be taken into account by each lexicographer. This study deals with the issue of selecting entry words for a specialised dictionary – mobile app (a multilingual Latvian-Latin-English, German-Russian dictionary of botany terms with Latvian as a main language) and describes the ways the initial prototype of the dictionary (including only 500 entry words (see Sviķe & Šķirmante 2019)) was improved, in particular by analysing the problems of selecting entry words to create a small-volume dictionary of special lexis with about 3,000 entry words.

Keywords: entry word selection, specialised dictionary, plant names, terminography

1. Introduction, Issues, and Objectives

Dictionaries play an important role in every culture. Especially when it comes to specialised (including subject-specific) dictionaries used by students, specialists, and translators. These dictionaries should be particularly useful for understanding terminology in the specific field, thus, the way of selecting terms to be included is particularly important. To be more useful to the target audience, dictionaries should meet users' needs. Achieving this is the task of a lexicographer. Although entry word selection has not been extensively studied, as this subject is quite complex (it is influenced by a particular field, its status in the society, the level of development of society, the skilfulness of the dictionary-maker, the languages

included, and other aspects), the importance of the subject is worth raising the topic. Especially when it comes to the dictionaries limited in scope.

This paper will provide an insight into the practice of terminography by addressing the issue of entry word selection for a specialised dictionary. The aim of the study is to describe, using descriptive and statistical methods, the methods of entry word selection for the newly developed electronic dictionary (mobile app) *The New Botanical Dictionary. Terms in Latvian, Latin, English, German and Russian*, where most of the entry words are plant names. In order to achieve this aim, the following objectives have been set:

1. Provide an insight into specialised dictionaries of botanical terms with Latvian as one of the dictionary languages.
2. Analyse the theory of entry word selection for dictionaries.
3. Characterise the principles and methods of entry word selection for a specialised small-volume dictionary of botany terms, describing the challenges and solutions.
4. Draw conclusions after studying the theoretical material and describing the practical part based on a real dictionary-development project.

Before reviewing the theoretical literature, the next section gives a brief overview of botanical dictionaries in Latvian.

2. An Insight and Status Quo of the Latvian Botanical Dictionaries

The beginnings of the Latvian botanical terminology can be traced back to the second half of the 19th century (Baltiņš 2006: 72) and the contribution of the first Latvian botanist Jānis Ilsters is undeniable. All the plant names (500 species) collected by Ilsters were published in 1884 as *Latviešu botāniski nosaukumi* (Latvian botanical names) in the collection of the Riga Latvian Society Knowledge Commission (in Latvian – *Rīgas Latviešu biedrības Zinību komisija*) Collection of Papers No II, 63-81 and No III 66-74, and already in 1883, the textbook with botany terms *Botānika tautas skolām un pašmācībai. Elementarkurss* (Botany for folk schools and self-study. Elementary course) was published. According to one of the largest dictionary studies in Latvia, 29 different zoology, botany (including names of taxa of organisms) term collections have been published between 1900 and 2007, thus these fields account for 6.9% of term dictionaries as to their frequency and are ranked 4th. (TTC 2007: 13). The botanical dictionary of Pauls Galenieks (*Botaniskā vārdnīca*, 1950) and the publications of the Latvian Academy of Sciences, e. g. the Dictionary of Agronomic Terms (*Agronomijas terminu vārdnīca*, 1973) and others should be mentioned here. For more information on the dictionaries of botanical terms in Latvian, see S. Sviķe & A. Stalažs' article (2019). However, from 2008 to the present day, only two dictionaries containing botanical terms in Latvian or Latvian and Russian (*Bioloģija. Skaidrojošā vārdnīca: mācību palīg līdzeklis* (Biology. Explanatory dictionary: teaching tool) by L. Ģerķe, A. Zvingēvics, 2008, Rīga: RAKA; *Latviešu-krievu, krievu-latviešu dabaszinību vārdnīca skolēniem: bioloģija, ģeogrāfija* (Latvian-Russian, Russian-Latvian dictionary of natural sciences for students: biology, geography) by M. Pupinš, I. Ščemeļova, 2009, Rīga: Zvaigzne ABC) have been recorded in the major lexicographic lists available (see Peina 2023: 156-160). Of course, botanical terms, especially plant names, are also included in general dictionaries, as well as in special encyclo-

paedias, specialised literature (handbooks), and periodicals, e. g. horticultural and floristics magazines. However, one resource would be more useful for translators – the specialised dictionary, where the most important lexis related to the terminology of the specific field can be quickly found. The need for new terminology resources in Latvian has already been demonstrated in several studies (Sviķe 2018; Peina 2022), and the selection of entry words for *The New Botanical Dictionary* is described in more detail in this article.

