Acta Universitatis Sapientiae, Philologica, **15**, 2 (2023) 166–181

DOI: 10.2478/ausp-2023-0023

Research Report: Termini Online Hungarian Dictionary and Database (TOHDD): A Dictionary for Hungarian Varieties Spoken in the Carpathian Basin[1]

Tibor M. PINTÉR
Károli Gáspár University of the Reformed Church in Hungary (Budapest, Hungary)
Department of Hungarian Linguistics
m.pinter.tibor@kre.hu
https://orcid.org/0000-0002-5212-4107

Katalin P. MÁRKUS
Károli Gáspár University of the Reformed Church in Hungary (Budapest, Hungary)
Department of English Linguistics
p.markus.kata@kre.hu
https://orcid.org/0000-0001-6826-7631

Attila BENŐ
Babeş–Bolyai University (Cluj-Napoca, Romania)
Department of Hungarian and General Linguistics
attilabe@yahoo.com
https://orcid.org/0000-0002-6931-2999

Abstract. The Termini Online Hungarian–Hungarian Dictionary and Database describes the lexicon of the Hungarian language as spoken in the countries neighbouring Hungary. It is considered to be a general dictionary of present-day Hungarian. Each entry contains authentic example sentences to illustrate the use of the headword, making it possible to examine the special use of a word or construction in a grammatical and pragmatic context. The lexicographical database is edited online in eight countries. The editors of the dictionary are members of the Termini Hungarian Language Research Network. Online editing makes it possible for the dictionary to expand – even simultaneously – as a result of activity in eight countries. In the present study, the authors review the novelties and peculiarities of the dictionary in some detail, touching on the following topics: dictionary structure, IT support, database character, multimedia elements, and labelling system.

Keywords: Hungarian language, Termini, dictionary, lexicography

1 Work on this paper was supported by the Domus Hungarica Scientiarum et Artium Programme of the Hungarian Academy of Sciences and a research grant obtained from the Faculty of

Introduction

The linguistic communities of minority Hungarians beyond the present-day borders of Hungary came into being in 1920, when significant parts of the Hungarian language area were annexed to newly formed non-Hungarian states as a consequence of the Treaty of Trianon (1920). The number and share of Hungarian-speaking people in some of the countries neighbouring Hungary are still significant this day: 1,248,623 in Romania (2011) (6.2% at the state level, 17.9% in Transylvania) (Péntek and Benő 2020: 59, 61), 458,467 (8.5%) in Slovakia (2011) (Horony 2016: 47), 254,000 (3.9%) in Serbia (2011), about 141,000 in Ukraine (2015) (11.28% in Subcarpathia)[2] (Kapitány 2015: 234–236). The number of Hungarian speakers is around 10,000 in Austria (in Burgenland and in Vienna, the latter being home to Hungarian immigrants), 14,408 in Croatia, and 5,544 in Slovenia (Kapitány 2015: 236–237). As a consequence of the minority situation, most Hungarian speakers living in these seven countries are bilingual: they speak Hungarian as their mother tongue and also the official language of the state, which is the native language of the majority group (Romanian, Slovak, Ukrainian, Serbian, etc.). Their active bilingualism creates a language contact situation which makes the lexicons of these varieties special, affected by language contact phenomena, a greater number of regionalisms, and lexical conservatism as a tendency to preserve older administrative terms. (For an exhaustive sociolinguistic description of these Hungarian-language varieties, see Fenyvesi 2005).

The Treaty of Trianon did not only divide the Hungarian-language community with political borders, creating linguistic minority groups in their own historic homeland but also led to divergent development in the use of the Hungarian language in Hungary and in the states neighbouring Hungary (Huber 2020: 17). The various *state varieties* of Hungarian formed different *standard varieties* and varying norms of speech of these minority language communities due to the close links to each respective majority culture, institutional and legal system. Therefore, Hungarian can be regarded as a *pluricentric language* (Lanstyák 1996, Csernicskó and Kontra 2018: 104). But this pluricentricity characteristic of minority varieties of Hungarian is "lacking the appropriate status" (Muhr 2012: 33) since the codification of lexical items used in regional standard function is not fully recognized due to the dominant prescriptive and exonormative model of codification in Hungary (Huber 2020: 23–25).

In November 2001, at the initiative of the Hungarian Academy of Sciences, linguistic research centres were established in Hungarian-speaking areas

Humanities and Social Sciences, Károli Gáspár University of the Reformed Church in Hungary (Theoretical and Experimental Linguistics Research Group [TELIR], reg. no. 20736B800).

