Theresa Kruse/Ulrich Heid

LEARNING FROM STUDENTS

On the design and usability of an e-dictionary of mathematical graph theory

Abstract We created a prototype of an electronic dictionary for the mathematical domain of graph theory. We evaluate our prototype and compare its effectiveness in task-based tests with that of Wikipedia. Our dictionary is based on a corpus; the terms and their definitions were automatically extracted and annotated by experts (cf. Kruse/Heid 2020). The dictionary is bilingual, covering German and English; it gives equivalents, definitions and semantically related terms. For the implementation of the dictionary, we used LexO (Bellandi et al. 2017). The target group of the dictionary are students of mathematics who attend lectures in German and work with English resources. We carried out tests to understand which items the students search for when they work on graph-theoretical tasks. We ran the same test twice, with comparable student groups, either allowing Wikipedia as an information source or our dictionary. The dictionary seems to be especially helpful for students who already have a vague idea of a term because they can use the resource to check if their idea is right.

Keywords Dictionary use; specialised languages; terminology; terminography; mathematics

1. Introduction

Wikipedia "will eliminate the market for traditionally conceived specialised online dictionaries, which will disappear as commercial products if they cannot offer anything to justify their market price," says Fuertes-Olivera (2016). With this in mind, we want to learn more about the demands users have for specialised online dictionaries. The field of our study is the domain of graph theory and we created a prototype of an electronic dictionary on the subject which we evaluate with students (N=113). The participants are given ten tasks that can be solved with the help of our prototype and we asked them to comment on their working procedures. For comparison, a second group with the same tasks was instructed to complete them using Wikipedia (N=182). In this paper, we discuss the comments of both groups and focus on microstructure and access structure. The method is similar to the user-centered design approach which has previously been used to assess the usability of dictionaries, also motivated by the function theory (Tarp 2008). In particular, we would like to answer the following research questions:

- How beneficial is our dictionary for solving language-based mathematics tasks?
- What can we learn from student feedback about challenges, search workflows, efficiency and user satisfaction with a view to the development of future terminological e-dictionaries?

The findings may be applied to terminological dictionaries in other domains and provide additional insights into what terminological dictionaries have to offer to compete with Wikipedia.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

2. Related work

An increasing body of literature on the usability of electronic dictionaries has emerged in recent years. We only cite some exemplary work. A general review of the status around 2010 to 2015 is provided by Töpel (2014) and Nesi (2015); for a comparison between printed and electronic dictionaries see Dziemianko (2012). The majority of the publications focus on general language learner dictionaries, only a few on the acquisition of terminology. We only cite empirical results that discuss the design of e-dictionaries and how users interact with them; pure log-file analyses, studies on pocket dictionaries and meta-surveys are not mentioned.

2.1 Microstructure

The microstructure of e-dictionaries has been the subject of several user-oriented analyses. They mainly examine which items users would like to have included, which elements they use and how these items should be presented. Laufer/Hill (2000, p. 74) show that "even though a variety of dictionary information was available, most students opted for definitions, translations, or both". Li/Xu (2015) confirm this result for definitions, but in their study examples are rated high as well. We can conclude that users mostly demand definitions, translations and examples, depending on the particular use situation. Khairiah (2021) investigates what users expect from e-dictionary definitions: "The respondents consider that the information category of taxonomy/scientific name and characteristics are the most significant in a definition." Concerning the layout of the microstructure, Dziemianko (2015) shows that functional labels in colour significantly increase the speed and effectiveness of online dictionary searches over those in black and white. Chan (2014, p. 39) shows:

> The layout of a dictionary entry is also an important factor determining learners' preferences. A dictionary entry may be excessively long and may contain much more information than what a learner wants to get. The situation will be compounded if a target word is polysemous. [...] The use of highlighting features will of course help focus learners' attention, but a clear and systematic layout should help facilitate information search. As evidenced from the results of the study, learners prefer dictionary information which is clearly presented.

2.2 Role of illustrations

The role of illustrations in e-dictionaries is another area of research. Based on voluntary disclosure and an eye-tracking study, Kemmer (2014) finds that users consult pictures and text equally, with a slight preference for text. Lew et al. (2018, p. 73) confirm this and find "no evidence that the presence of pictorial illustrations leads to the neglect of the verbal definition." However, in the instance of graph theory, where certain concepts have typical graphical representations this might not be the case.

2.3 Search strategies

Another significant aspect is how users interact with electronic dictionaries and in particular which usage patterns appear. Laufer/Levitzky-Aviad (2006, p. 151) identify "translation, definition(s) and example(s)" as the most typical pattern for a bilingual dictionary, closely followed by "translation only". Aust et al. (1993) measure efficiency by consultations per

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

minute and compare printed and digitized dictionaries, finding that digital and bilingual (unlike monolingual) dictionaries have superior efficiency. Laufer/Hill (2000, pp. 68/77) conclude:

> The number of times the word is looked up during a learning session bears almost no relation to its retention. We postulated, albeit cautiously, that what matters is greater attention during the lookup rather than the number of lookups.