As in all sciences, the study of the diversity of flora and fauna has evolved worldwide, and new insights into the diversity of organisms and their relationships have been gained over the decades. This has led to changes in the classification of many groups of plants. These changes have also affected many plant groups, especially those less studied in the past. In addition, it should be emphasized that the development of terminology has led to the development of linguistic and terminological traditions in Latvian terminology. Therefore, older sources should be used cautiously since not every plant name is still appropriate, as, for example, half a century ago. Translators, journalists, dictionary makers, and anyone whose work involves plant names must be informed of the latest developments in the field. This is currently ensured by the new terminology management system *Bioleksipēdija* (www.bioleksipedija.lv) which tracks changes in the names of organisms throughout history. (Stalažs et al. 2023) This resource is also used in the entry word selection for the dictionary described in this study. However, before a detailed description of the entry word, the theoretical literature on this topic should be reviewed and considered.

3. Theoretical Background

The term *entry word* in this study means a specific plant name and other subject-specific terms that make up the headword list of the particular dictionary. Entry word selection in dictionaries is not the subject of extensive and numerous studies, neither for general dictionaries nor for specialised ones, but the principles of entry word selection are some of the most important criteria for determining the quality of dictionaries. (Baldunčiks 2012: 116–117; Bergenholtz 2001, 12–14; Peina 2023: 53, 72; Sviķe 2016; Гринев-Гриневич 2007: 23–24) Also, in the front matter of the dictionaries, their authors provide very scarce information about the selection of entry words and other specific explanations.

In the front matter of the dictionaries of Latvian botanical terms, there is also no tradition of providing principles and methods of selecting the words to be included in the dictionary. In addition, the dictionaries also often do not give the number of entry words included. (Sviķe & Stalažs 2019) This is a typical feature of most dictionaries published since the mid-20th century, regardless of the type.

In the compilation of the dictionaries of the 21st century, if the language has a corpus, priority is given to the dictionary based on the corpus (Atkins & Rundell 2008: 3). This is a key principle in the so-called second era of dictionary development after using the entry word card method as in the past. Although currently there is the Latvian text corpora (www.korpuss.lv) available as a digital resource for researching the Latvian language diversity, the data of this general corpus, as the research of Jasmonts et al. (2022) shows, are not useful for studying the special lexis of botany. The practical data of the specialised dictionary compiled and described in this study shows that this general language corpus is not suitable for the selection of botany terms for the purposes of terminography either. Such an aim

would require a specialised corpus consisting only of texts in the specific field of botany. If the compilation of the dictionary does not have access to the corpus data base or the data are not useful, the basic method for entry word selection is using previously issued dictionaries (Bergenholtz 1992, 52; Landau 2001, 193; Цыренов 2013: 18); the word-list of this dictionary should be supplemented and new words should be added, or some words should be deleted if a larger volume is used in the new entry word list of the dictionary. There is currently no special corpus of Latvian botanical terminology available. However, the new multilingual database *Bioleksipēdija*, which contains a special lexicon for biology (more than 600,000 linked terms), proved to be useful in the process of selecting the terms for a botany dictionary.

The instructions for the selection of German-Malagasy dictionary entry words, where the corpus data is used, are described by Bergenholtz (1992: 49–59). In turn, working without access to the language corpus is analysed by Regina Hessky (1996: 5–20), as she describes the selection of words for a general German-Hungarian dictionary, where a general monolingual dictionary of the German language is taken as a basis that is of a larger volume than a bilingual one, and is reduced in conceptual circles in direction from outside to the centre, according to the hierarchic criteria. Looking further at the principles of creating small-volume printed general dictionaries, one must conclude that other small-volume dictionaries were used, but the choice of topics to be included in the dictionary depended on the local needs of the Latvian society (Soikane-Trapāne 1985: 389) because this dictionary was published outside Latvia. In the Soikane-Trapane's dictionary, the thematic groups of the words selected by the frequency criteria are supplemented by the principle of logics, only very few peripheral word groups are included (Soikane-Trapāne 1985: 390–392).

Also, the Czech linguist Marie Vachkova (2011: 31–33) describes the selection of entries for the large academic German-Czech dictionary *Großes Akademisches Wörterbuch Deutsch-Tschechisch* by using different tools but with the universal German dictionary as the source dictionary, supplementing its entries by using the corpus and Internet sources. The lists of special parts of the lexis – terms – were processed separately from the general list of vocabulary used in the dictionary, and some experts in particular fields were involved in the process of compiling. For the search of Czech equivalents, the native language sources were used, as well as other modern German-Czech dictionaries. The selection of specific vocabulary was based on the theory of centre–periphery developed by Josef Filipec, also the experts in the specific fields were involved and additional terminological sources were studied. Vachkova emphasises that special lexis was selected by taking into account the needs of Czech addressee (Vachkova 2011, 90). Filipec distinguishes five groups for a central–peripheral division of lexical units: centre, transition period group, periphery, a group of integral margins, as well as an outside system that already has other systems (Filipec 1996: 23–42). However, these entry word selection descriptions are more relevant to general dictionaries.

