EURALEX XIX

Congress of the European Association for Lexicography

Lexicography for inclusion

7-11 September 2021

Ramada Plaza Thraki
Alexandroupolis, Greece

www.euralex2020.gr

Proceedings Book
Volume 1

Edited by Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

EURALEX Proceedings

ISSN 2521-7100

ISBN 978-618-85138-1-5

Edited by: Zoe Gavriilidou, Maria Mitsiaki, Asimakis Fliatouras

English Language Proofreading: Lydia Mitits and Spyridon Kiosses

Technical Editor: Kyriakos Zagliveris

2020 Edition

Lemma Selection and Microstructure: Definitions and Semantic Relations of a Domain-Specific e-Dictionary of the Mathematical Field of Graph Theory

Kruse T., Heid U.

University of Hildesheim, Germany

Abstract

We design a bilingual electronic dictionary for the mathematical domain of graph theory. The target group of the dictionary are students in the field, and the dictionary should support them in both cognitive and communicative situations. Therefore, it will not only provide equivalents but also an ontology of the terminology. The dictionary is based on a corpus and the lemmas are selected by combining results of automatic extraction tools with the work of expert raters. For the microstructure, a domain-specific scheme is developed and presented. The lemmas are divided into nine categories (one for adjectives, one for verbs and seven for nouns). In addition, we introduce thirteen semantic relations for which information can be given in the microstructure, depending on the category of the lemma. The microstructure items for each semantic relation are introduced by means of a specific indicator phrase, as the target group might not be acquainted with the linguistic terminology.

Keywords: LSP-dictionary; microstructure; lemma selection; mathematics; pedagogical lexicography

1 Introduction

We plan to develop a bilingual e-dictionary for the mathematical domain of graph theory. Besides the equivalents in German and English, information on the relations between the concepts of the domain will be given, as an ontology forms the backbone of the dictionary. The aim of the dictionary is to meet cognitive as well as communicative needs. Therefore, aspects of domain-specific and pedagogical lexicography have to be combined in this project. Later on, we plan to determine if and how the dictionary influences the LSP-skills of the students compared to usual aids like Wikipedia.

Of course, one may wonder why we do not put our effort into the digitization of one of the already existing mathematics dictionaries. This might be sensible if the aim of the project simply was to have the dictionary as a product. Our project, however, also includes the development of a method to find the lemmas and the conceptual and/or semantic relations using linguistic patterns which are typical for the language of the domain (cf. Kruse & Giacomini 2019). The aim is to develop a generalizable method which makes it easier to create electronic corpus-based dictionaries for other sub-domains of mathematics as well.

In this paper, we will present the current state of the dictionary development regarding lemma selection and microstructure. The results may be applied to future mathematics LSP dictionary projects as well. In Section 2 we give an outline regarding the target group of the pedagogical dictionary. Section 3 gives an overview of mathematical lexicography by introducing the work of Eisenreich (2008) on printed dictionaries and presenting already existing electronic dictionaries for the domain of mathematics. In Section 4, we present our corpus. Section 5 introduces the process of lemma selection and a category system for classifying the lemmas. Based on that, we outline in Section 6 the planned microstructure of the dictionary focusing on the presentation of definitions and of conceptual relations. A conclusion and an outlook towards future developments is given in Section 7.

2 The User Group

The intended user group of the dictionary are mathematics students attending lectures on graph theory. Therefore, we consider our planned dictionary as a pedagogical dictionary. As Tarp (2011) has pointed out, there is some discussion on the question under which conditions a dictionary might be considered as pedagogical. Nevertheless, we use the terminology introduced by Gouws (2010) and regard the target group of our dictionary as semi-experts, as they already have basic mathematical knowledge. We classify them as advanced learners in the specialized language of the domain. The dictionary should help them in cognitive and communicative situations (cf. Fuertes-Olivera & Tarp 2014; Tarp 2008): They have to read and understand papers in English, which is generally their L2; and they have to give presentations or write theses in German, which is generally their L1. The target group as well as the functions of the dictionary were already described in detail in Kruse & Giacomini (2019).

