Evaluation System for Online Terminological Databases

DÓRA MÁRIA TAMÁS

Hungarian Office for Translation and Attestation Ltd., Termik Terminology Research Group at Károli Gáspár University of the Reformed Church of Hungary

ESZTER SERMANN

University of Szeged, Termik Terminology Research Group at Károli Gáspár University of the Reformed Church of Hungary

ABSTRACT

This paper attempts to create an extensive classification and evaluation system to offer objective viewpoints for the comprehension, description, classification, evaluation and review of central online termbases of larger organisations. Such a comprehensive classification already exists for lexicographical genres, i.e. printed and online dictionaries, but it is still missing for online termbases. The research is based on previous international and Hungarian literature, the study of several online examples, of existing partial classifications and our previous research. The experimental categorisation with a schematic version in the appendix offers a contribution to a deeper comprehension, the possibility of a thorough description and more attentive evaluation, and ultimately, a growing awareness of the use of online termbases. Providing a new insight may also inspire the development of these beneficial tools which facilitate communication in the field of specialised languages.

KEYWORDS: translation tools, terminological database, evaluation system, concept orientation, data categories.

ANOTACIJA

Straipsnyje siekiama pateikti visapusišką klasifikavimo ir vertinimo sistemą, kuri būtų objektyvus pagrindas siekiant suprasti, aprašyti, klasifikuoti, vertinti ir peržiūrėti svarbiausias didesnių organizacijų internetines terminų bazes. Tokios visapusiškos leksikografinių žanrų, t. y. spausdintų ir internetinių žodynų, klasi-

fikacijos jau esama, tačiau internetinėms terminų bazėms taikomos klasifikacijos nėra. Tyrimas grindžiamas Vengrijoje ir kitose šalyse paskelbta literatūra, kelių internetinių pavyzdžių ir dalinės klasifikacijos analize, autorių atliktais tyrimais. Bandomasis skirstymas į kategorijas taikant straipsnio priede pateikiamą schemą leidžia geriau suvokti, išsamiau aprašyti ir atidžiau įvertinti naudojimosi internetinėmis terminų bazėmis ypatumus, pagaliau – suvokti naudojimosi tokiomis bazėmis svarbą. Naujos įžvalgos gali paskatinti šias bendravimą specializuotų kalbų srityje lengvinančias priemones tobulinti.

ESMINIAI ŽODŽIAI: vertimo priemonės, terminų bazė, vertinimo sistema, orientavimasis į sąvoką, duomenų kategorijos.

INTRODUCTION

Today online termbases have become effective and almost indispensable tools for accurate professional communication and term usage. The termbase is a genre belonging to terminology as an applied linguistics discipline, and has unique features regarding its structure and content. "Terminology" is used in three senses (theory, methodology and collection of terms), its central category is the concept, and therefore it has a concept-oriented approach (Bessé et al. 1997; Fóris 2005). This can be seen in all three areas of terminology: in theoretical research, in methods used to analyse terms for conceptual comparisons, and organising data and terms in termbases (Tamás 2018).

Although termbases have some common features, they have been created for different purposes (e.g. language policy, company purposes, or as part of the work of international organisations), they contain concepts of different subject fields (e.g. legal, economic, medical etc.) and at different levels of detail, so it is easy to see that they present many differences. This is why it is not simple to create a generally accepted evaluation system that covers all aspects.

After surveying the relevant literature, this paper aims to offer a tentative system for examining and evaluating large online central databases according to relatively objective and comprehensive criteria. The evaluation criteria developed for printed and electronic dictionaries (Fóris, Rihmer 2007; Gaál 2012) formed the basis of our evaluation system. Dictionaries have a different approach as for principles of compilation and user interface, therefore new criteria or criteria filled with new content are necessary for studying and evaluating termbases.

1. LITERATURE REVIEW

In our previous studies we have created a definition of terminology databases, which proposed that a termbase is a collection of electronically stored terminology data, created from an onomasiological approach, based on mapping the conceptual system of the subject field. It contains terms and their definitions of one or several subject fields in one or more languages (Sermann, Tamás 2010: 113). As we observed, in the relevant international literature the terms *terminological databank and database* are often used as synonyms (e.g. Hartmann, James 1998), however in various studies *databank* is considered as a wider concept that includes *databases* (e.g. Cabré 1998; Sager 1990). Based on ISO's definition, it is a "database containing terminological data collection" (https://ttbs.isolutions.iso.org/obp/ui#:term:3.29). We have compared the characteristics of termbases and those of electronic dictionaries (Sermann, Tamás 2010) and have concluded that electronic dictionaries and termbases have some identical features and also some differences, the latter being summarised in Table 1.

Table 1. Table summarising the features of online electronic dictionaries and termbases
(Tamás 2014: 139; based on Sermann, Tamás 2013)

ONLINE ELECTRONIC DICTIONARY	TERMBASE
semasiological principles	onomasiological principles
central element: the word or linguistic sign	central element: the concept which is designated by the term and expressed in the definition
lists meanings	renders data to the concept (the definition)
extra information: e.g.: pronunciation (speaker icon), example sentences taken from the context dictionary, conjugation of verbs	extra information: information on the level of equivalence (full, partial or lack of equivalence), and on the equivalent provided (term, functional equivalent, term candidate, standardised, harmonised, recommended etc.)

The evaluation criteria for printed dictionaries (Fóris, Rihmer 2007; Fóris 2018) cover background information on the dictionary, information

Dóra Mária Tamás | *Evaluation System for Online Terminological Databases*
Eszter Sermann

on the content (mega-, macro-, micro- and mesostructure), content elements (lexicological basics and linguistic information), and data of dictionary usage. The megastructure of a dictionary refers to the presence or lack or quality of the front matter (foreword, contents, instruction) and the sources used; the macrostructure is the organisation of lexical entries (e.g. alphabetical or thematic); the microstructure is the inside structure of lexical entries; and the mesostructure is the system of cross-references.

The evaluation criteria for electronic dictionaries (Gaál 2012) contain elements to match the features of the reference tool, and these include the technical parameters of the dictionary, its embeddedness, the features of the search engine and the displaying of results.