The main criterion for selecting keywords for a specialised dictionary is the evaluation of the importance of terms, which is based on the concept of relevance described by Pedro A. Fuertes-Olivera and Sven Tarp (2014: 204). Term selection uses the so-called knowledge of subject field experts, which often turns out to be much more useful in practice compared to the selection of commonly used terms. In particular, the concept of the importance of terms can be applied directly to special dictionaries and the selection of entry words in a specialised dictionary. Fuertes-Olivera & Tarp (2014: 204) point out that the most fre-

quently used words (terms) are not useful and essential in the development of specialised dictionaries, but more attention should be paid to the selection of the most important words (terms), focusing especially on the potential user of the dictionary being developed and his specific needs.

Bergenholtz admits that no theoretical and methodological approaches have been developed regarding the functions of different types of dictionaries, and the so-called *external* (entry word) selection and *internal* (information selection) can be distinguished. (Bergenholtz 2001: 12-14) This distinction also proved useful in this study and is described below according to this division.

In summary, the question of selecting a special lexis is not studied enough and remains one of the main problems in lexicography and terminography. Various aspects, such as the addressee of the new dictionary, the specificity of the field, the content and function of existing dictionaries, should definitely be taken into account.

4. Methodology and Selection of Entry Words

Regarding the development of any dictionary, it should be concluded that determining the target group of users or potential addressees of the dictionary is one of the basic requirements that must be taken into account even before planning a dictionary project (Barz, Bergenholtz & Korhonen 2005: 15, Fuertes-Olivera & Tarp 2014: 45-57). Specialised dictionaries, unlike general dictionaries, reflect a certain set of lexical units selected on the basis of specific principles. These principles depend heavily on the potential dictionary user's wishes and preferences as well as on the function of the dictionary. A survey of potential users was carried out prior to the development of the described dictionary (Sviķe 2018), and the comments and wishes of potential users have been taken into account as far as possible. In the analysis of specialised bilingual dictionaries, Sandro Nielsen divides the terminographic selection process into five stages: method selection, domain selection, selection of data or information material in the working languages of the dictionary, entry word selection from source language data material, and selection of entry word equivalents from target language data material (Nielsen 1994: 130). For the selection described in this study, a combined data selection method is used, where the data is selected according to the encyclopaedic approach. The choice of the field is determined by the terminology field of the dictionary – botany. As the entry word selection is based on the basic plant name list already developed, its extension as an external selection and a selection of additional information (as an internal selection) are described below. The principles of the external selection of the entry can be divided into two basic steps:

1. a selection of terminographic sources (for the main dictionary language and equivalents);
2. a selection of sources for additional micro-structural information (e. g. definitions, examples, hyperlinks).

The resources used can be divided into print resources as well as digitised ones, including “using the internet as a complementary tool (to be handled with care)” (Scholze-Stubenrecht 2013: 1373), also the types of texts that are most frequently used in practical terminographical work, and those referred to by Stolze-Stubenrecht (2013: 1366–1368),

which often helps in making important decisions about the selection of the relevant information. The selection of entry words in the dictionary is based on approx. 90 sources (Sviķe et al. 2024), including reference literature (specialised encyclopaedias, dictionaries, etc.), botany textbooks and other special publications, popular and popular science literature, and the newly developed terminology management system *Bioleksipēdija*, the usefulness of which is described in the subchapters of this paper.

4.1 External Selection

A list of 500 plant names (500 units), developed for medium-sized general bilingual dictionaries (Sviķe 2015), is used as the basis for the first prototype of *The New Botanical Dictionary. Latvian-Latin-English-German-Russian Terms*. The first prototype of the botanical dictionary was prepared in 2017 (Sviķe & Šķirmante 2019), then the prototype of the mobile application was updated, and the entry word list for the dictionary was expanded to 890 words in 2018; in 2023 the final version of the dictionary was finalised for publication with about 3,000 entry words in Latvian providing their equivalents in Latin, English, German, and Russian. 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’, No. lzp-2020/1-0179, and published in the Google Play Store and the Apple Store in January 2024.

An encyclopaedic approach and a combined method of compiling a list of the plant names has been used in general: the excerption of the most common and important plant names from the lists in encyclopaedias, dictionaries, special literature, textbooks, and internet sources, as well as consulting field experts (Baldunčiks 1982: 98).