2 Data from the census of 2021 are still unavailable for this region. For this reason, we use data from the previous censuses.

neighbouring Hungary in Romania, Slovakia, Ukraine, and Serbia. In the same year, linguistic individual study bases came into being in Austria, Slovenia, and Croatia.[3] The tasks of these linguistic research centres include the sociolinguistic study of Hungarian spoken in the regions neighbouring Hungary, its varieties, lexicological and terminological issues, analysis and exploration of language problems in the context of language planning and language management.

Based on this network organization, these linguistic offices established the Budapest-based Termini Association in 2013, creating a legal framework for the already existing one-decade-long cooperation (http://termini.nytud.hu).

From the very beginning, one of the main research topics of these research centres was the collection and study of the special vocabulary of the Hungarian minority varieties. As a result, the Termini Online Hungarian Dictionary and Database (TOHDD) was established in 2007, and it has been continuously developing ever since.

The lexicographic database supplies lexical data concerning contact phenomena (widespread direct and indirect loanwords) and independent lexical creations present in the minority varieties of Hungarian. In addition to these items, the dictionary also contains common regional lexical elements characteristic of these regions and vernacular words as well. Since lexicalized forms are not limited to certain geographical and social dialects or registers only, the dictionary can be considered a general dictionary of today's Hungarian language. By documenting older linguistic forms used today in minority situations, the online dictionary reflects linguistic tradition, and by presenting neologisms, it registers current linguistic change and renewal.

The dictionary also functions as a digital database. The lexical data are placed in a relational database so that virtually any statistical analysis can be performed on them. Various glossaries and statistical statements ensure rich processing opportunities. In addition, the large number of example sentences forms a structured living language corpus that offers many additional research opportunities for digital word processing (Benő et al. 2020: 156–158).

The concept of the dictionary is based on treating the Hungarian language as unified, disregarding political borders (defragmenting the language, as it were). The aim of the linguistic defragmenting or de-bordering programme is to collect

3 Actually, these language centres started operating in 2002: the Attila Szabó T. Language Institute in Cluj/Kolozsvár, Romania; the Gramma Language Office in Dunajská Streda/Dunaszerdahely, Slovakia; the Antal Hodinka Institute in Beregovo/Beregszász, Ukraine; and the Hungarian Scientific Research Society in Subotica/Szabadka, Serbia, whose role was later taken over by the Verbi Language Research Workshop. The other neighbouring countries (Slovenia, Austria, Croatia) also had at least one locally-based researcher closely collaborating with them. This linguists' network was later expanded to include Imre Samu Language Institute, which studies the language use of Hungarian communities in Slovenia and Austria, and, most recently, the Glotta Language Institute in Croatia.

and enter elements from the entire Hungarian language area into new dictionaries (Péntek 2009: 104–105), since until the political changes of 1989 the codification of standard Hungarian did not take into account the regional standard varieties of Hungarian spoken outside Hungary by more than 2.5 million Hungarian native speakers.

In this respect, lexicographic antecedents of TOHDD are those Hungarian dictionaries that display lexicographic data from the varieties of the Hungarian language spoken in regions neighbouring Hungary. The second edition of *Magyar értelmező kéziszótár* [The Concise Explanatory Dictionary of Hungarian] (Pusztay 2003) indicates a commitment in this direction since the lexical data of the dictionary had been expanded to include representative Hungarian words and meanings used in the Hungarian spoken in Romania, Slovakia, and Ukraine.[4] Since 2003, other Hungarian dictionaries have been expanding their lexicographical data taking into account the lexicon of Hungarian speakers living in minority situations in the countries neighbouring Hungary. Such a lexicographical work is the Hungarian orthographic dictionary (Laczkó and Mártonffy 2004), which displays place- and institution names used in the minority varieties of Hungarian. The dictionary of recent lexical borrowings [*Idegen szavak szótára*] (Tolcsvai 2008) also makes use of the loans collected by linguists of the Termini Hungarian Linguistic Network in the regions outside Hungary. Thanks to further collaborations, new entries, meanings, and synonyms used in bilingual contexts were proposed for *Értelmező szótár+* [Hungarian explanatory and etymological dictionary] (Előry 2013).[5]

The lexical database has *practical benefits*. It makes it possible for Hungarian speakers to learn the words and phrases of the regional varieties used in different countries neighbouring Hungary.

The words included in the dictionary are *legitimized* in a sense, but "legitimizing" should not be confused with their *codification* (with their incorporation into the standard variety). Legitimatization here means the acknowledgment of the existence of these lexical items, but it does not mean that all such words would be used in high-level, formal-style writing.

In addition to lexicological and semantic research, the lexicographical database can also be used in *morphological*, *syntactic*, *pragmatic*, and *translation* studies since the lexical items appear in context.