Chen (2017) confirm these results. Heid/Zimmermann (2012) compare a user interface that resembles a search engine to one that allows users to specify their needs in much detail:

> A comment made by most participants to the study was that they had needed some time to get acquainted with the profile-based dictionaries; this shows that these are not exactly conformant to users' a priori expectations. Our users however also noted that once the principles had been understood, the profile-based dictionaries were indeed more effective and efficient than the one-shot dictionary.

Lorentzen/Theilgaard (2012, pp. 654 f.) confirm the desire for a search engine-like user interface. Boonmoh (2021) let users freely choose between a translation tool and an e-dictionary and all of them used Google Translate in addition to other resources.

2.4 Method: work with user comments

Hult (2014) analyses and classifies dictionary users' comments into five categories: criticism, suggestions for improvement, praise, other and nonsense. If a comment fit into several categories, it is divided. She also examines how often different parts of the microstructure are mentioned in the comments and if so whether positively or negatively. Miller et al. (2017) analyse user comments qualitatively and discover audio, easy to understand, easy to use, example sentences, free access and pictures as popular features while font size, offline use, number of entries and example sentences leave room for improvement. He (2017) uses a grounded theory approach to classify comments and finds the following categories, with sub-categories listed in parentheses: comprehension (definition, collocation, illustration), production (thesaurus, example, pragmatic information, usage notes, morphology), information accessibility (headword selection, neologism, pronunciation, encyclopedic information). Farina et al. (2019) use think-aloud protocols to examine how the dictionaries in use could be improved. They find that users choose dictionaries based on search engine rankings, that part of speech and lemmatisation are neglected during the search process and that users desire a sufficient amount of examples. Corris et al. (2000) divide user comments into four categories: (i) attitudes of users and makers to dictionaries, (ii) exhaustiveness, (iii) functionality and (iv) practical considerations. The results given in sections 2.1 and 2.3 are largely reflected in the aspects arising from qualitative study categories. This is especially true when it comes to the value of examples.

2.5 User skills in dictionary use

Based on research with business students, Gromann/Schnitzer (2016) find a general lack of awareness of current specialised resources and alternative ways for processing specialised texts. Chan (2017) shows that although bilingual dictionaries determine noun countability and associated article use, learners often misinterpret dictionary information on this prop-

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

erty, resulting in article errors or wrong countability judgments. Laufer (2011, p. 45) observes that "some learners could not find some target collocations in the dictionaries even though they were included there" and "sometimes learners were not aware of the fact that the collocation was unfamiliar to them and did not seek dictionary help". Chen (2017, p. 225) states:

> The participants showed inadequate dictionary use skills. They were reluctant to use the hyperlink function of the electronic dictionary to further look up relevant information or too careless to notice it, unable to distinguish between entry subsenses, inclined to choose the sense listed at the beginning of an entry, and apt to lose patience when faced with overcrowded entry information.

3. The dictionaries

3.1 Electronic dictionary of graph theory

We developed an electronic dictionary for the domain of graph theory (*GraWB* for *Graphentheoretisches Wörterbuch*). The process of lemma selection is discussed in Kruse/Heid (2021) and the microstructure in Kruse/Heid (2020). In the prototype used in the study, 452 German and 160 English lemmas have entries. The information in the microstructure depends on the headword, as each term is assigned to a semantic category such as *graph type*, *algorithm*, or *graph property*. Definitions are given for most of the German and some of the English terms. Access paths are a lemma list and a search field. Category-based access is planned, but not yet implemented. Currently, the dictionary is only accessible after login, not via a search engine.[1]