The selection of representatives of plants as a special botanical lexis and their names to be included in the general bilingual dictionary has a three-stage structure:

- 1) selection of plants (referents¹),
- 2) selection of plant names denoting the referents,
- 3) selection of equivalents in the target language. (Sviķe 2016: 68)

In the process of compiling the basic list of plant names (500 units), the selection was mostly conducted without a direct reference to one language, thus it can be relatively divided into two actions:

- 1) selection of plants to be represented in the optimum list of plants, i. e. plant names, and
- 2) selection on the right side of the dictionary focusing on plant names in the Latvian language that are mostly terminological.

The onomasiological approach used to create this list is to move from the referent to its names. It should be noted here that, unlike a general dictionary, which includes a large range of synonyms for a given plant name (but this depends largely on the size of the dictionary, the frequency of use of these synonyms, and other considerations), a terminological dictionary includes only scientific names and terminological plant names in dictionary

1 “specific realia which in the specific meaning is named by a language unit” (VPSV 2007: 321)

languages. This particular dictionary also attempts to find a compromise and, in some cases, adds up to two synonyms, which are very popular (accompanied by the appropriate label).

Therefore, first and foremost comes the representation of plants, next is the selection of equivalents in Latvian with their scientific or Latin plant name. To adapt this basic list to a multilingual botanical dictionary, special terminographic sources in all dictionary languages (English, German, Russian) were selected. The criteria for their selection are works by reliable authors, which are large enough when selecting the material for a small-volume dictionary and they should be really reliable resources, consulted by experts in the field. For German it is basically *Botanisches Wörterbuch* by R. Schubert & G. Wagner (2000) and its newest editions, *Zander: Handwörterbuch der Pflanzennamen* by W. Erhardt et al. (2014); for English it is *New Flora of British Isles* by C. A. Stace (2010) and for Russian it is *Научные и народные названия растений и грибов. Научно-популярное издание* by И. В. Бугаев (2010). The vernacular names and scientific or Latin names are checked using the systematic tree of organisms of the *Bioleksipēdija* (2024), which allows to verify the exact Latin name of the taxon. Various resources available on the internet were certainly used, in particular to obtain the latest and most up-to-date information; a Google search was used to check the statistics of the plant names used.

Since plant names are characterised by a large diversity of vernacular names, one of the problems was choosing the terminological and only one or two of the most popular vernacular names, if any. Primarily, the terminological names were chosen as the first ones, but in some cases also popular vernacular names of frequent use were added, e.g. for *Leucanthemum vulgare* the Latvian terminological name *parastā pīpene* and also *margrietiņa* as a second popular name with the label *pop.* (for – ‘popular’). Google search shows the following results for the frequency of use: 3,130 results for *parastā pīpene*, but 66,100 for *margrietiņa*; so *margrietiņa* is used 20 times more frequently. The English, German, and Russian names have been chosen on the basis that they are the first to appear in at least two authoritative specialised sources selected. If not, an expert in the field has been called in to confirm which should be chosen as the first equivalent.

As the basic list of plant names is mostly plant genus names (as they are used in common language and compiled for a general dictionary), the second version of the dictionary was based on adding plant genus names until the basic list consisted of 890 entry words with equivalents in all dictionary languages. The same principles as for the basic list of plant names are used here (see Sviķe 2016). For the third and final version (up to 3,000 entry words), however, the names of plant species and cultivars were also included, forming a more specific part of the botanical lexicon, away from the common lexicon. According to Filipec, approach selection is made in concentric circles (see Fig. 1) from the centre (names of plant genus) to the middle circle and periphery (names of plant species and cultivars). The main and initial criterion for the external selection (up to a maximum list of 3,000 entry words) was to select the major diversity of the referents denoted by the plant species and also cultivars – herbaceous plants, bushes, trees, lianas and mixed life-form plants – for the selection of their names to include in the dictionary.