Table 2. Main evaluation criteria for printed and electronic dictionaries
(Fóris, Rihmer 2007; Fóris 2018; Gaál 2012)

PRINTED DICTIONARIES (FÓRIS, RIHMER 2007; FÓRIS 2018)	ELECTRONIC DICTIONARIES (GAÁL 2012)
Background information	Background information
Structural information on the content: mega-, macro-, micro- and mesostructure	Structural information on content: mega-, macro-, micro- and mesostructure, novelties: *technical parameters, embeddedness, price, features of the search engine and results.*
Information on the content: phonology, grammar and semantics	Information on the content: phonology, grammar and semantics, novelties: *audio files*
Data on dictionary use: usability, access, innovative features and wider social value	Data on dictionary use: usability, access, innovative features and wider value – *novelties: user friendliness, updates and reliability*

Some papers (Tamás 2010, 2014; Sermann 2011; Novák 2018) described specific termbases, others created various classifications and evaluation systems (Sager 1990; Arntz et al. 2014; Tamás 2014; Schmitz, Drewer 2017), but these are not suitable for comprehensive analysis in education, the training of translators or writing reviews. In this paper we use the following classification system for databases (see Table 3):

Table 3. A tentative classification of termbases (based on Tamás 2014: 111)

TYPE OF TERMBASE	CONCEPT ORIENTED	TERM AUTONOMY	NUMBER OF DATA FIELDS
SIMPLE	necessarily applicable (or with some hybrid features)	not necessarily applicable (resemblings to a word list)	their number does not reach the minimum (e.g. only term and subject field)
TRADITIONAL	necessarily applicable	applicable	traditional number of data fields present, a definition in at least one language
COMPLEX (sometimes called knowledge base or terminology information system	necessarily applicable	applicable	traditional number of data fields present, and further pieces of knowledge (concept maps, corpora)

According to the above classification e.g. the Termin (https://jogi-ter-minologia.im.gov.hu/), the termbase of the Ministry of Justice of Hungary, belongs to the category of simple termbases. It has entries in four languages (Hungarian, English, French, German), in a simple, table format. It contains information on the domain and subdomain, with occasional explanations added recently (including: source, definition, example sentences), but does not feature any other data field, and the entry does not mark a main term (which would be important for term autonomy) only showing one version. In its present form, the UNTERM (United Nations Multilingual Terminology Database, http://untermportal.un.org), also belongs to this category. This interinstitutional database was made public in 2014, and only offers a limited amount of information on terms. EUROTERMBANK, a database created on an EU initiative to facilitate the accession of new member states from 2004, also had just a few data categories (e.g. term type, grammatical part of speech, collection, domain, editors) (https://www.eurotermbank.com/) and in its new version contains collections of different domains. Another simple database is Euskalterm, created by the Basque Centre for Terminology and Lexicography (UZEI) (http://www.euskadi.eus/euskalterm/). The Centre merged their previously published terminological dictionaries to create the database in 1987, and in 2001 declared it to be the Basque national termbase. Its structure

Dóra Mária Tamás | *Evaluation System for Online Terminological Databases*
Eszter Sermann

resembles a word list, some data sheets lack a definition, but all of them show a domain, the source of terminology data and the year of recording the data (for details see Sermann 2013). The existence of these termbases is often gap-filling, of vital importance and great value, especially for minor languages.

The traditional type of termbases include the bistro (Novák 2013; Tamás 2013), which is a terminology database initiated by Bolzano, an autonomous county, and maintained by the EURAC Research Centre (http://www.eurac.edu/en/research/autonomies/commul/Pages/default.aspx). The database contains legal and public administration terms in Italian, German and Ladin (http://bistrosearch.eurac.edu/). The entries are detailed, contain several data categories, provide a definition in at least one language, and the results of a search are shown in such a way that concepts are displayed separately. One of the oldest databases is the Swiss Termdat (www.termdat.ch), which was established in 1987 and has similar features (Novák, Tamás 2013). The termbase of the EU, IATE (https://iate.europa.eu/home) also belongs to this category, but in this database definitions can be substituted by contexts. The Serbian database promoting Serbia's accession to the EU, Evronim (http://prevodjenje.mei.gov.rs/evronim/index.php?jezik=engl) also has mostly features of a traditional termbase. The SAPTerm (www.sapterm.com) has a user interface that shows some new features too, and also belongs here (see Tamás 2015) (http://www.eurac.edu/). Cercaterm, the termbase of the Catalan centre for terminology (TERMCAT) can also be regarded as a traditional terminology database (https://www.termcat.cat/ca/cercaterm). Its main language is Catalan, and it shows equivalents in Spanish, French, English and Italian. Entries contain the definition of the concept, a term in Catalan, its grammatical category and some other language equivalents.

Complex termbases entail concept maps and corpora as well. One example is the Eohsterm knowledge base (www.eohsterm.org) (Tamás 2010), which contains a concept map and the aligned texts of the analysed laws as well, but has not been updated since its creation. WIPO Pearl (https://www.wipo.int), the termbase of the UN World Intellectual Property Organisation, is an excellent example of using a concept map. It contains data in ten languages, and terms can be searched using the linguistic search or the concept map search. In addition, in case of a not yet worked out term, one can find the results of machine translation with indication.

We have tried to analyse a wide range of termbases, altogether 17 termbases, though it could be completed by tools available in certain areas, like Russia or Asia (for instance China).

It is important to bear in mind that this evaluation system concerns the user interfaces of large central online databases, and does not entail the study of termbanks created by terminologists vs. translators, the individual file or even the server based termbases, a description of the steps to create these, the analysis thereof and the description of complex terminology management systems (Arntz et al. 2014; Bowker 2015: 307; Drewer et al. 2014; Schmitz, Drewer 2017). The above mentioned publications took the needs of everyday use as their starting point and did not generalise their evaluation criteria for more general research purposes.

2. EVALUATION SYSTEM FOR ONLINE TERMINOLOGICAL DATABASES

2.1. Background information on central termbases

The study of the general background to a database may start out from a review of the termbases of the given domain. Previous versions may reveal the historical background of the termbase, and if it has a predecessor (for example as in the case of EURODICAUTOM and its successor IATE). This can be studied from various aspects. For example, Sager (1990: 165–167) looks at technological advancement and distinguishes first generation termbases (word oriented, designed for internal use, from the 1970's) and second generation termbases (that appeared in the 1980's, when the technological development enabled concept orientation and term autonomy to take place).