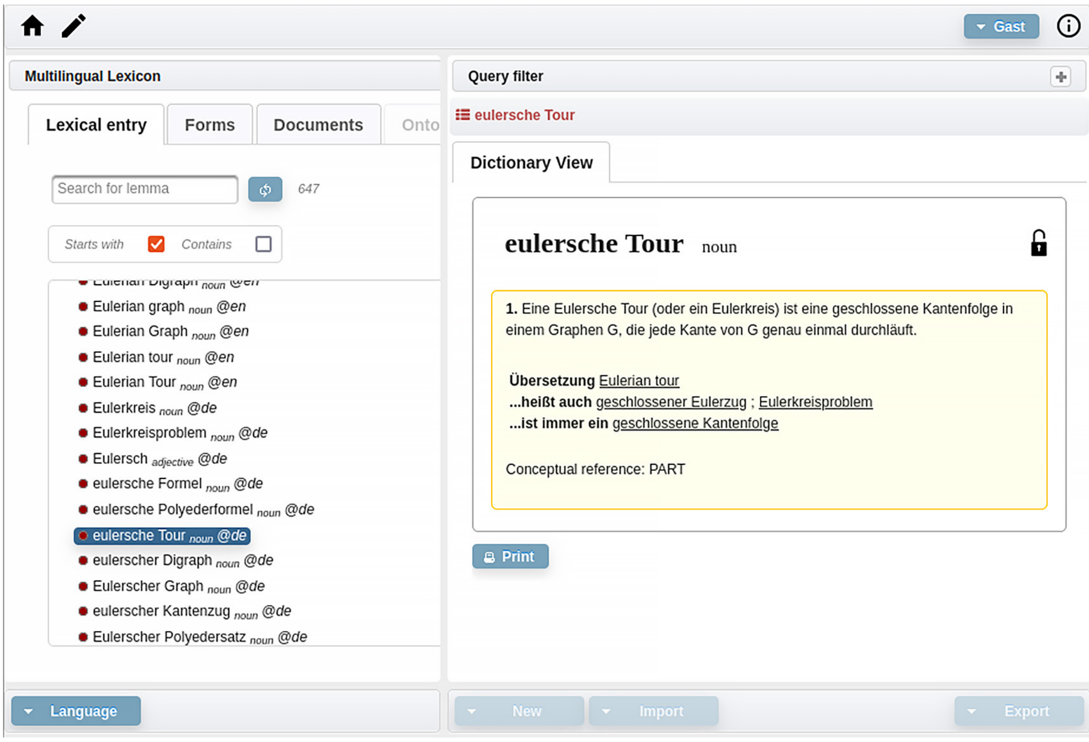

Fig. 1: Interface of GraWB

[1] https://lexo.hosting.uni-hildesheim.de/LexO/faces/loginView.xhtml.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

The dictionary is implemented in the LexO framework (Bellandi et al. 2017). LexO was created for terminologists and is optimized for desktop or tablet use. It has a panel and a tab view, however, we only use the panel view for our research. Figure 1 contains a screenshot of the user interface and an example entry. GraWB does not contain any illustrations yet. We identified a few issues during a cognitive walk-through with a usability expert, which we communicated to the developers, but which are outside the scope of this paper.

3.2 Wikipedia entries for graph theory

Wikipedia is a well-known online encyclopedia. Nevertheless, it has not been the topic of many (meta-)lexicographic analyses (Fuertes-Olivera 2009; Mederake 2014), likely because as an encyclopedia, its article structure differs massively from that of a typical (monolingual) specialised dictionary.

Although most of the articles include links to equivalents in other languages, the contents of the articles differ greatly because they are written collaboratively and independently. The links can be, however, used to detect translation equivalents. The articles consist of continuous text divided into (sub-)sections. Terms with their own article are highlighted and linked. Articles can include illustrations, videos, or sound files.

The articles can be mainly accessed via a search field and the above-mentioned links. As Wikipedia articles rank highly in most search engines, they can also be accessed this way. Inside Wikipedia, access paths like lists of articles related to a specific topic, e.g. *list of graphs* or *glossary of graph theory*, are available. As anyone can modify most parts of an article, the encyclopedia is not static but its articles may change in contents and form over time.

On a content level, the main difference between Wikipedia and GraWB is that Wikipedia does not focus on a particular domain but provides information on a vast range of domains. This may lead to cognitive overload in some users. Additional effort is needed when users have to extract the needed information from the texts.

4. Experimental setup

4.1 Study design and participants

We conducted our study with two groups of students: one in summer 2020 and the other in summer 2021. An online survey was active for around six weeks each time. The participants in both studies were given identical tasks with the exception that the first time they were only allowed to use Wikipedia as a resource (N=182), while the second time only GraWB (N=113). In the second study, the NASA task load index[2] was employed as well.

[2] https://humansystems.arc.nasa.gov/groups/TLX/.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