3 Lexicography and Mathematics

Eisenreich (2008) gives an overview of mathematics lexicography with a focus on printed dictionaries. Nevertheless, some of his results are also relevant for electronic dictionaries. He states that mathematics dictionaries tend to be out of date rather soon after publication as constantly new terminology comes up. An electronic dictionary seems to be appropriate to deal with this obstacle, as it can be updated much more easily than a printed one. Further, Eisenreich (2008) recommends to focus only on a sub-domain of mathematics for writing a dictionary, since there exist several terms which have multiple meanings depending on the particular sub-domain. Recognizing this problem, our project is restricted to the sub-domain of graph theory.

It is difficult to make a clear separation between encyclopaedic and terminological works when dealing with LSP-dictionaries (Adamska-Sałaciak 2012). There are overlaps between the two areas and they are not clearly distinguishable. Nevertheless, we try to give an overview of existing mathematics lexicographic works (see Table 1). The main focus is on German and English resources. The list is by far not complete, especially as there exist a lot of private projects. Eisenreich (2008) has divided his overview into the following categories: (1) monolingual dictionaries in the narrower sense, (2) overall mathematics reference works, (3) elementary mathematics for the general public, (4) multilingual dictionaries. The same categorization may be applied for electronic dictionaries, but it seems reasonable to divide category (3) into didactic resources for pupils or the lay public vs. works for an academic audience. Furthermore, we will merge categories (1) and (2), as they are difficult to distinguish. Additionally, we want to look at content and form as two different dimensions. Therefore, we first distinguish between monolingual, bilingual and multilingual resources. The second dimension concerns the lemma selection and the purpose: school, academic, general public. Thirdly, we checked whether the dictionaries cover our topic of graph theory. So, in all of the resources, we looked up the term *graph* to see to which degree graph theory is considered in the particular work. In the context of scientific textbooks about the domain of graph theory we expect to find this word describing a graph in the discrete sense, consisting of edges and nodes; whereas in school mathematics it will rather refer to the graph of a function, meaning its representation in the plane like the parabola for $f(x)=x^2$, because graph theory is not part of school education at the moment. This assumption turned out to be true: Graph theory is, if at all, only covered in dictionaries for academic purposes.

The category *purpose* is based on the self-portrayal of the dictionaries. Of course, there are a lot of reference books for mathematics, such as collections of formulas, which at present appear either in print or with increasing frequency in digital form (Schmidt-Thieme & Weigand 2015). But following the terms of Wiegand (1998) these are non-lexicographic resources. Therefore, they are left out of this overview. Private publishers are not named, companies are. Some resources also combine properties of a simple dictionary and a general learning tool.

Furthermore, this overview only contains dictionaries with a proper user interface. For example, PDF documents such as online word lists, are not part of this overview, as they cannot really count as electronic dictionaries. The considered dictionaries either offer monolingual definitions or a list of terms, but not both.

Another dictionary or rather an encyclopaedia not mentioned here is Wikipedia, as we only list works where the author was named. Wikipedia does not fulfil this criterion. As it is an open collaborative resource, it might be difficult to trust the information from an academic perspective.

4 Corpora

To compile our dictionary, we built two comparable corpora consisting of academic texts the students use during their studies of graph theory. Therefore, the selection was based on the bibliography used in the course as well as on a survey we carried out with mathematics students. In the survey we asked them which sources they use. The result was that most of them consult Wikipedia (Kruse & Giacomini 2019). However, in order to maintain the quality of the dictionary we only included scientific publications in our corpora.

The English corpus contains eight books and 26 scientific papers (about one million tokens) and the German corpus consists of the lecture notes as well as of (parts of) nine textbooks (about 700.000 tokens). Each corpus comprises about 30.000 word types.