The *theoretical significance* of the research programme cannot be neglected either. It is important to examine how a language adapts to the specific needs of

4 Approximately 250 Hungarian words and meanings were included from Slovakia, 91 from Romania, and 38 from Ukraine (Kiss 2004).

5 According to M. Pintér (2017: 166), these dictionaries contain entries from TOHDD in the following distribution: Hungarian explanatory and etymological dictionary: 55 entries, 115 meanings, 850 synonyms; the concise explanatory dictionary of Hungarian: 273 entries; Hungarian orthographic dictionary: 739 words; the dictionary of recent borrowings: 1,501 headwords.

speakers living in eight different countries, what types of contact phenomena are introduced in each community, what it is that can enter easily into the receiving language system, and what enters only with certain modifications (Lanstyák et al. 2010: 56–58).

Technology in the editorial process

Digital lexicography is considered to be part of digital humanities: lexicographic work is nowadays driven by computer-assisted technology supporting the so-called computer-assisted lexicography (see Wooldridge 2004: 69). In the case of TOHDD, digital humanities are used to build the elements of microstructure or build connections between entries or producing glossaries from the datasets. With the help of database structure, human lexicon is described as a structured text enhanced with several functions. For this reason, semi-structured mark-up languages (such as XML) and databases are used to store the envisaged elements. As described in Dziemianko (2012: 319), these characteristics of the digitized texts were outlined at the beginning of the era of electronic dictionaries (see De Schryver 2003).

TOHDD is technically driven by the phpMyAdmin database management software using an SQL-database, which enables queries of selected data or searches in entries, providing user dictionaries based on selected datasets. This solution has proved to be useful, as structured data and multiuser data access are used. The database is connected by the PHP programming language to a web GUI and handles the user rights of the clients (registered and unregistered users), enables simple browsing, creates lists, statistics, and edits the entries. A user interface built for editors is straightforward to use, helping editors with dropdown lists and simple technical issues. The webpage is in Hungarian, no localizations have been done, as the dictionary covers only Hungarian lexical data (except for etymological information).

Because the database is distributed, editors can edit entries and add descriptive suggestions seen only by other editors (the database also enables sending emails to other editors – with the embedded HTML version of the chosen entry). This feature is useful, as the dictionary is edited from eight different countries. Modifications appear in the dictionary immediately after editing, which distinguishes digital lexicography from the "classic" one. The technical background used in TOHDD is nowadays widely used in terminography and lexicography, but at the time of its beginnings (in the early 2010s) it was still unique in Hungary.

The database and the online interface are not only useful for quick editing and effective cooperation between editors, but they have also proved to be good for providing statistics and data mining – this is a feature that is not available in paper dictionaries. As TOHDD is used online, multimedia data, such as images

and audio files, can be linked to its entries. The database setup allows providers to gain information on lexical data, trying to collect the most exhaustive information on lexical and conceptual content.

The dictionary can be used by two types of user profiles: registered and non-registered users. Non-registered users have fewer options for queries than registered users (editors and linguists). Non-registered users can only use the *Search* function (*Keresés*), allowing them to set the search in three ways. First, users can set restrictions to the position of character strings (characters typed into the edit box): they can be searched in the headword list, meaning field, or example sentences. Second, strings can be searched as a beginning of a word, as a set of characters in the middle of a word, or as a set of characters at the end of a word. Third, users can define the number of example sentences to be shown to the user: zero, all, or only two sentences. In contrast, registered users can view a more detailed setting of the search tool, making the dictionary much more interactive, but this setting does not contain all fields represented in the database either, only those which contain important linguistic data. Linguistic information to be used in lexicology, lexicography, sociolinguistics, or data mining will only be visible after the query of represented data (e.g. the log data remain covered).

Technology has changed the way a dictionary is compiled: the process used for compiling a dictionary is conducted in a distributional database where all modifications can be viewed immediately in the dictionary, and modifications leave footprints in the statistics and the log files (entries can have their own "history"; their versions can be viewed in their "life period"). Editors can be informed about the changes through the database, and changes can be done anytime, as there is no paper format consuming extra costs for printing.

Interactive dictionary

As Lew states, the efficiency of e-dictionaries can be measured only according to the expectations towards them, and paper dictionaries are as useful as the digital ones: the differences between them are mostly based on the conditions of the needs of a particular user in a particular situation (Lew 2012: 344, 360). However, access to data is quicker and more effective (in comparison to the paper dictionaries), and additional materials linked to entries (such as hyperlinks and multimedia material) are also enabled.

TOHDD is not simply a dictionary but rather a database of language use, varieties, morphological data, and lexicon – it can be used for data mining, and its content can be visualized according to several criteria. TOHDD makes it possible to create glossaries and lists, and the classical dictionary view can also be generated.