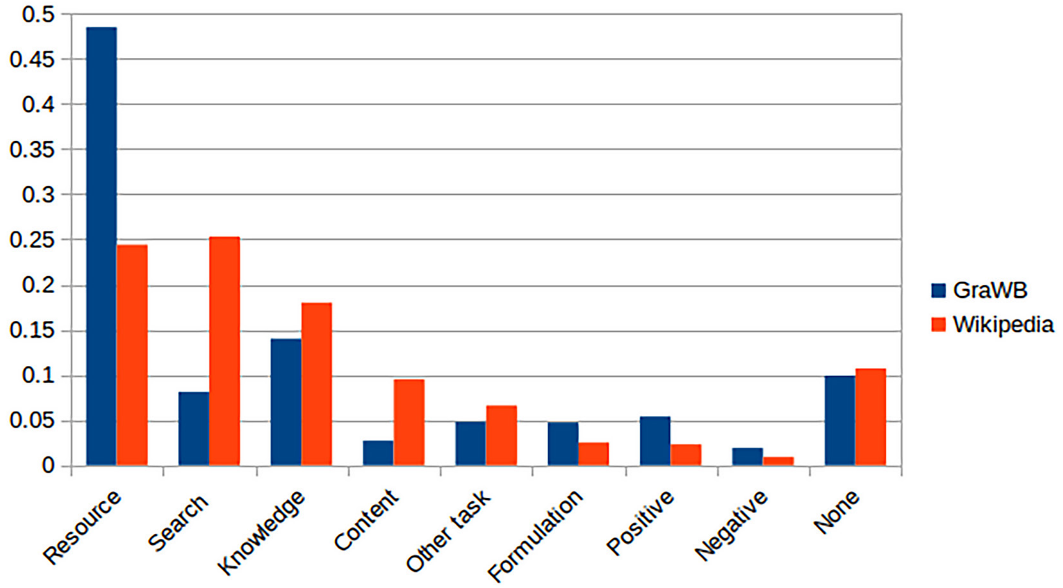

Fig. 2: Distribution of the comments over the categories

The tasks were written in German and distributed through a lecture on algebra and number theory at the University of Hildesheim attended by teacher students. The participants accessed the online survey individually. Both groups had taken similar courses in their study program and had already basic mathematical but no specific knowledge of graph theory. We also inquired about first languages, graph theory courses attended and study semester, to gain a better understanding of the impact of background knowledge. In the second survey, the participants also had to indicate whether they had already taken part in the first study which was the case for about 25% of the subjects but we do not find a significant difference (t-test) for the number of search terms, the time used and the points awarded between the two groups. We suggested taking the survey on a desktop-like surface because it allows switching between browser tabs more easily.

4.2 Tasks and types of searches

The study consists of six tasks with subtasks. They address different aspects of terminology. In task (1), the participants have to give the properties of graphs which are only represented graphically (namely a binary tree, a cycle graph and the Petersen graph), i. e. no search term is given. This allows us to evaluate whether the students describe the graphs only in general language or if they use terminology. Task (2) shows which terms the students use to describe a mathematical concept they do not know by name. Both tasks cannot be solved by mere look-up, but rather by following paths through the conceptual domain, some sort of 'ontological inference'. Tasks (3) and (4) relate to the classification of mathematical terms by Vollrath (1978): We chose a term with a homograph in general language with the same meaning as well as a term where the general language meaning differs considerably to examine how the participants apply these terms. Task (5) addresses search strategies, as we ask the participants to name algorithms for a given problem. Task (6) is to translate two sentences from English to German which together contain seven target terms. The participants obligatorily had to comment on each task. In section 5, we delve deeper into tasks (1.1), (3.1), (4) and (6.1).

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

4.3 Methods for data analysis

We combine numerical scoring and an analysis of the users' textual comments to evaluate our data. Each response is given a score; the authors did all of the scoring work task-wise. In tasks (1) and (2), 1 point is given if the answer provides a graph-theoretical perspective, 0 points for a non-graph-theoretical perspective and 0.5 points for answers in between. In tasks (3) and (4), 1 point is given for a correct answer, 0 points for a wrong answer and 0.5 points for a semi-correct answer. In task (5), 1 point is given if at least two algorithms are named, 0.5 points for only one and 0 for none. In task (6), one point for each target term can be earned.

The lengths of the submitted comments range between 1 and 910 characters; the average length is 86.17 with a standard deviation of 82.19 and a median of 64. Each comment is assigned to one of the following categories:[3]

- design and content of the resource, e. g. *Article nicely manageable*
- search strategy, e. g. *I found the term by following the blue highlighting*
- prior knowledge, e. g. *I picked it up somewhere*
- mathematical content, e. g. *Looks like stochastic*
- reference to another task, e. g. *derived from the previous question*
- references to the formulation of the particular task, e. g. S*omewhat unclear to what extent the question should be answered...*
- general positive remark, e. g. *No problem with the task*
- general negative remark, e. g. *Outch, I'm so bad*
- none/other, e. g. *No comment*

Figure 2 shows the distribution of the comment categories. The majority of the comments are concerned with the resource and the search strategies used. For those, the following subcategories were found:

- no use of the resource because it would not help anyway, e. g. *Again, a translation app/ website is the better option*
- lack of a helpful search term or access structure, *e. g. I found it very difficult to find a search term for this.*
- nothing found, e. g. *I somehow couldn't really find anything.*
- found a helpful article but did not understand it, e. g. *I did not understand the entries*
- description of a successful search, e. g. *The explanation was quickly found*
- general remark on the resource, e. g. *Unfortunately, no example on the page, then it would certainly be easier to understand.*
- The categories *nothing found, no access structure, not understood* mainly apply to comments by participants who did not provide a high rated answer.

5. Results and interpretation

5.1 Effectiveness: usefulness for the tasks

In terms of quantitative evaluation, we calculate the Spearman correlation between the number of searches and rewarded points (Spearman 1904). We get 0.2516 ($p<0.01$, two-tailed) for Wikipedia and 0.2136 ($p<0.05$, two-tailed) for GraWB when all sub-tasks are

[3] The example comments have been translated into English by the authors.

counted equally, demonstrating a small correlation with either resource. The results change when we calculate the coefficient separately for task (6) and the others. With Wikipedia, we have 0.2398 ($p<0.01$, two-tailed) for tasks (1) to (5) and −0.1193 (not significant) for task (6) while it is 0.1560 (not significant) for tasks (1) to (5) and 0.2180 ($p<0.05$, two-tailed) for task (6) using GraWB. We only detect small correlations, however, the positive influence of GraWB is stronger in the translation tasks, as could be expected because it is – unlike Wikipedia – optimized for such use situations. In the following, we analyse four tasks in detail.