Another criterion would be if the termbase has a connection with other termbases: due to institutional background, data could be transferred within the organisation, there could be sub-centres (see the inter-institutional database of the EU, the IATE; Termdat, UNTERM).

A further criterion of the study of the general background could be the classification of the database based on subfields of terminology (language policy, translation and standardisation). In some cases the database is created under governmental supervision to fulfil the purposes of language policy (such as bistro, Termdat, Termium Plus, TermCat). In other cases

the aim is the creation of an effective translation tool (e.g. IATE), or the terms created during the standardisation process are made public in the termbase (e.g. DINTerm; https://www.din.de). Often, databases are multi-purpose and they serve various functions at the same time.

When describing the immediate background to the termbase, we may scrutinise the features of the organisation that produced the database: its type, background, national or international scope, domains and available funding. One criterion could be whether the organisation that created the database is one of the typical groups that usually create termbases (Tamás 2019a: 267):

– international organisations (UN – UNTERM, UN WIPO – Wipo Pearl, EU – IATE, Eurotermbank);
– public administrative bodies (Federal Chancellery of Switzerland – Termdat, Hungarian Ministry of Justice – Termin);
– research institutes and universities (EURAC Research – bistro);
– translation agencies (the Canadian Translation Bureau – Termium Plus).

The next step could be to study the target audience of the database (native speakers, speakers having a majority or minority status, translators or experts of a specific domain, e.g. engineers, lawyers, doctors etc.).

Furthermore, based on the intended purpose of the database, termbases may be distinguished on the basis of their descriptive features of informative nature and prescriptiveness of normalising nature (Tamás 2019b: 110).

2.2. Technical parameters of a termbase

Technical parameters include those that are accessible for viewers, i.e. the user interface, and do not include those that require knowledge about the internal structure of the database or the background to its creation, unless they are published otherwise. Being familiar with the internal structure of the database, we can decide whether the termbase fulfils the hierarchical three-fold structure required by the terminological approach (entry level – language level – term level); or if it has a structure of a database network, a simple database or a relational database (Sermann, Tamás 2010). We also need the background information to determine the extent to which the full data content of the termbase is accessible, or if the data is stored in one or more data repositories, and in what way.

Large central online databases usually have their own bespoke way of displaying the content. However, there may be hybrid solutions, e.g. the

bistro termbase of South Tyrol, where the detailed information stored in the SDL Trados Multiterm software may also be accessed behind the user interface. Starting from this year, the Verwaltungsglossar (https://www.sprachressourcen.at/verwaltungsglossar/), created in cooperation with the University of Vienna and containing 700 terms can already be accessed in Excel, TBX and pdf format, in SDL Trados Multiterm file format and through the online Quick Term software, and, what is more, also in the form of an app. Other databases allow users to download some of their content (see Download IATE: https://iate.europa.eu/home). Databases are becoming more modern and technologically sophisticated, and may allow users to access the data using various devices, for instance a smart phone or tablet in the case of WIPO Pearl (https://www.wipo.int/reference/en/wipopearl/news/2019/news_0003.html).

2.3. Information on the content of the termbase

2.3.1. GENERAL FEATURES OF THE DATABASE

For the purpose of studying the general features of termbases, we have grouped them relating to the applicability of the terminological approach and the structure of databases into two: one group for the more general, main features and one for the more detailed and specific ones.

In the first group we started with analysing whether the linguistic data have been recorded in the database keeping the concept-oriented approach in mind. In some cases we might witness the semasiological approach (based on the word or linguistic sign) being mixed with the concept-oriented approach, and as a consequence term autonomy is not fully observed or lexicographical tools are added (see the word list characteristics of EUROTERMBANK or the additional list of dictionaries on the website of the SAPTerm or Termium Plus database for further purposes, but not constituting part of the termbase) (see also Sermann, Tamás 2013; Tamás 2015). The Termium Plus termbase does not offer a separate hitlist with the indication of domains as a stepping stone, but offers a data sheet containing record 1, record 2 and so on. Based on the above we can distinguish between purely onomasiological termbases and termbases that have some kinds of semasiological features as well in their make-up.

Dóra Mária Tamás | *Evaluation System for Online Terminological Databases*
Eszter Sermann

Still discussing general features, in chapter 2, Table 3 of this paper, we group termbases into three categories based on their structure and the details included in entries: simple, detailed traditional, and complex termbases that contain extra information such as a corpus or concept maps.

2.3.2. DETAILED ANALYSIS OF THE STRUCTURE OF TERMBASES

Based on the analysis of their structure, termbases can be grouped in several sub-groups. We have matched these tentatively with the dictionary categorisation found in Table 2, and established what the equivalent feature is to the megastructure, macrostructure, microstructure and mesostructure in termbases. We have decided to apply this set of criteria created for vocabularies despite the different content of termbases. As we observed, the hierarchical structure corresponding to the level of microstructure is not necessarily visible on the visualisation surface, for example the term or the definition can be set on two different levels (though it can be contained by the user's manual). On the other side, users socialised on either traditional or online dictionaries could have in the case of a transparent structure more reference points and the differences were unequivocally revealed. This could be solved by a more transparent surface for external users.

In line with this, at the level of the megastructure, we can examine whether we find any information about the use of the termbase, such as a user's manual, help page, frequently asked questions, description of the purpose of the termbase, the domain, target audience, work method, number of languages and the multilingual coherences between them, domain and sub-domains; or any other informative tool, such as a forum or chat service to help users (see SAPTerm: chat, SAP Community, SAP Help Portal). For instance, the IATE database informs us of the number of terms and entries, the number of searches done on the previous week, and modifications (new entry, deletion, editing).

Databases also inform users of the copyright aspects of the termbase. The user interface of a termbase is considered a collection, and therefore is protected by copyright, with special emphasis on its structure and organisation of data, and consequently its screen display cannot be copied or published freely.