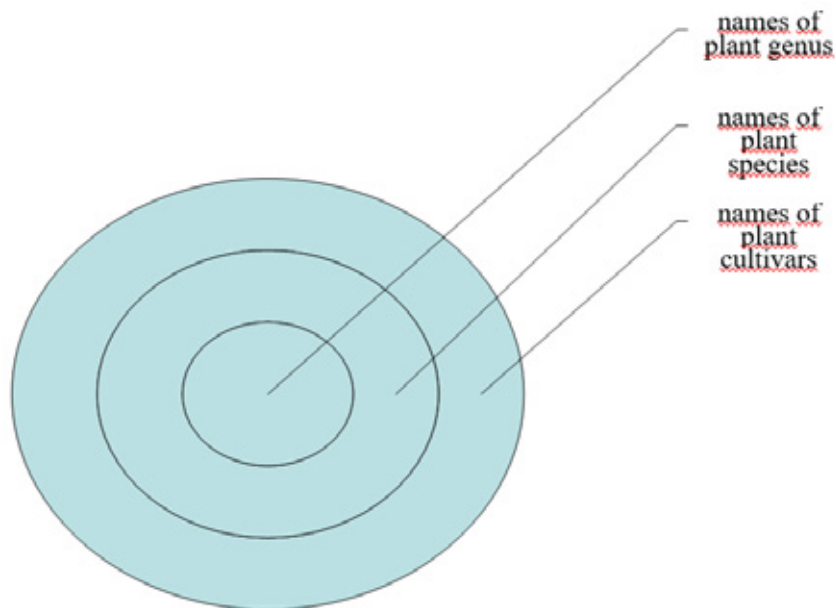


Figure 1: External selection according to Filipec's centre-periphery theory

One of the biggest problems in selecting plant species names and, in particular, cultivar names, was that many plants have a disproportionately high diversity. It was found, for example, for cultivar names of genus *Hosta*, *Hemerocallis*, *Paeonia*, *Phlox*, *Thuja*, *Berberis*, and *Acer*. The cultivars of these plant genus are characterised by their great diversity, which means that their names must be carefully selected for the dictionary. Only cultivar names whose referents differ in colour, plant height, flowering time, and other aspects have been selected for inclusion in the dictionary. Initially, a very large variety of cultivars was selected, but due to disproportionality, however, a large number of the selected cultivar names were deleted, leaving a maximum of 20 different cultivar names denoting as many different and distinct referents as possible. Table 1 shows the percentage of the change in the number of genus, species, and cultivar names included in the basic list and the final version of the dictionary. Table 1 shows that the final version includes much more specific cultivar and species names due to the wish of potential users: that one of the dictionary functions should be a selection of specific plants according to certain characteristics – flower or fruit colour, plant height, life-form, etc., as provided by the filter search function in the dictionary.

Basic list of plant names (500 plant names)	Final list of entry words for the New Botanical Dictionary (about 3,000 plant names)
plant genus names – 90% plant species names – 10%	plant genus names – 14% plant species names – 35% plant cultivar names – 45% names of other taxa and botany terms – 6%

Table 1: Selected taxon names in the basic list and final list of entry words

The selected botany terms are represented by interactive images containing only a few basic terms of plant structure, so specific botanical terms are not included in the dictionary. In the external selection phase, it should be noted that the resource *Bioleksipēdija* was very useful to check for errors in scientific names, both in the specialist literature and in the educational literature. The *Bioleksipēdija* provides excerpts from the latest scientific literature on the names of organisms, as it is possible to view the names in a timeline. Although errors have been corrected in the external selection process (e. g. specification of species names, correction of grammatical gender for plant names in Latvian, German, and Russian), some plant groups have not yet been inspected. This is the case for Iris, where no revision has been carried out and the plant names of this group are included in the dictionary as given in the contemporary plant catalogues. Since the online encyclopaedia in the Latvian language *Latvijas daba* (www.latvijasdaba.lv) contains mainly wild plant names, this botanical dictionary includes a limited number of wild plant names, but focuses more on other groups, such as ornamental plants and indoor plants.

4.2 Internal Selection

This subsection describes the information selection issues and solutions that make up the microstructure of the dictionary. The main language of the dictionary is Latvian, so most of the information provided in the dictionary is in Latvian, and some sections are only available in Latvian. The microstructure of the dictionary consists of the parts given further, and it includes information that can be divided into three groups: information relating to entry words and their equivalents (1), information about the referents denoted by entry words (2), and additional information – a definition or a link to an external resource with an additional description of the referent (3).

A search word or words (plant name) and its equivalents in other languages; grammatical references (gender and number) for German, Latvian, and Russian equivalents. Feminine gender: f – *femininum*; masculine gender: m – *maskulinum*, neuter: n – *neutrum*, singular: sg – *singularis*, plural: pl – *pluralis*. Entry words are highlighted with the corresponding colour that indicates the taxonomic category, e. g. genus, species, cultivar. Explanations for the colour markings are given in the dictionary. The Latvian names and some plant names in other languages may have explanations added in parentheses, e. g. the label *pop.* (from Latin – *popularis*) with the meaning ‘popularly known’: *ciklamenas* pl, f; *Alpu vijolītes* (*pop.*) pl, f. The German and English plant names may contain indications about regional use – two-letter country codes in accordance with the ISO 3166-1 standard are given in the parentheses: (DE) – mostly in Germany, (AT) – in Austria, (CH) – in Switzerland, (US) – in the United States, and (GB) – in Great Britain. General philological dictionaries and linguistic encyclopaedias were useful for the selection of grammatical labels, while specialised literature and plant name catalogues or databases available on the internet were mainly used to add explanatory notes on the geographical use of plant names.