5.1.1 Task: describe a given graph

In task (1.1), the participants had to name properties of a binary tree, presented to them as a picture. In the Wikipedia study, 47% used a graph-theoretical description, but 75% did in the GraWB-study, while 32% took a stochastic approach with Wikipedia and only 11% with GraWB. The rest used a combination. The reason for the difference is probably that Wikipedia users got side-tracked by information about the use of graph-theoretical concepts in other domains like stochastic, while GraWB only presents graph theory.

In the following, we list the top five search terms used in that task for each of the two studies. In parentheses, we give the absolute number of participants who used this search term, followed by the average rating of the term (Likert scale, *1=very helpful*, *5=not helpful at all*) and a translation. Most terms relate to *tree*, although some are graph-theoretical and some are stochastic, as evidenced by the answers.

Wiki: Baumdiagramm (50, 2.42, *tree diagram*), Graphentheorie (29, 2.04, *graph theory*), Graph (18, 3, *graph*), Baum (Graphentheorie) (8, 1.75, *tree* (*graph theory*)), Eigenschaften Baumdiagramm (7, 2.57, *properties tree diagram*)

GraWB: Baum (50, 2.04), Graph (27, 3.37), binärer Baum (11, 2.18, *binary tree*), Kante (11, 2.91, *edge*), Knoten (9, 2.56, *node*)

In both trials, the majority of those who use a dictionary provides a graph-theoretical perspective (63% for Wikipedia and 80% for GraWB) with GraWB reaching a higher percentage, as expected. For those who did not use a dictionary we find a wider disparity although we had expected similar results: In the Wikipedia study 66% of those not using a dictionary got 0 points but only 28% without GraWB did. A reason could be the small number of participants not using a dictionary (65 for Wiki, 32 for GraWB) which makes it difficult to draw further conclusions.

5.1.2 Task: apply a term – part 1

In task (3), the participants had to answer the question *What is N adjacent to?* for an indicated node *N* in a labeled Petersen graph. The German formulation of the task (*Wozu ist N adjazent?*) contains the interrogative adverb *wozu*, due to the syntactic construction of the adjective (*x ist adjazent zu y*). This syntactic information is not derivable from the respective Wikipedia articles. The answers of three participants in each study indicate that they misunderstood *wozu* as *for which purpose*, which may be seen as evidence for the need to give constructional information also in specialized dictionaries. In the following, we present again the search terms and values:

Wiki: adjazent (127, 1.35), Nachbarschaft (Graphentheorie) (19, 1.05), Adjazenzmatrix (10, 2.1), adjazent⟶Nachbarschaft (Graphentheorie) (6, 1.5), Pentagramm (3, 3), Nachbarschaft (3, 2.67)

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

GraWB: adjazent (100, 1.15), adjazente Ecke (20, 1.25), adjazenter Knoten (14, 1.29), benachbart (5, 2), Ecke (4, 1.25)

We assume that those students who give *Nachbarschaft* (*Graphentheorie*) used *adjazent* as a search term because Wikipedia automatically forwards *adjazent* to *Nachbarschaft* (*Graphentheorie*). In the Wikipedia study, 98.35% used the resource, while 90% did in the GraWB study. In both studies, those who did not use a dictionary solved the task correctly with prior knowledge, except for two GraWB-users who claim that they could not find anything helpful. We see that the resources support finding the meaning of terms only used in terminology with Wikipedia having a slight advantage as an already established resource which students are familiar with.

5.1.3 Task: apply a term – part 2

In task (4), the participants answered the question *What are the leaves of the graph?* for a given labeled graph. *Leaf* is an example of a mathematical term that is homonymous with a general language word, but its meaning cannot be easily derived.

Wiki: Blätter und innere Knoten in der Graphentheorie (32, 1.25), Baum (Graphentheorie) (20, 2.2), Blätter Graph (17, 2.29), Blätter Graphentheorie (17, 2.53), Blätter Graphen (16, 3.31)

GraWB: Blatt (87, 1.60), Blatt eines Baumes (20, 4.15), Grad der Ecke (15, 1.33), Blätter (8, 4), Baum (8, 1.63)

When searching Wikipedia for *Blätter Graph* or *Blätter Graphen* the article *Baum* (*Graphentheorie*) is suggested while most search engines return the Wikipedia article *Blätter und innere Knoten in der Graphentheorie* as a prominent result. The first two options in GraWB when typing *Bl* are *Blatt* and *Blatt eines Baumes.* Unfortunately, the GraWB-search does not include lemmatization, therefore *Blä* yields no results. In the Wikipedia study, 94% used the dictionary and 84% did in the GraWB study.