On the level of the macrostructure, we can analyse the search options offered by the termbase: simple or complex, and within the latter, based on the term, the domain and sub-domain, fragment, or the linguistic search

may also be complemented with a search in the concept map (e.g. the WIPO Pearl) or in the corpus. The database may offer various filters, for example, in bistro we can search individual areas of legislation, languages and geographical distribution (i.e. spoken in Germany, Austria or Switzerland). The database also offers the possibility to save our filter settings; and the integrated feature to detect the language may also be useful.

As a result of the search, most databases display a hitlist, and we may study the method used to display results and the organisation of data. For instance Termium Plus does not offer a separate list with the indication of domains as a stepping stone, but a data sheet containting record 1, record 2 and so on. At this point we already have some information about the nature of the data: is it only textual or are there illustrations or multimedia as well? It might also offer to search Google images (such as WIPO Pearl). It might also enable interoperability with other data repositories (for instance the IATE and the EURLEX legal repository, the WIPO Pearl and the Patenscope database of patents).

At the level of microstructure, the number of data categories is a clear indicator of the extent of detail in the database. It can be seen whether it contains minimal information or if it is a well developed, detailed data repository; and this information is grounds for evaluation for the user.

When we examine data categories, from a structural point of view we can differentiate between mandatory and optional data fields. It is always a crucial aspect whether the definition can be substituted by the context.

The relevant literature, standards and best practices (see Drewer et. al. 2014) can help us decide which are the most important data categories as per the principles of compilation, and we can also compare existing termbases (e.g. if possible also on the basis of their hierarchical organisation – respecting the entry level, language level and term level, for details see: Drewer et. al. 2014). Arntz et al. (2009: 233) claim that a simple termbase contains the following data fields: subject field, language, ID number, other codes, the term, linguistic features, definition, context, notes, synonym, data of the editor, date, source. Arntz et al (2014: 229–238) however contains a more complex description, that already meets the requirements of modern terminology management.

Owing to the concept oriented nature of a termbase, the definition undoubtedly plays a central role beyond the term, while equivalence is not always a separate category. The other frequent data categories include:

Dóra Mária Tamás | *Evaluation System for Online Terminological Databases*
Eszter Sermann

- term status (main term or head term, synonyms in a separate entry or not) and types of relations with related terms (superordinated, subordinated or co-ordinated terms);
- reliability and validated status;
- the use of various labels: geographical distribution; document type (e.g. passport such as in Termium Plus); special terms related to the confederation or cantons can be "flagged" (in Termdat); acceptance (recommended, forbidden, standardised); whether it was confirmed by the terminology committee (in bistro); archaic (in IATE: obsolete);
- illustrations and other multimedia;
- linguistic features;
- context and collocations.

If we look at ISO standards on preparing termbases we can find recommendations specifically about the data categories in the ISO 12620:2019 standard titled *Management of terminology resources – Data category specifications*. In the version published in 1999, the following ten data categories were defined: the term, information on the term, equivalence, domain, descriptive data categories on the concept, conceptual relations, conceptual structures, notes, information on the documentation, administrative information. The key data categories are: the main term, the input date and the source (Sermann 2013). Later with the purpose of unification the *Data Category Registry* was published on www.isocat.org parallel to the standard of 2009. This was an open forum which listed a total of approximately 590 data categories, including overlaps, (for details see Schmitz 2012; Tamás 2014: 103). Currently the www.isocat.org website is replaced by http://www.datcatinfo.net, which is a *Data Category Repository (DCR)* developed according to the forthcoming ISO 12620:2019. This version itself does not contain concrete data models, only guidelines and requirements, acknowledging the differences in approach for developing language resources of different domains and needs which lead to variations in data categories. Though unification remains important to enable data exchange using the TBX file format, published on www.datcatinfo.net:
- data categories of the core module: date, note, term;
- data categories of the basic module: context, definition, external cross reference, grammatical gender, geographical usage, project subset, related concept, related term, responsibility, source, term location, term type, transaction type, xGraphic.

An initiative of unified use of data categories of the creators of large databases in the organisation called JIAMCATT (see EU, UN, IMF; France Terme, Termdat, Termium Plus) aims to establish a common platform that provides a single central access to terminology resources, tools and databases (Maslias 2014; https://ec.europa.eu/education/knowledge-centre-interpretation/events/jiamcatt-2019_en).

Concept orientation and term autonomy are not the only aspects in studying whether editing principles have been respected. Other aspects include data elementarity (i.e. if the termbase contains one data element in each data field) and data granularity (i.e. if the level of detail is balanced and the number of sub-categories is not too high) (Reinke 2012: 102; Arntz et al 2014: 239–241; ISO 16642: 2017).

In line with terminology as a discipline using different work methods and trends, the user interface of termbases also reflects various approaches. For instance, definitions may be in one or several languages. In some cases, all of the languages may use the same concept (such as in unified EU or international terms), and in this case featuring a definition in one language is sufficient. Another scenario might be when the aim is to show the findings of contrastive analysis. Different aims may also require different data categories. If the aim is to support standardisation (as for DINTerm), reference to standards is crucial, if the aim is to unify and harmonise term use or for translation oriented terminology work the following are important: indicating the status of terms, functional equivalents or term candidates, providing contexts or collocations. For instance bistro contains system-bound legal collocations linked to the main term (Chiocchetti 2019). Finally, if the aim is to serve language policy, a contrastive analysis of concepts is crucial (c.f. legal terms in bistro). However, there might be data necessary in case of some languages (e.g. articles in German) or data that depend on the domain (illustrations or multimedia might be necessary for medical or engineering domains).

As for the structure, the arrangement of data categories and the amount of data in each field may be different, or the first result may show a shorter version of the text and the user can expand the text if needed (such as in the case of bistro). This way the user can get a clearer picture of the entries.

At the level of the mesostructure, we can study the referencing system, i.e. the presence and appearance of internal links (within and across en-

tries) and external links (links to external websites). The use of internal links, i.e. links within the entry and to other entries, is part of the cross-referencing system, and this may be done in case of concepts by other forms, such as via a concept map. The display of external links may be different; the questions here are whether it can be opened in a straightforward way, if there is any description of its content (address, pathway), if it is permanent or if the use of one-off links is permitted.