Pictorial illustrations indicating that a plant belongs to a particular plant group: an indication of life-form (e. g. herbaceous plant, tree, bush) or a particular use of the plant (e. g. nectar plant, aquatic plant, or indoor plant); pictorial illustration indicating an audio file with a description of the plant, e. g. the entry *Quercus robur* or *Tilia cordata*, pictorial illustration indicating an image of the plant described in the entry. The most useful resources

for selecting this information were the planting guides (e. g. *Apstādījumu veidotāju ceļvedis Latvijas stādaudzētavu sortimentā* (A guide for landscapers in the Latvian nursery assortment) by Anita Neilande as well as Plant identifiers – catalogues published by the Latvian Fund for Nature (e. g. *Kas aug dabiskās pļavās?* (What grows in natural grasslands?)), which give detailed descriptions of the plants themselves.

There are two drop-down menus in the entry: *Tēzaurs* and *Information*. *Tēzaurs* drop-down menu contains data that is automatically retrieved from the online electronic resource *Tēzaurs.lv*, e. g. an automatically retrieved definition. There is a hyperlink to the particular section on *www.tezaurs.lv* (online access). The drop-down menu *Information* includes the definition provided by the dictionary author or other additional information, e. g. a link to the image and description of the search word on the website *www.wikipedia.com*. Most of the information provided there is in English (online access). (Sviķe et al. 2024)

Although most of the necessary additional explanatory and grammatical information could be found in the selected sources used in the process of the dictionary compilation, the definitions were the most difficult for the internal selection. The difficulties were encountered in word and information selection with examples (e. g. choosing equivalent variants, correcting automatically retrieved definitions), and the solutions for compiling a small-volume specialised electronic dictionary were needed. The automatically retrieved definitions from *Tēzaurs* (*www.tezaurs.lv*) have many inaccuracies and even errors, so they need to be corrected and aligned with their information layout. (see more Sviķe & Šķirmante 2019: 13–14)

Since artificial intelligence (AI) solutions can also be used in the so-called third generation of lexicography practice, an experiment has been carried out to get the solutions for writing definitions using ChatGPT. For writing the new definitions for plant genera, an attempt was made to use the AI tool ChatGPT 3.5 (free version), which was queried to develop a definition for a given plant genus according to the given criteria. The answer in Table 2 below shows that ChatGPT cannot perform such a specific task – the terminology used in the definition is inaccurate (the underlined words do not exist in Latvian), the definition contains both grammatical and content errors. ChatGPT offered similar definitions with the same errors also for other plant genera – *Crataegus*, *Hedera*, *Buxus*, etc.

Query for ChatGPT 3.5 (OpenAI, 2023)	Answer of ChatGPT 3.5 (OpenAI, 2023)
Uzraksti definīciju augu ģintij “ceriņi” pēc šādas shēmas: augstāka taksona apzīmējums, šīs augu ģints pārstāvju dzīvesforma (piemēram, koks, krūms, kokaugs, liāna, lakstaugs), lapu raksturojums, ziedu raksturojums, augļu raksturojums, specifiskā pazīme, kas atšķir tos no līdzīgajiem. (Translation into English: Write a definition of the plant genus “lilac” according to the following scheme: designation of the higher taxon, life form of the members of this plant genus (e.g. tree, shrub, woody plant, liana, deciduous plant), characteristics of leaves, characteristics of flowers, characteristics of fruits, specific feature that distinguishes them from their counterparts.)	“Ceriņi” ir augu ģints, ko augstākajā taksonomiskajā hierarhijā iekļauj “Ceriņu dzimta” (Syringa). Šī augu ģints pārstāvji ir lieli krūmi vai mazi koki. Lapas ir pretstatēs, vienkāršas, otršķautnes un biezas. Ziedi veido kupenas vai ķēdes formas galotnēs, un tie ir īpaši aromātiski. Katrs zieds sastāv no četrām kāršu segmentiem, četrām krūmītēm un divām darzenēm. Augļi ir kapsulas, kas ietver nelielas sēklas. Specifiskā pazīme, kas atšķir ceriņus no līdzīgiem augiem, ir to raksturīgais ziedu aromāts, kas ir ievērojams un bieži izceļas kā viens no galvenajiem identifikācijas pazīmes. (Translation into English: “Lilacs” is a genus of plants included in the “Lilac family” (Syringa) in the highest taxonomic hierarchy. Members of this plant genus are large shrubs or small trees. The leaves are opposite, simple, secondary, and thick. The flowers form clusters or chains at the apex and are particularly fragrant. Each flower consists of four carpel segments, four racemes, and two tarsels. The fruits are capsules containing small seeds. The specific feature that distinguishes lilacs from similar plants is their characteristic floral scent, which is remarkable and often stands out as one of the main identifying features.)