5.1.4 Task: translate a sentence

The participants were asked to translate the sentence *A square grid graph may have a spanning tree* to German with the target terms *square grid graph* and *spanning tree.*

Wiki: spanning tree (41, 1.98), square grid graph (20, 2.7), grid (15, 2.53), square grid (13, 2.85), A square grid graph may have a spanning tree (12, 4.12)

GraWB: spanning tree (58, 1.57), square grid graph (57, 1.47), Gittergraph (9, 2.22), Spannbaum (6, 3), square grid (4, 4).

As expected, the target terms are the most popular search terms. When searching German Wikipedia for s*panning tree*, its equivalent *Spannbaum* appears among the top five suggestions. Analogously, *square grid graph* leads to *Gittergraph.* Some participants searched for the complete sentence (cf. translation engine metaphor) but were not very satisfied with the results. The participants are slightly more satisfied with the outcomes in GraWB, but this is probably because Wikipedia is not optimized for translation. Only 52% used Wikipedia in this task, while 74% used GraWB.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/
Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society.
Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

5.2 Design aspects of the dictionary

In this section, we present further aspects mentioned by the participants. One example, already introduced in section 4.3, is the access path, as several participants struggle to come up with a suitable search term. Different reasons cause this problem: Some users have no notion of how to complete the task and others have a concept in mind but do not know the corresponding terminology. A graphical access structure with pictures could help both. At least in graph theory, such approaches appear feasible, while it remains to be investigated if it also applies to other subdomains.

Another idea is to include non-terminological quasi-synonyms of terms in the lemma list and introduce links to the relevant terms. For example, the term *isomorphic* denotes an identity of graphs even if they are visualized differently. An entry *same as*, appropriately marked as a general language item, could link to an entry for *isomorphic*, relating an everyday non-terminological expression to the actual term. It has to be evaluated from a didactic perspective if such a device aids or hinders acquiring the right terminology as it can be problematic for those who acquire new terms as opposed to those who recall terms they have already learned.

The users comment differently on links: Some regard them as helpful, while others are annoyed by clicking on several links until they reach their goal. Alternative access structures may help to reduce the number of links that must be used, e. g. clusters like those introduced by *Wortgeschichte digital*.[4]

As graph theory strongly relies on visualizations, it is natural that the users would have liked the dictionaries to include (more) illustrations. While no positive effect of illustrations could be found for learning languages (cf. section 2.2) it could be explored in further research if that also holds for terminology.

We further see preferences for using similar strategies as in search engines while other participants indicate that they developed a routine of dictionary use in the course of the study. Sentiment analysis of user feedback shows that the overall satisfaction is highly task-dependent. In tasks (1), (2) and (5), the comments are slightly more positive in the Wikipedia study, while in the others, GraWB has more positive comments.

6. Limitations of the study

Due to the pandemic, we could not use tracking software but had to rely on voluntary disclosure by the users. Thus, we cannot be sure if they visited other articles but did not give them in the form they had to fill, as some comments mention translation tools for task (6) in the Wikipedia study. Another caveat is that some users gave their actual search terms, while others gave the first article they were forwarded to or clicked on. Further, some comments indicate that the users access Wikipedia articles through a search engine. While within Wikipedia, results only depend on the search term, most search engines personalize the ranking of the results and thus we can only vaguely retrace the search paths. Furthermore, some search results are highlighted by the engines with a picture, mostly a Wikipedia article, so that users are tempted to click on it without investigating further results. Finally, we

4 https://www.zdl.org/wb/wortgeschichten.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

did not vary the order of the tasks and therefore carry-over effects may appear. For replication, it has to be considered that Wikipedia articles change and that all participants may access different versions.

7. Conclusion and future work

The present study can be the starting point of a design-based research study on the development of terminological resources in mathematics. That way the ideas given by the students can be further investigated. The same applies to other access paths where articles are suggested after an associated search term. Further ideas are to develop the mixed-methods approach in either a purely quantitative or a purely qualitative study. Not much work has so far gone into the user-centered design of specialized dictionaries. The present study, although carried out on a particular subdomain of mathematics, led to several conclusions that may be valid more generally.

The comparison between our dictionary of graph theory and Wikipedia showed that a specialized dictionary avoids that users get side-tracked by information on related domains; and that the particularized microstructure of a dictionary often allows for more efficient access to individual items than the explanatory texts of the encyclopedia and the different types of links contained therein. We also learned that both types of resources require access by lemmas or related items and that students who do not know which search term to use have difficulties in accessing the necessary information. The encyclopedia articles tend to be somewhat more helpful in such situations, but are still far from satisfactory. Whether the inclusion of lay terminology into the dictionary, accompanied by references to specialized terms of the domain, is an appropriate solution remains to be investigated. Likewise, more research is necessary to understand whether some kind of picture-based search is feasible for a limited and visually rather straightforward domain like graph theory, to cater for such situations.