2.4. Information on the usage of the termbase

Information on usage involves user friendliness of the termbase, the reliability of the data, the frequency of updates, the innovative nature of the database, its importance for the profession and value for society.

The user always prefers to open a tool that is easy to use, clear and provides the information quickly. From this aspect, the user friendliness of the database, which must be experienced across various platforms (c.f. technical parameters) is of crucial importance. If the database is meant to be understandable for the general public, it is advisable to avoid abbreviations (this was a major improvement in the current version of Termdat over its earlier version), searches should be easy and the use of visuals is also recommended. One aspect of visuals is how the reliability of data is shown, so that the user immediately obtains clear information on frequency and updates. This is not only applicable to the reliability of data, but also to the quality of the data of the termbase. When making judgements on quality, we need to rely on the current form of the termbase, and it is not always apparent whether the data are indeed up-dated, checked and validated. The up-to-dateness can be shown in various forms, for instance through the date of recording, or moreover the validated state in a explicite mode for example: in the IATE stars indicate reliability, in the bistro termbase a five-point scale shows the same, and standardised terms have the top score; the WIPO Pearl uses a four-point scale for the same purpose.

In order to increase reliability, the selection and proper documentation of sources is crucial (for example, preference can be given to documents that have a permanent, DOI identifier). It is also important whether the termbase producer is working with and internal and/or an external network of experts (SAPTerm).

Another aspect of reliability is the editing methods applied in editing data (does it offer a possibility to submit suggestions, whether internal

and/or external experts are consulted, whether the editing is crowdsourced, does it contain results from machine translation as WIPO Pearl). For instance, the SAPTerm aims to establish a network of experts in each country (Tamás 2015); or whether the editors have cooperation agreements with universities (as IATE and WIPO Pearl). Editing may be done at a central location or locally and then uploaded to a central database (IATE is edited by the translation services of the EU, the Termdat is edited both centrally and in cantons, the UNTERM contains data of various bodies of the UN). In the latter cases the filtering of duplicate entries is a major challenge (see IATE and UNTERM).

Löckinger (2019: 9) examined the research methods of document editors, and evaluated the results of online questionnaires concluding that users make their judgements on quality primarily based on the content (whether it is up-to-date and professional) and usability (user friendliness, display and categorisation of information).

The innovative nature of the database may be reflected in a new kind of display or extra information (concept map, corpus, machine-translation results, video tutorial, e-course), or in whether the data repositories are linked (such as Google images, the IATE and EURLEX database, WIPO Pearl and Patentscope).

Last but not least, the value of the termbase for the profession and its importance for society should also be examined. The questions here are whether the termbase is useful for various expert groups, if it has any relevance for language policy and terminology policy, if the creator of the termbase offers further education, e-training, workshops or opportunities for cooperation (e.g. Termdat, IATE, WIPO).

CONCLUSIONS: USABILITY OF THE EVALUATION SYSTEM

The criteria for termbases is developed on four categories:
– background information,
– technical parameters,
– information on the content, and
– information on the usage.

The above tentative evaluation system reveals that termbases built on a relatively unified terminological approach and principles still present a wide range of features, because the user requirements against various domains and aims place the emphasis on different editing principles.

In addition, these features are closely related to each other, and often hard to separate (such as editing principles, documentation, reliability and quality). In line with the above we have encountered two kinds of difficulties: on the one hand, the common features are less obvious than in the case of dictionaries, and on the other hand some features belong to more than one criterion. The application of the set of criteria for vocabularies has revealed the differences between the genres of these translation tools of lexicography and terminology and the main characteristics of termbases alike. Nonetheless, the analysis can't be considered as concluded. With a further specification of evaluation criteria, a possible development of a more evident structure on online surfaces and the necessary evolution over time of termbases, it will be possible to work out a more specific description.

Still, we believe that this evaluation system can be useful for obtaining a more thorough understanding and more conscious use of these tools in education, when writing reviews and for translators alike, as already revealed in the case of vocabularies. Applying this evaluation system step-by-step, the user might discover new functions of the tool. It also helps users form and share an objective expert opinion of termbases both orally and in the form of reviews. As we have emphasised at the beginning of the paper, this evaluation system is meant to be tentative. We hope that its comprehensive nature of our analysis might even inspire developers. The genre of termbases itself is closely linked to practical requirements, however it also enables creativity.

REFERENCES

Arntz R., Picht H., Mayer F. 2009: *Einfuhrung in die Terminolgiearbeit*, Hildesheim, Zürich, New York: Georg Olms Verlag. (6. Aufl.).
Arntz R., Picht H., Schmitz K.-D. 2014: *Einführung in die Terminologiearbeit*, Hildesheim / Zürich / New York: Georg Olms Verlag.
Bessé B. de, Nkwenti-Azeh B., Sager J. C. 1997: Glossary of Terms Used in Terminology. – *Terminology* 4, 117–156.
Bowker L. 2015: Terminology and Translation. – *Handbook of Terminology*. H. J. Kocker, F. Steurs (eds.), Amsterdam / Philadelphia: John Benjamins Publishing Company, 304–323.
Cabré M. T. 1998: *Terminology. Theory, Methods and Applications* (Terminology and Lexicography Research and Practice I), Amsterdam / Philadelphia: John Benjamins.
Chiocchetti E. 2019: *Terminology Work in South Tyrol: New Approaches, New Termbase, New Domains. Scientific, Administrative and Educational Dimensions of Terminology*. Held at: the 3rd International Conference on Terminology, Lithuania, Vilnius, 17–18 October 2019.
Drewer P., Pulitano D., Schmitz K.-D. 2014: *Terminologiearbeit. Best Practices 2.0*, Köln: Deutscher Terminologie-Tag e.V.
Fóris Á. 2005: *Hat terminológia lecke* [*Six lectures in terminology*], Pécs: Lexikográfia Kiadó.

Fóris Á. 2018: *Lexikológiai és lexikográfiai ismeretek magyar (mint idegen nyelv) tanároknak. [Lexicology and lexicography for teachers of Hungarian (as a foreign language)].* Károli Gáspár Református Egyetem, Magyarnyelv-tanári segédkönyvek, Budapest: L'Harmattan.