The user interface determines the queries in various ways. Users can gain data (make their own dictionaries) and information with the help of three views. Its speed depends on the local computer, browser, and Internet connection, i.e. on circumstances independent of the dictionary and its system. Views (query interfaces and possible formats of the dictionary entries assigned to users) are connected to user rights and registration. Registered users can exploit the database more deeply and view more options in queries. The interface built for non-registered users offers only the basic options, enough for browsing data and reading the lexicographic information as described above (*Keresés*/Search function).

The enhanced search tool *Szójegyzékek* (Glossaries) is built for registered users (especially with linguists in mind) and serves as a tool for data mining or helps dictionary editors in the editorial process. Queries facilitated through dropdown lists can be done in any part of an entry. *Glossaries* can provide four types of lists, focusing on specific data within the entries. Selected entries can be organized into glossaries according to the time period of the creation/modification (*Időszak*/Time period), entries listed connected with any editor (*Kutatási listák*/Editorial lists), parts of the microstructure rendered as lists for editorial work (*Fejlesztési listák*/Development lists), or glossaries to view the work of any editor (*Dokumentációs listák*/Documentation lists). The structure of entries or lists presented in the Glossaries section can be set according to data in microstructure or metadata in entries.

As has been mentioned above, editors can add comments to any entry (and send them to any other editor for further correction). Comments can be viewed only by the editors (as they are written for work purposes – this feature is hidden from the public, only editors can use it). Unfinished entries can be set as hidden entries viewed only by editors – in the *Glossaries* function, editors can generate a list showing all hidden entries and work on them.

Lists and glossaries are generated for the lexicographic enhancement of the entries, and they also provide information on varieties of Hungarian spoken outside Hungary. As the database and its front end (user interface) are flexible, any glossaries and lists used for quantitative research can be changed or new options for generating lists/glossaries can be added according to the needs of the editors.

With the help of lists and views, users and editors of the dictionary can select the information they want to see, while all the other, distracting information is excluded. Data in the entries are rendered with the help of colours, font types, and tool tips, helping editors and readers access the most accurate linguistic information.

TOHDD is a unique lexicographic tool in Hungarian lexicography. Its uniqueness concerns the coverage of varieties, its detailed microstructure and editorial process allowing simultaneous compilation from eight countries. The database-based dictionary is especially built for online use, although its entries

were published in several monolingual Hungarian dictionaries published by major publishers. TOHDD is the only dictionary focusing on the varieties of Hungarian spoken outside Hungary.

Multimedia elements

Today, the growing popularity of digital dictionaries is undeniable. Their biggest advantage is that access to information is easy and fast. Furthermore, information can be accessed from anywhere at any time. Digital storage also enables a wide range of multimedia tools to be used, so dictionary users can not only read the "traditional" dictionary entries found in paper dictionaries, but visual and acoustic elements are also incorporated. The use of the image as an element supporting the description of meaning is a traditional feature in the history of dictionaries (cf. picture dictionaries). If we look back at the history of Hungarian dictionaries, we find Comenius's pictorial "dictionary" as early as 1653, which is a great forerunner of the use of visual possibilities (P. Márkus 2019). Another good example is the first edition of The Concise Explanatory Dictionary of Hungarian (*Magyar értelmező kéziszótár*, Juhász et al., 1972), which explained certain headwords not only through definitions but also by means of pictures. Pictorial illustrations in dictionaries are incorporated in order to assist the user with the process of decoding meaning. They were considered typical in encyclopaedias and, therefore, encyclopaedic by nature since illustrations are concerned with the world, not with linguistic signs (Rey-Debove 1971). However, modern dictionaries include not only linguistic but encyclopaedic information as well. The latter type of information can be provided in written and visual form.

In TOHDD, pictures which can be inserted separately for each meaning in the dictionary entry are mostly used to represent nouns and concrete objects, especially for culture-specific ones. Captions and the sources of the illustrations can also be added. The captions do not necessarily repeat the lemma but are useful for descriptive and explanatory purposes. This can be seen in the entry for *blokk* 'block of flats', 'apartment block', where under the equivalent 'block of flats' the caption reads "Typical socialist-era block of flats in Szered".