Table 2: ChatGPT 3.5 definition writing query and answer

However, this proved ineffective, as the Chat GPT model is not trained with special botany texts in Latvian. This method could not be used for this specialized dictionary, so the definitions from *Tēzauris* were retrieved automatically, but edited and improved manually. For comparison, the author’s definition of the genus “*ceriņi*” (lilacs) in Latvian is as follows: *Ceriņi. Olīvu dzimtas (Oleaceae) ģints. Vasarzaļi krūmi. Lapas – ovālas vai sirdsveida. Ziedi – balti, violeti, sārti, dzeltenīgi, smaržīgi. Ziedkopas – konusveida daudzziēdu skaras. Augļi – plakanas, koksnainas pogaļas.* (Sviķe et al. 2024)

Conclusion

Taking into account the fact that the range of dictionaries of special botanical terms in combination of Latvian (English, German, and Russian) and Latin is quite modest, such a dictionary as a mobile application with the possibility to use this resource offline is of great

practical importance. The study describes the entry word selection in the specialised dictionary of botany terms and provides the principles of the selection of words – plant names, as well as offers an insight into the macro – and microstructure of the dictionary. The study showed that despite a certain range of data available in various digital resources for researching the diversity of the Latvian language, such as the Latvian text corpus (www.korpus.lv), they are not useful for terminology research and for dictionary compiling in such a narrow field. (Jasmonts et al. 2022) Before the compilation of this dictionary, the author took into account almost all suggestions and desires of potential users, as the intended user of the dictionary should be strictly taken into account. Nevertheless, this does not exclude the possibility that the dictionary may also be used by a non-expert user who is unknown to the compiler of the dictionary.

The new open-access interactive multifunctional database management system for special lexis of biology under the domain *Bioleksipedijs.lv* (see Stalažs et al. 2023: 52–67) proved to be much more useful in the selection of entry words (frequency of names in a set of specialised excerpts – organism names, exact scientific names, etc.). The study illustrates the difficulties in entry word and information selection with examples (e. g. choosing equivalent variants, correcting automatically retrieved definitions), and describes solutions for compiling a small-volume specialized electronic dictionary. However, as far as the entry word selection is concerned, a fairly precise procedure has to be defined for the inclusion of a set of entry words appropriate for the users of such a small-volume dictionary. This paper studies and promotes the theoretical meaning of selecting special lexis units – plant names and basic botany terms – having a direct link to a terminographical practice for a future compilation of a specialised dictionary. The description of the entry word selection in this article can be considered a sample illustration of selecting entry words for a small-volume specialised electronic dictionary, and might serve as a methodological tool for choosing other large groups of special lexis, e. g. animal names.

References

- Atkins, B. T. S., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Baldunčiks, J. (1982). Anglicismi angļu-latviešu vārdnīcās. In: S. Lagzdina (Ed.) *Latviešu valodas kultūras jautājumi*. 17. Rīga: Avots, 1982, 96–100.
- Baldunčiks, J. (2012). *Vārdnīcu izstrāde Latvijā (1991-2010)*. Pētījums J. Baldunčika vadībā. Rīga: LVA.
- Baltiņš, M. (2006). Latviešu terminoloģijas attīstība: mantojums, problēmas un perspektīvas. In *Letonikas pirmais kongress. Valodniecības raksti*. Rīga: Latvijas Zinātņu akadēmija, pp. 72–79.
- Barz, I., Bergenholtz, H. & Korhonen, J. (2005). *Schreiben, Verstehen, Übersetzen, Lernen. Zu ein – und zweisprachigen Wörterbüchern mit Deutsch*. Frankfurt am Main: Peter Lang.
- Bergenholtz, H. J. (2001). Lexikografie ist Selektion ist Selektion ist Selektion. In J. Korhonen (Hrsg.) *Von der mono – zur bilingualen Lexikografie Für das Deutsche*. Frankfurt am Main u.a.: Peter Lang, pp. 11–30.