Fóris Á., Rihmer Z. 2007: A szótárak minősítési kritériumairól. [On the Evaluation Criteria for Dictionaries]. – *Fordítástudomány* 9(1), 109–113.

Gaál P. 2012: Szempontrendszer online szótárak minősítéséhez. [Evaluation System for Online Dictionaries]. – *Magyar Terminológia* 5(2), 225–250.

Hartmann R. K. K., James G. 1998: *Dictionary of Lexicography*, Exeter: Routledge.

Löckinger G. 2019: *Fachsprachliche Nachschlagewerke für technische Redakteurinnen und Redakteure – Ergebnisse einer emprischen Studie.* Edition. 15. Jahrgang, 1. Ausgabe. 5–9.

Magris M., Musacchio M. T., Scarpa F. 2002: *Manuale di terminologia. Aspetti teorici, metodologici e applicativi*, Milano: Editor Hoepli.

Maslias R. 2014: Combine EU Terminology with Communication and Ontology Research. – *Terminology and Knowledge Engineering, 2014*. Jun 2014, Berlin, Germany. Available at: https://hal.archives-ouvertes.fr/hal-01005876.

Novák B. 2013: A bolzanói bistro olasz–német–ladin jogi terminológiai információs rendszer bemutatása. [Bistro, the Italian-German-Ladin information system for legal terminology]. – *Magyar Terminológia* 6(1), 61–83.

Novák B., Tamás D. 2013: A svájci jogi-közigazgatási terminológiai adatbank, a TermDat. [TermDat, the Swiss Terminology Database of Law and Public Administration]. – *Magyar Terminológia* 6(2), 207–231.

Novák B. 2018: *A terminológiai munkafolyamat a minőségi jogalkotásban. Magyar–olasz összehasonlító vizsgálat az alkotmányjogi terminológia területén. [The Role of Terminology Work in Quality Law Making. Hungarian–Italian Comparative Terminology at the Field of Constitutional Law].* Doctoral Thesis University of Pécs, Faculty of Law.

Kranebitter K., Ralli N. 2019: *Piccola guida per sviluppare strumenti terminologici.* Talk held at: XXIX Convegno Ass. I. Term, Terminologie e vocabolari: lessici specialistici e tesauri, glossari e dizionari. Accademia della Crusca, Villa Medicea di Castello, Firenze, 30–31 May 2019.

Reinke U. 2012: Terminologiearbeit fur und mit Sprachtechnologie. – *Terminologieprozesse und Terminologie-werkzeuge.* F. Mayer, K.-D. Schmitz (eds.), Akten des Symposions. Heidelberg, 19–21. April 2012, Deutscher Terminologie-Tag e.V, Köln: SDK Systemdruck Koln GmbH & Co. KG, 101–118.

Sager J. C. 1990: *A Practical Course in Terminology Processing*, Amsterdam / Philadelphia: John Benjamins Publishing Company.

Schmitz K.-D., Drewer P. 2017: *Terminologiemanagement. Grundlagen, Methoden, Werkzeuge*, Berlin: Springer Vieweg Verlag.

Schmitz K.-D. 2012: Using International Standards for Terminology Exchange. – *Terminologija* 19, 33–38. Available at: http://lki.lt/wp-content/uploads/2017/06/Terminologija-19-ilovepdf-compressed.pdf [accessed 05.08.2019].

Sermann E., Tamás D. 2010: Hogyan definiálhatjuk a fordítói adatbázist? Egy olasz és egy spanyol fordítói terminológiai adatbázis vizsgálata. [How to Define the Translation Database? Study of an Italian and a Spanish Translation Oriented Termbase]. – *Nyelvek találkozása a fordításban. Doktori kutatások Klaudy Kinga tiszteletére.* K. Károly, Á. Fóris (eds.), Budapest: ELTE Eötvös Kiadó, 101–115.

Sermann E. 2011: A terminológiai szabványosítás szerepe a lokalizáció folyamatában. [The Role of Termi-nology Standardization in Localisation]. – *Fordítástudomány* 13(2), 72–87.

Sermann E. 2013: *A terminológiai szabványosítás és a terminológiai harmonizáció fordítási vonatkozásai. [The Translational Aspects of Terminology Standardisation and Terminology Harmonisation].* Doctoral dissertation (Manuscript), Budapest: ELTE.

Sermann E., Tamás D. 2013: Elektronikus szótár vagy terminológiai adatbázis? [Electronic Dictionary or Termbase?]. – *Társadalmi változások, nyelvi változások. A XXII. MANYE Kongresszus előadásai. Szeged 2012. április 12–14.* Sz. Tóth (ed.), Budapest–Szeged: MANYE–Szegedi Egyetemi Kiadó Juhász Gyula Felsőoktatási Kiadó, 450–454. Available at: http://mek.oszk.hu/11700/11730/.

Sermann E. 2019: A terminológiai munkafolyamatok és a fordítás szempontjából releváns ISO-szabványok. [ISO Standards Relevant for Terminology Work and Translation]. – *A szabványosítás fordítási és termi-nológiai vonatkozásai.* Á. Fóris, A. Bölcskei (eds.), Budapest, L'Harmattan.

Sermann E. 2019: Terminológiai adatbázisok mint a nyelvi korlátok áthidalásának eszközei. Magyar nyelven (is) elérhető néhány terminológiai adatbázis tartalmi és szerkezeti vizsgálata. [Termbases as the Means to

Surpass Language Barriers. Content and Structural Analysis of Some Termbases (also) Available in Hungarian]. – Acta Historiae Litterarum Hungaricarum. *Acta Universitatis Szegediensis* 34(3), 149–169.

Tamás D. 2010: La Knowledge Base EOHS Term, ossia un vocabolario Online Creativo. – *Corvinus nyelvi napok, "Kreativitás és nyelv". 2009. szeptember 23–24.* G. Cs. Dávid, A. Zelényi (eds.), Budapesti Corvinus Egyetem IOK Alkalmazott Nyelvészeti Kutató- és Továbbképző Központ, Budapest: Aula, 349–355.