One obstacle in the creation process of the corpus was to deal with mathematical formulas. Due to different source file formats it was not possible to use a single workflow. Therefore, one has to keep in mind that as a result of these differences, the number of tokens for the same formula may vary in different texts. Yet, this is of no concern as the focus of this project is on the terminology and not on the formulas or the corpus itself.

5 Lemma Selection and Semantic Categorization

Our process of corpus-based lemma selection consists of different steps. We first extract definition patterns from the corpus which are typical for the mathematics language (Pagel & Schubotz 2014). Each of these patterns expresses a certain semantic relation which can be used in the further development of the dictionary (Kruse & Giacomini 2019). This pattern-based approach will be combined with data produced by other term extraction tools (e.g. Rösiger et al. 2016). The merged results are assessed by three expert raters (inter-rater reliability to be computed). This will lead to the final lemma list.

The selected lemmas will be classified according to nine different categories. The microstructure for the entry of each lemma will depend on the category of the lemma. The categories are: PARTS OF GRAPHS, TYPES OF GRAPHS, PROPERTIES OF GRAPHS, ALGORITHMS, MAPPINGS, THEOREMS, PROBLEMS, ACTIVITIES and PERSONS.

The dictionary will cover nouns, verbs and adjectives. The latter two are each assigned a single category, according to their respective function. Adjectives are used to express PROPERTIES OF GRAPHS, typically in the form of adjective+noun combinations. In the entries, we will distinguish cases where objects always have a given property (indicated to the user by the key phrase *X is always ADJ*) from those where an object may or may not have a given property (*X can be ADJ*). There are rather few verbs with a terminological meaning in the domain of graph theory. They express ACTIVITIES (or states) and will be presented like in a valency dictionary, with an indication of the possible (categories of) subjects and complements. For example, the verb *inzidieren* ("be a neighbour of") has *Kante* ("arc") as a typical subject.

Nouns are classified by the categories TYPES, PARTS, ALGORITHMS, MAPPINGS, THEOREMS, PROBLEMS and PERSONS. Examples for TYPES OF GRAPHS are *tree* or *Petersen graph*. Our notion of TYPE OF GRAPH is based on the structure of the graphs (with/without circles, bridges, etc.). PARTS OF GRAPHS are lemmas such as *node*, *edge*, *path* – so all the objects of which a graph consists or rather all the terms being in a meronymic relationship with the term *graph* or with another lemma from the category TYPES OF GRAPHS.

The categories ALGORITHMS, MAPPINGS, THEOREMS and PROBLEMS should be self-explanatory. For example, they apply in cases that a theorem is given a proper name, such as the *Handshaking-Lemma*. Thus, not all theorems found in the corpus will have an entry in the dictionary.

PERSON NAMES are part of the dictionary, in case that a category, e.g. a THEOREM or a TYPE OF GRAPH, is named after a person. Probably, there will not be a lot of information on the persons available in the corpus. Therefore, we plan to link these entries to other databases dealing with mathematicians.

Name	Purpose	Form	Graph theory	Publisher	URL
Encyclopedia of Matheamtics	a	m EN	covered	Springer / European Mathematical Society	https://www.encyclopediaofmath.org/index.php/Main_Page
Encyclopedia of Triangle Centers	a	m EN	no, other focus	private	https://faculty.evansville.edu/ck6/encyclopedia/glossary.htm
epi Wörterbuch	?	b DE-EN	no	private / spirito GmBH	http://www.informatik.oelinger.de/dictionary/index.html
Illustrated Mathematics Dictionary	s	m EN	no	private	https://www.mathsisfun.com/definitions/index.htm
Lexikon der Mathematik	a	m DE	covered	Guido Walz / Springer Spektrum	https://www.spektrum.de/lexikon/mathematik/
Mathematik online Lexikon	a	m DE, m EN	covered	Universitäten Stuttgart und Ulm	https://mo.mathematik.uni-stuttgart.de/lexikon/
Mathematisches Lexikon	s, a	m DE	no	Universität Wien	https://www.mathe-online.at/mathint/lexikon
Mathematisches Wörterbuch / Math Dictionary	?	b DE-EN	no	private	https://www.henked.de/maple/woerterbuch.htm
Math Glossary, Math Terms	?	m EN	covered	private	https://www.cut-the-knot.org/glossary/atop.shtm
Math spoken here	?	m EN	no	private	http://www.mathnstuff.com/math/spoken/here/1words/words.htm
Mathworld Wolfram	a	m EN	covered	Wolfram Research	https://mathworld.wolfram.com/
SchulMatheLexikon	s	m DE	no	Vorhilfe.de e.V.	https://www.matheraum.de/wissen/SchulMatheLexikon
UniMatheLexikon	a	m DE	no	Vorhilfe.de e.V.	https://matheraum.de/wissen/UniMatheLexikon?mrsessionid=aa46eb31c21ae22eb45e2930f26a487c24689235