While the translation advantages of the dictionary are expected, the fact that some users deal with the dictionary the same way as with an online translation system (enter full sentences, wait for the translation result) was less expected, yet it shows the extent to which generic online tools influence user expectations. Future specialized e-dictionaries will indeed have to offer specific advantages (e. g. accuracy, ease of access, classification, functions of an ontological categorization, equivalents) over generic online tools to be recognized and used by the public.

In terms of methodology, we see a need for refined methods for user-centered analyses of specialized dictionaries: In a quantitative study, further data could be collected, e. g. not based upon voluntary disclosure but on log files and by measuring the exact time for each task or search. A qualitative study could include extensive interviews with the participants to gain further insights into their working procedures. These would not only be interesting from a lexicographic but also from a didactic perspective. Another approach could be to let the tasks be done twice by the same group as a pre- and post-study. It might be also interesting to introduce the dictionary to the participants but let them choose if they use Wikipedia, our resource, or any other device.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

References

Aust, R./Kelley, M. J./Roby, W. (1993): The use of hyper-reference and conventional dictionaries. In: Educational Technology Research and Development 41 (4), pp. 63–74. http://doi.org/10.1007/BF02297512.

Bellandi, A. et al. (2017): Developing LexO: a collaborative editor of multilingual lexica and termino-ontological resources in the humanities. In: Proceedings of Language, Ontology, Terminology and Knowledge Structures Workshop (LOTKS 2017). Montpellier, France. https://aclanthology.org/W17-7010.

Boonmoh, A. (2021): Use of dictionaries and online tools for reading by Thai EFL Learners in a naturalistic setting. In: Lexikos 31 (1). http://doi.org/10.5788/31-1-1645.

Chan, A. Y. W. (2014): Using LDOCE5 and COBUILD6 for meaning determination and sentence construction: What do learners prefer? In: International Journal of Lexicography 27 (1), pp. 25–53. http://doi.org/10.1093/ijl/ect034.

Chan, A. Y. W. (2017): The effectiveness of using a bilingualized dictionary for determining noun countability and article selection. In: Lexikos 27 (1). http://doi.org/10.5788/27-1-1399.

Chen, Y. (2017): Dictionary use for collocation production and retention: a CALL-based study. In: International Journal of Lexicography 30 (2), pp. 225–251. http://doi.org/10.1093/ijl/ecw005.

Corris, M. et al. (2000): Bilingual dictionaries for Australian languages: user studies on the place of paper and electronic dictionaries. In: Heid, U. et al. (eds.): Proceedings of the 9th EURALEX International Congress. Stuttgart, pp. 169–181.

Dziemianko, A. (2012): On the use(fulness) of paper and electronic dictionaries. In: Granger, S./Paquot, M. (eds.): Electronic lexicography, pp. 319–342. http://doi.org/10.1093/acprof:oso/9780199654864.003.0015.

Dziemianko, A. (2015): Colours in online dictionaries: a case of functional labels. In: International Journal of Lexicography 28 (1), pp. 27–61. http://doi.org/10.1093/ijl/ecu028.

Farina, D. M. T. C./Vrbinc, M./Vrbinc, A. (2019): Problems in online dictionary use for advanced Slovenian learners of English. In: International Journal of Lexicography 32 (4), pp. 458–479. http://doi.org/10.1093/ijl/ecz017.

Fuertes-Olivera, P. A. (2009): The function theory of lexicography and electronic dictionaries: Wiktionary as a prototype of collective free multiple-language internet dictionary. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.): Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools today. (= Linguistic insights, 1424-8689 v. 90). Bern/New York, pp. 99–134.

Fuertes-Olivera, P. A. (2016): A cambrian explosion in lexicography: Some reflections for designing and constructing specialised online dictionaries. In: International Journal of Lexicography 29 (2), pp. 226–247. http://doi.org/10.1093/ijl/ecv037.

Gromann, D./Schnitzer, J. (2016): Where do business students turn for help? An empirical study on dictionary use in foreign-language learning. In: International Journal of Lexicography 29 (1), pp. 55–99. http://doi.org/10.1093/ijl/ecv027.

He, Y. (2017): Dictionary user research in the digital era: value of user-generated content. In: Proceedings of ASIALEX 2017, pp. 676–693. https://asialex.org/pdf/Asialex-Proceedings-2017.pdf.

Heid, U./Zimmermann, J. T. (2012): Usability testing as a tool for e-dictionary design: collocations as a case in point. In: Fjeld, R. V./Torjusen, J. M. (eds.): Proceedings of the 15th EURALEX International Congress. Oslo, pp. 661–671.