Tamás D. 2013: Az olaszországi Bolzano autonóm megyében élő kisebbségek jogainak védelméről és a bistro terminológiai projekt létrejöttének hátteréről. [On the Protection of Rights of the Minority Communities Living in the Autonomous County of Bolzano, and the Bistro Terminology Project]. – *Magyar Terminológia* 6(1), 41–60.

Tamás D. M. 2014: *A gazdasági szakszövegek fordításának terminológiai kérdéseiről.* [*The Terminological Aspects of the Translation of Economic Texts*], Fordítástudományi értekezések I, Budapest: ELTE BTK Fordító- és Tolmácsképző Tanszék.

Tamás D. 2015: Az SAPterm terminológiai adatbázis vizsgálata. [The Study of SAPterm Termbase]. – *A fordítás titkos ösvényein. Doktori kutatások Klaudy Kinga tiszteletére II.* K. Károly, Á. Fóris (eds.), Budapest: ELTE Eötvös Kiadó, 211–230. Available at: http://www.eltereader.hu/media/2018/05/A_forditas_titkos_osvenyein_READER.pdf.

Tamás D. M. 2018: Miért rögzítsünk jogi szakszavakat terminológiai adatbázisokban? [Why Should We Record Legal Terms in Termbases?]. – *Magyar Jogi Nyelv* 2(2), 30–32. Available at: https://joginyelv.hu/beszamolo-a-jog-es-terminologia-cimu-konferenciarol/.

Tamás D. M. 2019a: A IUSTerm jogi és közigazgatási terminológiai adatbázis háttérmunkálatainak tapasztalatairól. [IUSTerm Terminological Database of Law and Public Administration: A Glimpse Behind the Scenes]. – *Diszciplínák találkozása – nyelvi közvetítés a XXI. században.* Sz. Szoták (ed.), Budapest: OFFI Zrt, 260–281. Available at: https://150.offi.hu/kiadvanyok/diszciplinak-talalkozasa-nyelvi-kozvetites-21-szazadban.

Tamás D. M. 2019b: *A terminológiai adatbázisok mint a terminológia-stratégia eszközei. [Terminology Databases as Means of Terminology Strategy].* Talk held at: Terminológia-stratégiai kihívások a magyar nyelv vonatkozásában. [Challenges of Terminology Strategy in the Hungarian Language Area], Budapest: OFFI–L'Harmattan, 105–119 (In press).

SOURCES

12620: 2019 Management of Terminology Resources – Data Category Specifications. 3th edition [2019.05].

12620: 2009 Terminology and Other Language and Content Resources – Specification of Data Categories and Management of a Data Category Registry for Language Resources.

12620: 1999 Computer Applications in Terminology – Data Categories.

ISO 16642: 2017 Computer Applications in Terminology – Terminological Markup Framework. 2nd edition [2017.11].

ONLINE SOURCES:

http://www.datcatinfo.net [accessed 05.08.2019]

https://ec.europa.eu/education/knowledge-centre-interpretation/events/jiamcatt-2019_en [accessed 05.08.2019]

http://www.eurac.edu/en/research/autonomies/commul/Pages/default.aspx

www.isocat.org [accessed 10.03.2013]

https://www.termcat.cat/en/recursos/productes-multimedia/termtest-dictionary-self-assessment-test [accessed 05.08.2019]

https://www.wipo.int/reference/en/wipopearl/news/2019/news_0003.html [accessed 14.03.2019]

TERMBASES:

bistro http://bistrosearch.eurac.edu/ [accessed 14.03.2019]

DINTerm https://www.din.de [accessed 14.03.2019]

Cercaterm https://www.termcat.cat/ca/cercaterm [accessed 06.08.2019]

Evronim http://prevodjenje.mei.gov.rs/evronim/index.php?jezik=engl [accessed 05.08.2019]

eohsterm www.eohsterm.org [accessed 14.03.2019]

EUROTERMBANK (https://www.eurotermbank.com/) [accessed 05.08.2019]

EUSKALTERM (http://www.euskadi.eus/web01-apeuskal/eu/q91EusTermWar/kontsultaJSP/q91aAction.do) [accessed 05.08.2019]

France Terme (http://www.culture.fr/franceterme) [accessed 05.08.2019]

IATE https://iate.europa.eu; https://iate.europa.eu/home [accessed 14.03.2019]

SAPTerm www.sapterm.com [accessed 16.03.2019]

Termdat www.termdat.ch [accessed 15.03.2019]

Termium Plus https://www.btb.termiumplus.gc.ca [accessed 14.03.2019]

Verwaltungsglossar https://www.sprachressourcen.at/verwaltungsglossar/ [accessed 05.08.2019]

WIPO Pearl https://www.wipo.int [accessed 15.03.2019]

Termin https://jogi-terminologia.im.gov.hu/ [accessed 14.03.2019]

UNTERM (http://untermportal.un.org) [accessed 05.08.2019]

APPENDIX

EVALUATION SYSTEM FOR ONLINE TERMINOLOGICAL DATABASES

I. Background information on central termbases

A) GENERAL BACKGROUND

1. A survey of the termbases at the given domain:
   – historical background, date of creation, evolution, newly implemented or improved version of a former database;
   – diachronic (former versions) or synchronic description;
   – relationship with other termbases within or outside the organisation.
2. Grouping based on aim:
   – language policy and terminology policy aims;
   – translation-oriented termbase;
   – standardisation purposes.

B) IMMEDIATE BACKGROUND

1. The characteristics of the organisation that produced the termbase:
   – the background to the organisation, whether it has a national or international scope, domain, funding and type of organisation: international organisation, public administrative agency, research institute, university, translation agency.
2. Target users:
   – native speakers, speakers having a majority or minority status;
   – translators;
   – experts of a specific domain, e.g. engineers, lawyers, doctors etc.

3. Intended purpose:
  – descriptiveness;
  – prescriptiveness.

II. Technical parameters of a termbase

A) THE SOFTWARE
  – bespoke;
  – hybrid;
  – commercially available.

B) ACCESSIBILITY OF THE TERMBASE
  – open or restricted;
  – downloading of data allowed, supported platforms (e.g. smart phone, tablet).