Table 1: Electronic mathematics dictionaries. In the purpose column, academic is abbreviated to a, school to s; ? means that the purpose is not given. The form is described as either monolingual (m) or bilingual (b); the particular languages are indicated.

6 Microstructure

Our intended microstructure consists of two main parts: definitions and relations. Before we present their role in our dictionary, we give an overview of different types of definitions considered in lexicography based on the work of Lew and Dziemianko (2006).

6.1 Definitions

Lew and Dziemianko (2006) discuss three types of definitions which are used in lexicography: single clause *when*-definitions, contextual definitions and analytic definitions. We go through them and see how far they apply for our case and with which advantages and disadvantages they come.

6.1.1 Analytic Definitions

First, we examine analytic definitions (or logical definitions). They are the most classical ones following the Aristotelian scheme. Mathematical definitions in textbooks are generally written in the following defining format, cf.:

A *graph G* is an ordered pair *(V(G), E(G))* consisting of a set *V(G)* of *vertices* and a set *E(G)*, disjoint from *V(G)*, of *edges*, together with an *incidence function* ψ_G that associates with each edge of *G* an unordered pair of (not necessarily distinct) vertices of *G*. (Bondy & Murty 2008: 2; emph. in original)

This definition provides the genus proximum of *graph*, namely *ordered pair*. This definition style is almost always used for nouns.

Adamska-Sałaciak (2012) had a closer look at this kind of definitions and describes some downsides coming with their usage. The first problem she investigates is circularity because the genus proximum might be defined itself at some other place in the dictionary and in the end becomes the definiens. This is especially a problem in general language lexicography because not every word has a clear definition, e.g. due to connotation, collocational meaning, etc. But in terminology, especially in mathematics, this is not very likely to happen, as mathematics typically relies on definitions of the objects it works with, and on logical relationships between such objects. So, in our case there is no need to worry about this issue from a lexicographer's perspective.

Secondly, Adamska-Sałaciak (2012) deals with obscurity which occurs when the words in the definitions are even harder or less common in texts than the lemma itself. That might also apply for our dictionary since the user may have to look up words used in the definition, but as they are a prerequisite to understand the subject from a cognitive perspective, this is a risk we absolutely have to take.

Similarly, a third issue addressed by Adamska-Sałaciak (2012) will not be very likely to happen in mathematics for most of the lemmas: gaps in hierarchy resulting in missing hypernyms. If we go back to the basic definitions of a mathematical domain, the words used are taken from the general language, as is the case above with *pair,* of which the user should have an intuitive understanding. In general, most of the mathematical definitions rely on set theory which is basically an idea of the cognitive concept of being inside or outside something.[1] Nevertheless, not all definitions can be based on the indication of hypernyms: Adamska-Sałaciak (2012) suggests to use hyponyms in these cases. We will come back to this proposal in Section 6.2.

Another point of criticism are the abbreviations used by lexicographers which might not be always understandable to the user. Most of them date back to printed dictionaries which had a notorious lack of space. As we create an electronic dictionary, space is not a problem and such abbreviations will not be used.