Following Hupka (1989), who classifies visual elements into several types, TOHDD uses three main types. The first type includes images which show just one specimen of a kind (*single illustration*). In these cases, it is recommended to choose the most typical elements. The illustration in the entry of *baklazsán* 'aubergine', 'eggplant' is a good example, where the shape, form, and colour of the vegetable are sufficient for interpretation. The dictionary also contains brand names, which can also be included in this group. Images of brand-related products can be used to present these products in a clear and concise way (e.g.

ticsinki 'salty sticks', where the packaging and appearance of the product speak for themselves).

ticsinki (fn)
(Gaszt) Fv *(ált) (közh) (biz)* ropi ♠ Fv *én viszek magammal mexicornt, ticsinkit,* sajtostekercset (ez olyan anyukám féle finomság). szerintem valamit kifelejtettem :-). (http://zene.sk/forum) ♠ Fv *Természetesen azok, akik nem szeretik, ha labdával futkároznak a férfiak, akik egy pillanatra sem ültek le férjük, vőlegényük, udvarlójuk, szeretőjük mellé, hogy „ticsinkit"* rágcsálva drukkoljanak a képernyő előtt ennek vagy annak a csapatnak. (http://mek.oszk.hu; Bodnár Gyula: Szappanopera a medencében)



In the next type, the object is presented in its normal surroundings (*structural illustration*). This type can provide the reader with a lot of extra information (e.g. *dodávka* 'van', where the vehicle can be seen in front of a warehouse entrance, in the driveway, illustrating its commercial vehicle character).

dodávka (fn) ~´k, ~´t, ~´ja
(Gépk) Fv *(ált) (közh) (biz)* furgon ♥▪ Fv *Dodávkával hozták az árut.* (ht-kut) ♥▫ Fv *Dodávkával hozatta haza a szőnyeget, a személyautójába nem fért bele.* (f.n.)



The third type of images shows the object in operation (*functional illustration*), since the meaning is best illustrated with the demonstration of the function (e.g. *bublifuk* 'soap bubble blower', where a child is blowing soap bubbles).

Because of the typological characteristics of the TOHDD, it does not use other types of images, such as nomenclatural illustrations, which are particularly common in learner's dictionaries, as they support vocabulary development. There is also no need for the type illustrating actions and processes of various kinds (*sequential illustration*). Sequential illustrations are used to graphically

explain different types of processes such as the stages of insect development in encyclopaedias.

In addition to the visual content, the dictionary also includes audio material, as some of the example sentences in the entries illustrate spoken language usage. The audio recordings are selected from the editors' collections. They can be read as written examples and listened to as audio files in the entries. The dictionary contains relatively little audio material since it is a new area of development and is still in an experimental phase.

Editorial and structural details

Entry structure

The microstructure of the dictionary comprises three parts, which are typographically distinct from each other in the dictionary: the head, the body, and the foot. There are two types of elements in the head: elements that are included in all the dictionary entries (obligatory elements: headword, part of speech specification, and inflected forms) and elements that are only included in certain types of entries (optional elements: spelling variants or pronunciation).

The dictionary indicates the pronunciation or spelling variant(s) of the headwords that have been borrowed in their unchanged form during language contact. Spelling variants, listed in round brackets after the headword, are pronounced as the corresponding lemma, differing only in their written forms. The dictionary only indicates the pronunciation of the headword if it differs from the written form or if it does not follow the pronunciation rules of the Hungarian language. Pronunciation is presented after each headword in square brackets. A pronunciation variant is provided for headwords if there is a sound absent from the Hungarian sound system or the spelling of the sound is different from that used in Hungarian (e.g. the specific speech sounds of Serbian and Ukrainian or the diacritical marks of Slavic languages). Because pronunciation is less regulated than spelling, there may be more than one spoken form of a headword, and these are indicated in the entry of the standardized headword.

The versatility of the microstructure is demonstrated by the richness of the information in the main body of the entry, which contains two separate sets of information: information needed to understand the entry word and example sentences to illustrate its use and meaning. The interpretation of the lemma is aided by the elaboration of the meanings and senses as well as the 46 status labels (subject, regional, temporal, and stylistic labels). Status labels are used to specify the scope of the use of a headword or a meaning.

TOHDD being a monolingual, explanatory dictionary with etymological references, it has two main types of definitions: synonym definitions and lexicographic definitions (cf. Svensén 2009). The dictionary's example sentences are also a means of presenting meanings as accurately as possible since meaning can be clarified through example sentences. Furthermore, the stylistic and linguistic value of the headword can be demonstrated through the different types of example sentences, which are increasingly moving away from literary texts towards real-life examples. Earlier Hungarian dictionaries abound with examples from literature, while today's dictionaries try to draw not only from literary texts but also from real-life language use. Corpora can be a great source in this process because corpus-based dictionaries are no longer an unattainable dream but have become one of the hallmarks of modernity; as TOHDD is a dictionary of language use, utilized primarily for sociolinguistic (and lexicological) research, and the primary classification of example sentences is based on their authenticity. The editors aim to use authentic – spoken or written – example sentences and avoid the use of lexicographer's made-up examples. In this respect, example sentences in the dictionary are of three types: (1) authentic written language example sentences, (2) authentic spoken language example sentences, and (3) non-authentic example sentences (whose sources cannot be verified). These categories are further subdivided into subcategories, specifying the extralinguistic circumstances in which the example sentence was created. Based on the above, the types of example sentences in the dictionary present the following typology (see April 2022; statistics in brackets): authentic written language (52,654); authentic spoken language (2,703); undocumented and thus not authentic (204); meta-linguistic, authentic written, lay speaker (3,150); meta-linguistic, authentic written, linguist speaker (588); meta-linguistic, authentic spoken, lay speaker (691); meta-linguistic, authentic spoken, linguist speaker (91); constructed by a lay speaker (18,220); constructed by a linguist (7,253); translated text (115); other (447).