III. Information on the content of the termbase

A) GENERAL FEATURES OF THE DATABASE
  1. Concept oriented:
    – purely onomasiological approach;
    – hybrid approach, containing semasiological elements as well.
  2. Simple, traditional or complex termbases:
    – number of structural elements (termbase, concept map, corpus);
    – level of detail in entries: type and number of data categories.

B) DETAILED STRUCTURE OF THE DATABASE
  1. At the level corresponding to megastructure:
    – directions for use and other forms of help: forum, chat service for users facing difficulties;
    – availability of copyright;
    – information on the number of concepts/entries and terms;
    – number of languages: monolingual, bilingual, multilingual;
    – number of domains and sub-domains.
  2. At the level corresponding to macrostructure:
    – search options: simple or advanced for terms, in the domain, subdomain, for word fragments; linguistic search only or concept map search as well; search in aligned corpus; links giving access to other data repositories;
    – display and organisation of search results (hit list or lack of);
    – nature of shown information (text, image, multimedia).

3. At the level corresponding to microstructure:
   – number of data categories and extent of elaboration;
   – types of data categories and its organisation:
     – display of data categories and clarity of entries: order and labelling of data categories;
     – respect for principles of terminology data modelling: beyond being concept oriented is there term autonomy, the level of realisation of a three-level hierarchical structure (entry level – index level – term level, if transparent through external descriptions), data elementarity and granularity;
     – mandatory and optional data fields;
     – data categories as per the work methods: is there a definition or not; the number of language variants of the definition; weather a definition can be replaced by a context; definition is for one concept or is there a contrastive purpose (e.g. existence of the data category of equivalence);
     – choice of data categories based on aims/subfield of terminology: standardisation, translation or language policy;
     – choice of data categories based on specific requirements of the domain (e.g. illustrations in an engineering termbase).
4. At the level corresponding to mesostructure:
   – cross references (links within the entries, links among entries);
   – external links (to other websites or repositories);
   – appearance of links (can be opened, accessibility, description of content).

IV. Information on the usage of the termbase

A) USER FRIENDLINESS OF THE DATABASE
   – understandability;
   – use of abbreviations;
   – visual elements;
   – easy to search.
B) UPDATES
   – frequency of updates;
   – general state of up-to-dateness (dates of recording).
C) RELIABILITY OF DATA
   – documentation: sources and their indication, information on reliability in each entry and its marking; date of recording for up-to-

dateness; star, icon, text to indicate reliability for the validated state; written sources, internal or external network of experts; the use of permanent or one-off links;

– editing options and reliability: organisational with internal and/or external experts, cooperation with other organisations, possibility for individual submissions, crowd-sourced and checked, results from machine translation with indication etc.

D) THE INNOVATIVE NATURE OF THE DATABASE

– contains extra elements: concept map, corpus;
– extra information: video tutorial, e-course;
– innovative display or lack thereof.

E) VALUE OF THE TERMBASE FOR SOCIETY AND ITS IMPORTANCE FOR THE PROFESSION

– aspects of language strategy and terminology policy;
– scope and type of experts interested;
– opportunities for training, e-course, workshop, or cooperation agreements (e.g. with universities).

INTERNETINIŲ TERMINŲ BAZIŲ VERTINIMO SISTEMA

Santrauka

Straipsnyje siekiama pateikti visapusišką klasifikavimo ir vertinimo sistemą, kuri būtų objektyvus pagrindas siekiant suprasti, aprašyti, klasifikuoti, vertinti ir peržiūrėti svarbiausias didesnių organizacijų internetines terminų bazes. Tokios visapusiškos leksikografinių žanrų, t. y. spausdintų ir internetinių žodynų, klasifikacijos jau esama, tačiau internetinėms terminų bazėms taikomos klasifikacijos nėra. Tyrimas grindžiamas Vengrijoje ir kitose šalyse paskelbta leksikografijos ir terminologijos sričių literatūra, ISO standartais, kelių internetinių terminų bazių, dalinės klasifikacijos analize, autorių atliktais tyrimais.

Terminų bazėms taikytinų kriterijų rinkinys parengtas atsižvelgiant į keturias svarbiausias kategorijas: pagrindinę informaciją, techninius parametrus, turinio informaciją ir naudojimą. Straipsnyje taip pat pateikiama schema, kurią būtų nesunku naudoti seminarų ar pamokų metu. Pamažu taikydamas tokį kriterijų rinkinį, vartotojas gali įgyti vis daugiau žinių ar net atrasti naujų funkcijų.

Preliminarioji vertinimo sistema atskleidžia, kad taikant beveik vienodus terminologinius metodus kuriamos terminų bazės turi tam tikrų ypatumų, nes įvairių sričių vartotojų poreikiai ir keliami tikslai numato skirtingus redagavimo principus. Be to, kai kurie ypatumai yra glaudžiai susiję ir juos dažnai sunkoka atskirti (pvz.: redagavimo principai, dokumentacija, patikimumas, kokybė).

Eksperimentinis skirstymas į kategorijas padeda geriau suvokti, išsamiau aprašyti ir atidžiau įvertinti naudojimosi internetinėmis terminų bazėmis ypatumus, pagaliau – suvokti, kaip svarbu kalbos specialistams tokiomis bazėmis naudotis. Tai skatintų vartotojus susidaryti apie terminų bazes objektyvią nuomonę ir ja dalytis žodžiu ar raštu, rašant apžvalginius straipsnius.

Gauta 2019-09-02

Dóra Mária Tamás
Hungarian Office for Translation and Attestation Ltd
Hungary, 1062 Budapest, Bajza utca 52

Termik Terminology Research Group at Károli Gáspár University
of the Reformed Church of Hungary
Hungary, 1146 Budapest, Dózsa György út 25
E-mail: tamas.dora@offi.hu; tamas.dora.maria@gmail.com

Eszter Sermann
University of Szeged
Faculty of Humanities and Social Sciences
Department for Italian Studies
Hungary, 6722 Szeged, Egyetem u. 2

Termik Terminology Research Group at Károli Gáspár University
of the Reformed Church of Hungary
Hungary, 1146 Budapest, Dózsa György út 25
E-mail eszter.sermann.lingo@gmail.com