6.1.2 Single Clause when-Definitions

With single clause *when*-definitions and full sentence definitions (FSD) a new format was established which is closer to the general language than analytic definitions. The beginning of this development might not date back to Aristotle, but even 30 years ago the following was stated:

Lexicographic definitions have a curious tendency not to stick in the mind, whereas the immediacy, the accessibility and the vividness of folk definitions often make them more memorable and consequently more likely to be of help in both decoding and encoding. (Stock 1986: 86f.)

What Stock (1986) here refers to as folk definitions were the bases for the development of FSD and single clause *when*-definitions.

According to Dziemianko and Lew (2006) single clause *when*-definitions are mostly used for the definition of nouns. In mathematical texts however, definitions with the use of *when* do not really appear. It is more common to use *if*, as in "A *graph* is *simple* if it has no loops or parallel edges" (Bondy & Murty 2008: 3; emph. in original). This definition style is mostly used to define properties of mathematical objects, expressed by means of adjectives. Thus, the actual definiendum is often an adjective+noun combination that denotes a subtype of a mathematical object, e.g. a certain type of graph.

Similarly, definitions of this form also appear to define verbs, e.g. in "If *e* is an edge and *u* and *v* are vertices such that $\psi_G(e)=\{u,v\}$, then *e* is said to *join u* and *v*" (Bondy & Murty, 2008: 2; emph. in original). Dziemianko and Lew (2013) and Lew and Dziemianko (2012, 2006) did several experiments on the usage of single clause *when*-definitions in pedagogical dictionaries and concluded:

One way in which dictionary users confronted with a single-clause definition might recognize that the definition defines a noun would be through their familiarity with the convention of using this definition type to explain nouns. The question is, however, to what extent this actually is a convention: can we be sure, for example, that such definitions are never used to define adjectives or verbs? There is no evidence to tell us this. (Lew and Dziemianko, 2012)

[1] For a comprehensive account of that idea see Lakoff & Núñez (2000).

As pointed out before, in mathematics the single-clause definitions are indeed used for adjectives and verbs. Therefore, we can conclude that mathematical definitions are somehow different (Vanetik et al. 2020). In our dictionary we will use the definition scheme established in mathematics.

6.1.3 Full Sentence Definitions (FSD)

The third type of definition are the FSD, which came up in mid 1990s when the Cobuild dictionary was published. The research carried out along with it mainly focuses on the acquisition of a foreign language (e.g. Allen 1996; Bogaards 1996; Herbst 1996). Though this definition form was highly praised, it did not really find its way directly into more dictionaries. Rundell (2006) tries to explain this fact as he is actually in favour of them: "They provide a much fuller picture of the target lexical items, yet without making unreasonable demands on users or requiring them to know any special conventions" (Rundell 2006: 326). This statement fully applies to our case, as our user group is not familiar with linguistic terminology or definition styles; thus, FSD may be a reasonable option. Nevertheless, Rundell (2006) also gives three major disadvantages of FSD: length, overspecification, and new conventions for old.
As our dictionary will be (only) electronic, length is not as important as for a printed dictionary because a clearly arranged layout can be used without any loss of space. Nevertheless, sentence length and sentence complexity should be kept in mind. Therefore, we will use indicator phrases in the microstructure which paraphrase the semantic relations by using expressions of general language. They are presented in detail in the next section.
Rundell (2006) also mentions anaphora resolution but as the target group will be familiar with either German or English or at least the grammar of an Indo-European language this can be ignored. Further arguments of Rundell (2006) against FSD address the general language and are not really relevant for the case of LSP. Additionally, as the mathematical definitions always include a specific meaning, overspecification is not an issue.