An authentic example sentence is one that is not taken from an elicited speech event, i.e. it is a statement that has been published in writing (either in an online forum or in a literary work) or is a documented oral statement made by a speaker. Another interesting feature of the labelling of example sentences is that the editors indicate not only the regional affiliation of the word but also the regional affiliation of the speaker and, in the case of authentic example sentences, the source.

The references in the foot of the entry include three types of information (each in a separate line): variants in the same region, analogous words in other regions, and other words of the dictionary from the same word family, regardless of their occurrence.

The dictionary is primarily intended for sociolinguists, but researchers from other fields of linguistics, such as contact linguistics, descriptive linguistics, as well as translation theory and practice, also use it. In addition, it is also important for a

variety of practical uses: the editors of the dictionary are particularly interested in enabling lay language users to learn specific words and meanings of other varieties of Hungarian (Lanstyák et al. 2010: 44). The lexicological (temporal, regional, and stylistic) and lexicographic description of the vocabulary (e.g. the structure of the entry) provides "layers of use" for the dictionary in several ways: as an explanatory dictionary of the Hungarian language, a general dictionary, a dictionary of foreign words, as well as a source of language statistics and data mining.

The wordlist of the dictionary aims to reflect primarily synchronic language use – this is one of the reasons why it also includes neologisms. In addition to the present-day Hungarian vocabulary, there are also some typical words from the interwar period and from the socialist era. Headwords or meanings marked *Hu*, i.e. elements used as archaic in Hungary are used in various varieties in bordering regions – explaining the archaic nature of marginal dialects. Although reflecting a synchronic situation, the obsolete, old-fashioned or somewhat old-fashioned *Hu* headwords in the dictionary are now commonly used words or meanings in some regions. As the wordlist of the dictionary is not in any way linked to a particular language variety, its general dictionary character is also obvious. The criterion headwords need to meet to be included in the dictionary is that the word, word combination, or meaning has to have widespread use (it should not be limited to a small geographical area or to any stylistic variety or register). As the entry words and meanings included in the dictionary typically originate in a state language but have become an integral part of the Hungarian language usage of the given region, the dictionary can also be seen as a kind of dictionary of recent loans (called foreign words, *idegen szavak*, in the Hungarian linguistic tradition). Among the headwords, there are some that are also used in Hungary but with a meaning that is specific to the given minority language variety and different from the Hungarian one. The vitality of TOHDD lies precisely in the fact that it functions as a database: users access the headword or meaning they are looking for through queries. The database feature of the dictionary has several advantages (which is perhaps why the editors do not plan to publish the dictionary as a book), providing easier editing and simpler, more customized presentation. This can guarantee the continued success of the dictionary.

Labelling system

The labelling system used in the dictionary is based on the linguistic labels of *T Magyar értelmező kéziszótár* [The Concise Explanatory Dictionary of Hungarian] – we tried not to deviate too much from them, as most Hungarian dictionary users are familiar with the labels of this dictionary and encounter them in other dictionaries, so users can interpret them most easily. The stylistic value of

each vocabulary item or meaning was defined along various sociolinguistically definable categories – these are the dimensions. The system currently consists of five main dimensions. In defining the dimensions, we have taken as our starting point the linguistic reality, but we have also been influenced by M. A. C. Halliday's classification (Halliday et al. 1964). The following dimensions are used:

1. The labels of the dialectal dimension refer to the stylistic value of lexemes as a result of their use being specific to certain geographical regions and social groups. Regional dialects are referred to as "regional". Examples of labels referring to social dialects are "educated" and "uneducated" – referring to the level of education of the language user – and "child language" and "youth language", which refer to the age group.

2. The register dimension is largely made up of the stylistic values under which headwords associated with certain specialized activities can be classified. This category includes, for example, "technical", "literary", and "slang". The fact that a word or a meaning is not linked to any specialized register but can occur in discourse belonging to any register is referred to by the qualification "commonly used".