Hult, A.-K. (2014): The authentic voices of dictionary users – viewing comments on an online learner's dictionary before and after revision. In: Abel, A./Vettori, C./Ralli, N. (eds.): Proceedings of the 16th EURALEX International Congress. Bolzano, pp. 237–247.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Kemmer, K. (2014): Rezeption der Illustration, jedoch Vernachlässigung der Paraphrase? Ergebnisse einer Benutzerbefragung und Blickbewegungsstudie. In: Müller-Spitzer, C. (ed.): Using online dictionaries, pp. 251–278. http://doi.org/10.1515/9783110341287.251.

Khairiah, D. (2021): Definition model for plant and animal AME lemmas in KBBI V: a user study. In: Proceedings of ASIALEX 2021, pp. 142–149. https://asialex.org/pdf/Asialex-Proceedings-2021.pdf.

Kruse, T./Heid, U. (2020): Lemma selection and microstructure: definitions and semantic relations of a domain-specific e-dictionary of the mathematical domain of graph theory. In: Gavriilidou, Zoe/Mitsiaki, Maria/Fliatouras, Asimakis (eds.): Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress, 7–9 September 2021, Alexandroupolis. Volume 1. Alexandroupolis, pp. 227–233.

Kruse, T./Heid, U. (2021): From term extraction to lemma selection for an electronic LSP-dictionary in the field of mathematics. In: Electronic lexicography in the 21st century: post-editing lexicography, pp. 572–587.

Laufer, B. (2011): The contribution of dictionary use to the production and retention of collocations in a second language. In: International Journal of Lexicography 24 (1), pp. 29–49. http://doi.org/10.1093/ijl/ecq039.

Laufer, B./Hill, M. (2000): What lexical information do L2 learners select in a CALL dictionary and how does it affect word retention? In: Language Learning & Technology 3 (2), pp. 58–76.

Laufer, B./Levitzky-Aviad, T. (2006): Examining the effectiveness of ‘Bilingual Dictionary Plus’ – a dictionary for production in a foreign language. In: International Journal of Lexicography 19 (2), pp. 135–155. http://doi.org/10.1093/ijl/eck006.

Lew, R. et al. (2018): Competition of definition and pictorial illustration for dictionary users’ attention: An Eye-Tracking Study. In: International Journal of Lexicography 31 (1), pp. 53–77. http://doi.org/10.1093/ijl/ecx002.

Li, L./Xu, H. (2015): Using an online dictionary for identifying the meanings of verb phrases by Chinese EFL learners. In: Lexikos 25. http://doi.org/10.5788/25-1-1295.

Lorentzen, H./Theilgaard, L. (2012): Online dictionaries – how do users find them and what do they do once they have? In: Fjeld, R. V./Torjusen, J. M. (eds.): Proceedings of the 15th EURALEX International Congress. Oslo, pp. 654–660.

Mederake, N. (2014): Artikel der Wikipedia aus lexikografischer und textlinguistischer Perspektive. In: Mann, M. (ed.): Digitale Lexikographie: ein- und mehrsprachige elektronische Wörterbücher mit Deutsch: aktuelle Entwicklungen und Analysen. (= Germanistische Linguistik 223–224). Hildesheim, pp. 229–249.

Miller, J./Kwary, D. A./Setiawan, A. W. (2017): Koalas, Kiwis and Kangaroos: the challenges of creating an online Australian Cultural Dictionary for learners of English as an additional language. In: Lexikos 27 (1). http://doi.org/10.5788/27-1-1405.

Nesi, H. (2015): Thirty years of user studies – and what we still need to find out. In: Li, L./Mckeown, J./Liu, L. (eds.): Proceedings of ASIALEX 2015 Hong Kong: words, dictionaries and corpora: innovations in reference science, pp. 1–8.

Spearman, C. E. (1904): The proof and measurement of association between two things. In: The American Journal of Psychology 15 (1), pp. 72–101. http://doi.org/10.2307/1412159.

Tarp, S. (2008): Lexicography in the borderland between knowledge and non-knowledge. General lexicographical theory with particular focus on learner’s lexicography. Tübingen.

Töpel, A. (2014): Review of research into the use of electronic dictionaries. In: Müller-Spitzer, C. (ed.): Using online dictionaries, pp. 13–54. http://doi.org/10.1515/9783110341287.13.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

Vollrath, H.-J. (1978): Lernschwierigkeiten, die sich aus dem umgangsprachlichen Verständnis geometrischer Begriffe ergeben. In: Bauersfeld, H./Otte, M./Steiner, H.-G. (eds.): Lernschwierigkeiten im Mathematikunterricht. Bielefeld, pp. 57–73.

Contact information

Theresa Kruse
University of Hildesheim
kruset@uni-hildesheim.de

Ulrich Heid
University of Hildesheim
heidul@uni-hildesheim.de

Acknowledgments

We would like to thank Andrea Bellandi, Simone Marchi and Silvia Piccini from the University of Pisa for their support by setting up LexO, as well as Jürgen Sander and Martin Kreh for their support in executing the study and Lea Wöbbekind for her insights on usability.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Proceedings of the XX EURALEX International Congress. Mannheim: IDS-Verlag.