6.2 Relations

Having all this in mind, we will now take a look at the second part of the microstructure, the relations between mathematical objects. As stated above, in mathematics, semantic and logical relations tend to be equivalent. In other domains it might be necessary to distinguish these two levels carefully.
In our dictionary we will use a kind of FSD when we explain the relations, since the intended user group of the dictionary is not familiar with linguistic terminology. We paraphrase the relations using expressions of general language: synonyms (*is also called*), hypernyms (*is always a*) / hyponyms (*examples are*), meronyms (*is part of*) / holonyms (*is composed of*), eponyms (*is named after*), pertonyms (*linguistically related*), mapped to (*is usually mapped to* / *is canonically mapped to*), alternatives, attributes (*possible properties*), analogies (*is analogous to*). Which relations apply for each category is shown in Table 2.
Most of these relations are known from lexical semantics and used in our dictionary in the standard way, but there are also some domain-specific ones: *mapped to*, *alternatives* and *analogous to*. *Mapped to* means a mapping in the mathematical sense. For example, to each *edge* a *weight* can be assigned. We differentiate between *usually* and *canonically mapped to*. A *canonical mapping* is one that occurs because it follows from the way how graphs (or other objects of the domain) are defined. For example, each edge *is canonically mapped to* two nodes because this is how graphs are defined. In contrast, weights are only *usually mapped to* edges because not every edge needs to have a weight. The mappings can be defined between GRAPHS or their PARTS.
There are two other relations, alternative and medium, which are both related to ALGORITHMS. An alternative can only exist for ALGORITHMS: there can be two ALGORITHMS to reach the same goal. For example, both, *Fleury's algorithm* and *Hierholzer's algorithm* can be used to compute an *Euler tour*. But the terms are not synonymous as the ALGORITHMS apply different techniques to reach their goal.
Usually, the texts of our corpus contain textual definitions following the established scheme of mathematics for lemmas from the categories PARTS OF GRAPHS, TYPES OF GRAPHS, PROPERTIES OF GRAPHS, MAPPINGS and ACTIONS. ALGORITHMS, PERSONS, PROBLEMS and THEOREMS will not be defined (see above). As the adjectives always have a noun they refer to, they will be given as a lemma together with this noun. This uniqueness applies within a certain field: In German, for example, *vollständiger Graph* and the proof technique *vollständige Induktion* would be regarded as two different lemmas.
The underlying ontology structure is implemented in the Web Ontology Language (OWL)[2] using the editor Protégé (Musen 2015). Thereby, the categories are used as classes and the relations are used as object properties. Therefore, we will use some of the terminology from OWL in the remainder of this section. For each relation we can indicate a possible source category (domain) and a target category (range). For example, if we have a look at the eponymic relation, indicated by *is named after* only PERSONS are a possible range, whereas members of all the other categories can serve as the domain.
In addition, we can differentiate between symmetric and non-symmetric relations. In our case, equivalents, synonyms, pertonyms, antonyms, analogies, alternatives and mappings are symmetric relations, the others are not. Symmetric relations can only be established between members of the same category.
Not for each lemma from each category all relations are relevant and thus described in the dictionary. The equivalents are always given, as they constitute an essential part of the dictionary. Synonyms are given wherever possible. Hyper- and hyponyms are given for the defined lemmas. For the others, ALGORITHMS, PROBLEMS and PERSONS, they do not

[2] OWL is a W3C-Standard. More information can be found on their web page https://www.w3.org/TR/owl2-overview/

really exist in a way which is relevant for the project: each member of the category would have the same hypernym, namely the name of the category. Of course, the category itself will be visible within the microstructure (see Figure 1).

	ALGORITHMS	MAPPINGS	PARTS	PERSONS	PROBLEMS	THEOREMS	TYPES	PROPERTIES	ACTIVITIES
isEquivalentOf	DR	DR	DR	DR	DR	DR	DR	DR	DR
isSynonymOf	DR	DR	DR		DR	DR	DR	DR	DR
isHypernymOf		DR	DR				DR	DR	
isHyponymOf		DR	DR				DR	DR	
isHolonymOf			DR				D		
isMeronymOf			DR				R		
isPertonymOf	DR	DR	DR	DR	DR	DR	DR	DR	DR
isAntonymOf			DR				DR	DR	DR
isMediumTo	D	R	R				R	R	
isAnalogueTo		DR	DR				DR	DR	
isAlternativeTo	DR								
isAttributeTo	R	R	R		R		R	D	
isMappedTo			DR						
isEponymOf	R	R	R	D	R	R	R	R	R

Table 2: Categories and Relations. It is indicated whether the particular category serves as domain *D* or range *R* for each relation. Underlined entries denote that the relation can only exist between members of the same category.