3. The stylistic variation dimension consists of registers defined by the formality of the speech situation; the stylistic value of individual headwords is determined by whether they are generally used in formal or informal speech situations. Examples include the style labels "formal" and "informal". Words that are not associated with a style but can be used in any speech situation are labelled "neutral".

4. In the dimension of temporality, there are labels such as "old-fashioned", "obsolete", and "historical". Since most of the headwords and meanings do not refer to temporality, for the sake of simplicity we do not refer to these with any special label.

5. The labelling system also includes, of course, "real" stylistic classifications, which express the speaker's emotions and positive or negative attitudes towards the reality indicated by the headwords. This dimension includes labels such as "euphemistic", "humorous", "ironic", "taboo", "pejorative", and "rude".

6. Finally, in addition to the dimensions mentioned above, in some specific cases, the frequency of words is also indicated by the label "rare", and the regional affiliation of the headwords is also marked.

Conclusions

TOHDD is a unique lexicographical project work in Hungarian lexicography. It is unique in terms of the varieties covered, unique in the elaboration of its microstructure, and unique in the process of its production since it is a database dictionary edited by linguists, done as an electronic dictionary of special varieties

of Hungarian. This dictionary is the only Hungarian dictionary to provide lexical elements and meanings of the Hungarian language used beyond the borders of Hungary, employing a highly sophisticated lexicographic and lexicological apparatus and information and communications technology.

Using the online interface dictionary, users can create their own dictionaries on the fly and compile glossaries and statistics by using its sophisticated query system searching the database in the background. The material of the dictionary is used and exploited not only by "lay" language users but also by linguists and researchers of Hungarian varieties, providing the basis for numerous quantitative and qualitative analyses. Although not intended for linear reading and not published as a book, it is useful for those who wish to learn more about the meanings and usage rules of specific vocabulary of the state varieties of the Hungarian language in the Carpathian Basin. The dictionary aims to provide a detailed picture of the lexical and conceptual content behind headwords.

Work on the dictionary is supported by the Hungarian Academy of Sciences. The annual projects focus primarily on the development of micro- and macrostructure, the correction and improvement of the dictionary entries, and the implementation of documentation activities to support the development work.

The TOHDD can be used in almost all areas of linguistic research into present-day Hungarian, as it is a systematic and easily searchable database of the living language since its data come from actual language use and communication.

References

A. Dictionaries

Előry, Vilma, ed. 2013. *Értelmező szótár*+ [Hungarian Explanatory and Etymological Dictionary]. Budapest: Tinta Könyvkiadó.

Juhász, József–István Szőke–Gábor O. Nagy–Miklós Kovalovszky, eds. 1973. *Magyar értelmező kéziszótár* [The Concise Explanatory Dictionary of Hungarian]. Budapest: Akadémiai Kiadó.

Laczkó, Krisztina–Attila Mártonfi, eds. 2004. *Helyesírás* [Orthography]. Budapest: Osiris Könyvkiadó.

O. Nagy, Gábor–József Juhász–István Szőke–Miklós Kovalovszky. 1972. *Magyar értelmező kéziszótár* [A Concise Explanatory Dictionary of Hungarian]. Budapest: Akadémiai Kiadó.

Pusztai, Ferenc, ed. 2003. *Magyar értelmező kéziszótár* [A Concise Explanatory Dictionary of Hungarian]. Budapest: Akadémiai Kiadó.

Tolcsvai Nagy, Gábor, ed. 2008. *Idegen szavak szótára* [Dictionary of Foreign Words]. Budapest: Osiris Könyvkiadó.

B. Other literature

Benő, Attila. 2020. Terminology and language planning for Hungarian spoken in Romania. In Ildikó Vančo–Rudolf Muhr–István Kozmács–Máté Huber (eds), *Hungarian as a Pluricentric Language in Language and Literature*, 113–126. Berlin–Vienna: Peter Lang.

Benő, Attila–Tihamér Juhász–István Lanstyák. 2020. A Termini „határtalan" szótára [The "Borderless" Dictionary of the Termini Research Network]. *Magyar Tudomány* 2: 153–163.

Csernicskó, István–Miklós Kontra. 2018. Határtalanítás a magyar nyelvészetben [De-bordering in Hungarian Linguistics]. *Fórum Társadalomtudományi Szemle* 1: 91–104.

De Schryver, Gilles-Maurice. 2003. Lexicographer's dreams in the electronic-dictionary age. *International Journal of Lexicography* 16(2): 143–199.

Dziemianko, Anna. 2012. On the use(fulness) of paper and electronic dictionaries. In Sylviane Granger–Magali Paquot (eds), *Electronic Lexicography*, 319–342. Oxford: Oxford University Press.