- Bergenholtz, H. J. (1992). Lemmaselektion in zweisprachigen Wörterbüchern. In G. Meder, A. Dörner (Eds.). *Worte, Wörter, Wörterbücher. Lexikographische Beiträge zum Essener Linguistischen Kolloquium*. Tübingen: Niemeyer, S. 49–65.
- Filipec, J. (1996). Probleme des Sprachzentrums und der Sprachperipherie im System des Wortschatzes. In J. Filipec (Ed.). *Studia lexicologica*. München: Otto Sagner, S. 23–42.
- Fuertes-Olivera, P. A., Tarp, S. (2014). *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminography*. Berlin/Boston: Walter de Gruyter.
- Hessky, R. (1996) Das deutsch-ungarische Wörterbuchprojekt. In: R. Hessky (Ed.). *Lexikographie zwischen Theorie und Praxis*. Tübingen: Niemeyer, S. 5–20.
- Jasmonts, G., Sviķe, S. & Šķirmante, K. (2022). New information extracting and analysis methodology for the terminology research purposes: The field of biology. In *CEUR Workshop Proceedings. Vol. 3160. IRCDL 2022: 18th Italian Research Conference on Digital Libraries*. Accessed at: <https://ceur-ws.org/Vol-3160/paper2.pdf> [06/01/2024].
- Landau, I. S. (2001). *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- OpenAI. (2023) *ChatGPT*. Large language model. Accessed at: <https://chat.openai.com/> [22/10/2023].
- Peina, E. (2023). *Mūsdienu latviešu terminogrāfijas teorētiska un pragmatiska analīze / Theoretical and Pragmatic Analysis of Modern Latvian Terminography*. PhD thesis. Liepāja: Liepāja University, Latvia.
- Peina, E. (2022). Mūsdienu latviešu terminogrāfijas nākotne: būt vai nebūt? [The Future of Latvian Terminography: To be or not to be?] In *Scriptus Manet*, Nr. 14. Accessed at: <https://dom.lndb.lv/data/obj/1289751.html> [07/01/2024]
- Scholz-Stubenrecht, W. (2013). The World Wide Web as a Resource for Lexicography. In R. H. Gouws, U. Heid, W. Schweickard and H. E. Wiegand (Eds.) *Dictionaries. The International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*. Berlin, New York: Walter de Gruyter. Pp. 1365– 1374.
- Soikane-Trapāne, M. (1985). *Latviešu valodas pamata un tematisks vārdu krājums*. Lincoln: Augstums Printing Service Inc.
- Stalažs, A., Šķirmante, K., Sviķe, S., Jasmonts, G. & Ziediņš, R. E. (2023). Experience of design and development of a new open access interactive multifunctional database management system for special lexis of biology. In *Studies about Languages / Kalbu studijos*, 42. Kaunas: Kaunas University of Technology, pp. 52–67.
- Sviķe, S. (2015). Augu nosaukumu atlase vidēja apjoma vispārīgajai vārdnīcai. In: J. Baldunčiks, A. Veisbergs (Eds.). *Vārdnīcas un valoda. Valsts valodas komisijas raksti*, 7. Rīga: Zinātne, 100–121.
- Sviķe, S. (2016). *Speciālā leksika vispārīgajās divvalodu tulkojošajās vārdnīcās: augu nosaukumi / Special lexis in general bilingual dictionaries: plant names*. Summary of Promotional Paper, Ventspils: Ventspils university of Applied Sciences.
- Sviķe, S. (2018). A New Dictionary of Botanical Terms: Data Analysis of a Lexicographic Survey. In *Economics World*. David Publishing Company Journal. May-June 2018, 6 (3). pp. 228–241.
- Sviķe, S., Šķirmante, K. (2019). Practice of Smart LSP Lexicography: The Case of a New Botanical Dictionary with Latvian as a Basic Language. In *Electronic lexicography in the*

- 21st century. *Proceedings of the eLex 2019 conference*. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o., pp. 1–17.
- Sviķe, S., Stalažs, A. (2019). Jaunās botāniskās vārdnīcas mikrostruktūra: tradicionālais, mainīgais un inovatīvais. In *Vārds un tā pētīšanas aspekti*: 23, Liepāja: LiePA, pp. 418–429.
- Sviķe, S. et al. (2024). *Jaunā botāniskā vārdnīca. Termini latviešu-latīņu-angļu-vācu-krievu valodā/The New Botanical Dictionary. Latvian-Latin-English-German-Russian terms/Neues Botanisches Wörterbuch. Lettisch-Latein-English-Deutsch-Russisch/Новый ботанический словарь. Латышско-латинско-английско-немецко-русские термины*. Ventspils: Ventspils Augstskola. Accessed at: <https://play.google.com/store/apps/details?id=lv.venta.jbv> [10/01/2024].
- Vachkova, M. (2011). *Das große akademische Wörterbuch Deutsch-Tschechisch. Ein erster Werkstattbericht*. Frankfurt am Main u. a.: Peter Lang.
- VPSV (2007). *Valodniecības pamatterminu skaidrojošā vārdnīca*. Rīga: Valsts valodas aģentūra, LU Latviešu valodas institūts.
- Гринев-Гриневич, С. (2007). О принципах создания терминологических словарей. In *Terminologija*, Nr. 14. Vilnius: Lietuvių kalbos institutas, 2007, с. 20–36.
- Цыренов, Б.Д. (2013). *Монголызычно русская лексикография (структура, принцип, семантизация)*. Улан-Едэ: Бэлиг.

Acknowledgement

This research has been funded by the Latvian Council of Science. The project ‘Smart complex of information systems of specialised biology lexis for the research and preservation of linguistic diversity’, No. Izp-2020/1-0179. The author thanks Natalja Malašonoka for improving the English language of the manuscript.