Another question is the order in which the relations should be presented in the dictionary. It might be useful to give the equivalent first or even visually marked, as the user often either wants an explanation or a translation. Next, it is useful to give synonyms as the users might recognize terms which they are already familiar with and therefore do not need any further explanations. The synonyms can be followed by the hypernyms in order for the user to learn that the concept looked up is a subtype or an example of another given concept. Mathematics language is structured in a strongly hierarchical way. Therefore, the given hypernym will always be the direct hypernym on the next higher level.
Further information can be arranged in blocks which the user can open on demand. One block contains holonyms /meronyms, pertonyms and antonyms as they are linguistically related with the term. The other block provides information on domain-specific relations as it contains terms related as mediums, analogies, alternatives, attributes and mappings. The equivalents, synonyms, pertonyms and eponyms may be there for all the terms independently from the category.

Euler-Tour - *Category: Parts*

Definition A closed trail in a graph which contains each edge exactly once.

German equivalent Eulertour
Synonyms Eulerian path, Eulerian trail
...is always a trail, path
Examples Eulerian circuit, Eulerian cycle
...is composed of edges, nodes
...is part of Eulerian graph
...can be computed with Hierholzer's algorithm, Fleury's algorithm
...is analogous to Hamiltonian path
...is named after Leonhard Euler
...is linguistically related to Eulerian graph, Euler

Figure 1: Showcase article. The items shown would be linked to the corresponding article.

7 Conclusion and Future Work

We have now shown a possible microstructure for an electronic LSP-dictionary for mathematics. This microstructure is based on a category system we developed for classifying our lemmas. Our next step is to implement this structure and to fill it mostly automatically. The category system can be applied to other mathematical domains as well.

8 References

Adamska-Sałaciak, A. (2012). Dictionary definitions: problems and solutions. In *Studia Linguistica Universitatis Iagellonicae Cracoviensis*, 129(41), pp. 323-339.

Allen, R. (1996). The big four: The year of the dictionaries. In *English Today*, 12(2), pp. 41-55.

Bogaards, P. (1996). Dictionaries for Learners of English. In *International Journal of Lexicography*, 9(4), pp. 277-320.

Bondy, A. & Murty, M. (2008). *Graph Theory*. London: Springer 2008.

Dziemianko, A. & Lew, R. (2006). When you are Explaining the Meaning of a Word: The Effect of Abstract Noun Definition Format on Syntactic Class Identification. In E. Corino, C. Marello, C. Onesti (eds.) *Proceedings of the 12th EURALEX International Congress*, Torino, 6-9 September 2006. Torino: Edizioni dell'Orso, pp. 857-863.

Dziemianko, A. & Lew, R. (2013). When-definitions revisited. In *International Journal of Lexicography*, 26(2), pp.154-175.

Eisenreich, G. (2008). Die Fachlexikographie der Mathematik: eine Übersicht. In L. Hoffman & H. Kalverkämper & H.Wiegand (eds.) *Fachsprachen*. Berlin / New York: de Gruyter, pp. 1959-1966.

Fuertes-Olivera, P. & Tarp, S. (2014). *Theory and Practice of Specialised Online Dictionaries. Lexicography versus Terminology.* Berlin / Boston: de Gruyter.

Gouws, R. (2010). The Monolingual Specialised Dictionary for Learners. In P. Fuertes-Olivera (ed.) *Specialised Dictionaries for Learners*. Berlin / Boston: de Gruyter, pp. 55-68.

Herbst, T. (1996). On the way to the perfect learners' dictionary: a first comparison of OALD5, LDOCE3, COBUILD2and CIDE. In *International Journal of Lexicography*, 9(4), pp. 321-357.