Fenyvesi, Anna, ed. 2005. *The Hungarian Language outside Hungary*. Amsterdam: John Benjamins.

Halliday, Michael Alexander Kirkwood–Angus McIntosh–Peter Strevens. 1964. *The Linguistic Sciences and Language Teaching*. London: Longman.

Horony, Ákos. 2016. Az anyanyelvhasználat jogi keretei Szlovákiában [The legal framework of using Hungarian as a mother tongue in Slovakia]. *Acta Humana* 4(3): 45–55.

Huber, Máté. 2020. Hungarian as a pluricentric language. In Ildikó Vančo–Rudolf Muhr–István Kozmács–Máté Huber (eds), *Hungarian as a Pluricentric Language in Language and Literature*, 19–32. Berlin–Vienna: Peter Lang.

Hupka, Werner. 1989. *Wort und Bild: Die Illustrationen in Wörterbüchern und Enzyklopädien. Lexicographica, Series Maior 22*. Tübingen: Max Niemeyer.

Kapitány, Balázs. 2015. Ethnic Hungarian in neighbouring countries. In Monostori, Judit–Péter Őri–Zsolt Spéder (eds), *Demographic Portrait of Hungary*, 225–239. Budapest: Hungarian Demography Research Institute.

Kiss, Jenő. 2004. Egy régi-új nyelvi sikerkiadvány: a Magyar értelmező kéziszótár [An old linguistic publication renewed: The concise explanatory dictionary of Hungarian]. *Magyar Tudomány* 5: 670–673.

Lanstyák, István. 1996. Gondolatok a nyelvek többközpontúságáról (különös tekintettel a magyar nyelv Kárpát-medencei sorsára) [Thoughts on the pluricentricity of languages, with special attention to the fate of the Hungarian language in the Carpathian Basin]. *Új Forrás* 6: 25–38.

Lanstyák, István–Attila Benő–Tihamér Juhász. 2010. A Termini magyar–magyar szótár és adatbázis [The Termini Hungarian–Hungarian dictionary and database]. *Regio* 21(3): 37–58.

Lew, Robert. 2012. How can we make electronic dictionaries more effective? In Sylviane Granger–Magali Paquot (eds), *Electronic Lexicography*, 343–362. Oxford: Oxford University Press.

M. Pintér, Tibor. 2017. Adatbányászat a kétnyelvűség mögött [Data mining beyond bilingualism]. In Attila Benő–Erika Gúti–Dezső Juhász–Szilvia Szoták–Erika Terbe–András Trócsányi (eds), *Tudományköziség és magyarságtudomány a nyelvi dimenziók tükrében. A VIII. Nemzetközi Hungarológiai Kongresszus (Pécs, 2016) három szimpóziumának előadásai* [Linguistic Dimensions of Interdisciplinarity and Hungarian Studies: Papers of the Three Symposiums of the 8th International Conference of Hungarian Studies, Pécs, 2016], 158–167. Budapest: Termini Egyesület.

P. Márkus, Katalin. 2019. Szótárak: Kezdetektől napjainkig [Dictionaries: From the beginnings to the present]. In Péter Furkó–Éva Szathmári (eds), *Népszerű tudomány, tudománynépszerűsítés: A Károli Gáspár Református Egyetem 2019-es évkönyve* [Popular Science and Popularizing Science: The 2019 Yearbook of the Károli Gáspár University of the Reformed Church], 146–155. Budapest: Károli Gáspár Református Egyetem.

Péntek, János. 2009. Termini: The network of Hungarian linguistic research centres in the Carpathian Basin. *Minorities Research* 11: 97–123.

Péntek, János–Attila Benő. 2020. *A magyar nyelv Romániában (Erdélyben)* [The Hungarian Language in Romania (Transylvania)]. Cluj-Napoca–Budapest: Erdélyi Múzeum-Egyesület–Gondolat Könyvkiadó.

Rey-Debove, Josette. 1971. *Etude linguistique et semiotique des dictionnaires frangais contemporains. Approaches to Semiotics* 13. De Gruyter Mouton. https://doi.org/10.1515/9783111323459.

Svensén, Bo. 2009. *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.

Tolcsvai Nagy, Gábor. 2021. The indigenous status of the Hungarian language community in the Carpathian Basin: A historical and contemporary interpretation. *Foreign Policy Review* 2: 47–61.

Wooldridge, Russon. 2004. 'Lexicography'. In Susan Schreibman–Ray Siemens–John Unsworth (eds), *A Companion to Digital Humanities*, 69–78. Oxford: Wiley-Blackwell.