Kruse, T. & Giacomini, L. (2019). Planning a domain-specific electronic dictionary for the mathematical field of graph theory: definitional patterns and term variation. In I. Kosem, T. Zingano Kuhn, M. Correia, J. P. Ferreria, M.Jansen, I. Pereira, J. Kallas, M. Jakubíček, S. Krek, C. Tiberius (eds.) *Electronic lexicography in the 21stcentury. Proceedings of the eLex 2019 conference,* Sintra, 1-3 October 2019. Brno: Lexical Computing CZ, s.r.o., pp. 676-693.

Lakoff, G. & Núñez, R. (2000). *Where mathematics comes from: how the embodied mind brings mathematics into being.* New York: Basic Books.

Lew, R. & Dziemianko, A. (2006). A New Type of Folk-inspired Definition in English Monolingual Learners' Dictionaries and its Usefulness for Conveying Syntactic Information. In *International Journal of Lexicography*, 19(3), pp. 225-242.

Lew, R. & Dziemianko, A. (2012). Single-clause when-definitions: take three. In R. V. Fjeld, J. M. Torjusen (eds.) *Proceedings of the 15th EURALEX International Congress*, Oslo, 7-12 August 2012. Department of Linguistics and Scandinavian Studies, University of Oslo, pp. 997-1002.

Musen, M. A. The Protégé project: A look back and a look forward. *AI Matters*, 1(4), pp. 4-12.

Pagel, R. & Schubotz, M. (2014). Mathematical Language Processing Project. In M. England, J. H. Davenport, A. Kohlhase, M. Kohlhase, P. Libbrecht, W. Neuper, P. Quaresma, A. P. Sexton, P. Sojka, J. Urban, S. M. Watt (eds.) *Joint Proceedings of the MathUI, OpenMath and ThEdu Workshops and Work in Progress track at CICM,* Coimbra, 7-11 June 2014. Aachen: CEUR Workshop Proceedings.

Rösiger, I. & Bettinger, J. & Schäfer, J. & Dorna, M. & Heid, U. (2016). Acquisition of semantic relations between terms: how far can we get with standard NLP tools? In *Proceedings of the 5th International Workshop on Computational Terminology, Computerm 2016,* Osaka, 12 December 2016, pp. 41-51.

Rundell, M. (2006). More than one Way to Skin a Cat: Why Full-Sentence Definitions Have not Been Universally Adopted. In E. Corino, C. Marello, C. Onesti (eds.) *Proceedings of the 12th EURALEX International Congress*, Torino, 6-9 September 2006. Torino: Edizioni dell'Orso, pp. 323-337.

Schmidt-Thieme, B. & Weigand, H. (2015). Medien. In R. Bruder et al. (eds.) *Handbuch der Mathematikdidaktik.* Berlin/Heidelberg: Springer Spektrum, pp. 461-490.

Stock, P. (1986). The structure and function of definitions. In M. Snell-Horny (ed.) *ZüriLEX '86 Proceedings, Papers read at the EURALEX International Congress,* University of Zürich, 9-14 September 1986. Tübingen: Francke, pp. 81-89.

Tarp, S. (2008). *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography.* Tübingen: Max Niemeyer.

Tarp, S. (2011). Pedagogical Lexicography: Towards a New and Strict Typology Corresponding to the Present State-of-the-Art. In *Lexikos*, 21(1), pp. 217-231.

Vanetik, N. & Litvak, M. & Shevchuk, S. & Reznik, L. (2020). Automated Discovery of Mathematical Definitions in Text. In N. Calzolari, F. Béchet, P. Blache, K. Choukri, C. Cieri, T. Declerck, S. Goggi, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (eds.) P*roceedings of the 12th International Conference on Language Resources and Evaluation, LREC 2020*, Marseille, 11-16 May 2020, Paris: ELRA, pp. 2086-2094.

Wiegand, H. (1998). *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie.* Berlin/New York: de Gruyter.