Planning a Domain-specific Electronic Dictionary for the Mathematical Field of Graph Theory: Definitional Patterns and Term Variation

Theresa Kruse[1], Laura Giacomini[1,2]

[1] Institute for Information Science and Natural Language Processing (IwiSt), Universität Hildesheim, Universitätsplatz 1, D-31141 Hildesheim

[2] Institute for Translation and Interpreting (IÜD), University of Heidelberg, Plöck 57a, D-69117 Heidelberg

E-mail: theresa.kruse@uni-hildesheim.de, laura.giacomini@uni-hildesheim.de

Abstract

We plan to create an electronic dictionary for the mathematical field of graph theory. The dictionary should help students to improve their usage of the mathematical terminology. Besides the alphabetical access, the dictionary will also provide thematic, onomasiological access; it will contain lemmas in German and English, related terms and equivalence statements. Presently, such a dictionary does not exist. The dictionary basis is formed by two corpora composed of textbooks, scientific papers and lecture notes, containing all the texts the students use in their graph theory course in German and English. In the current pre-lexicographic stage, our focus is on relations between terms and on patterns used in the corpus to express them. We collect the definition patterns in the corpus and plan to use them for term extraction. Thereby, we can extract the semantic relations at the same time. In this paper we explore in particular the synonymy relations from an orthographical, morphological and syntactic perspective and draw conclusions for data acquisition. It might be possible to apply our extraction methods later for creating dictionaries in other mathematical domains.

Keywords: terminology, mathematical; patterns; relations; term variation

1. An electronic dictionary for graph theory: brief overview

We plan to create an electronic dictionary for the mathematical field of graph theory. The dictionary shall be bilingual, German and English. The purpose of the dictionary is to help mathematics students to improve their academic writing regarding terminology. We extract terms from the texts using definition patterns and aim to associate with each pattern a particular semantic relation which we will then use to automatically create components of an ontology, as a backbone of the electronic dictionary.

In this paper, we first give an overview of the historical and linguistic aspects of graph theory and mathematics, respectively. The first step is to show that the language of

graph theory is a language for special purposes (Section 2). Section 3 deals with the planned dictionary itself. There will be a closer look at the target group, the composition of the corpus and at the planned structure concerning distribution, micro- and macrostructure as well as user guidance. Section 4 presents definition patterns, their creation and the semantic relations. Additionally, we introduce the topic of domain specific variants and provide a first analysis of their usage.

2. Historical and linguistic aspects of graph theory

In the following, an overview of the lexicographic aspects of mathematics is given. A complete theory of the multimodal structure of mathematical texts is still missing. Mathematical language is regarded as a symbolic language, and all conclusions are inherent to the language (Atayan et al., 2015). Nevertheless, mathematical texts have a macrostructure[11] in the sense of Roelcke (2010). The macrostructure consists of text types like definitions, theorems and proofs which came with the formalization of the mathematical language at the beginning of the 20th century (Atayan et al., 2015). According to Atayan et al. (2015), the language of mathematics, science and technology constitutes a linguistic variety.

The reasons for a particular term to be well-established are often historical and depend on influential publications. According to Hischer (2010), mathematical terminology uses words from the general language. That is the case for graph theory as well; for example *tree*, *complete* and *edge* also have mathematical meanings.

Graph theory is very young compared to other mathematical fields. The first problem of graph theory was the problem of the seven bridges in Königsberg, where the aim was to find a path through the city whereby every bridge is crossed only once (cf. Figure 1).

Leonard Euler proved in 1735/36 that this is not possible (Euler, 2009 (1736)). He called the mathematical field of this problem *Geometria situs* (geometry of position). More than a hundred years later Sylvester (1878) proposed the term *graph* for these structures. That was the first time the term *graph* appeared in this context. A further overview of the introduction of important terms in graph theory is given by Mulder (1992).

[1] It should be noted that the macrostructure of a language for special purposes differs from the macrostructure of a dictionary.

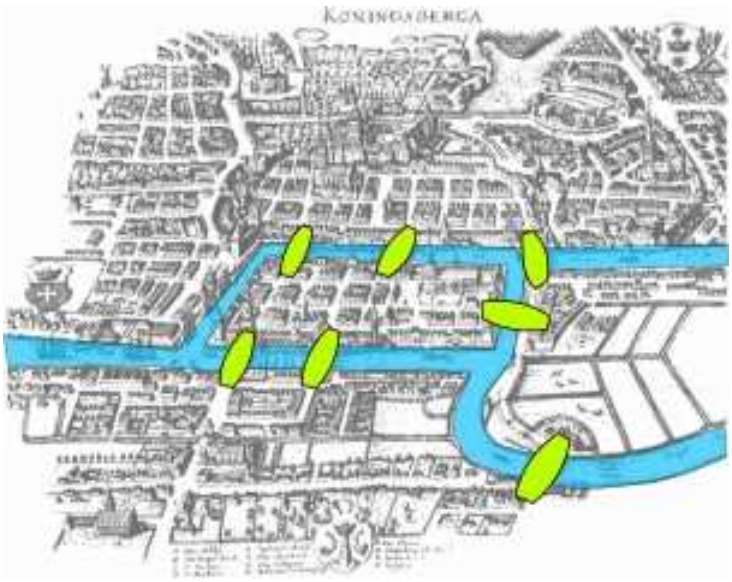

Figure 1: Map of Königsberg with the seven bridges to cross (Graphic: Bogdan Giuşcă).

In this paper, the language of graph theory is regarded as a language for special purposes (LSP) because typical characteristics of LSP can be identified (Roelcke, 2010). Some of our examples only apply to the German language, as Roelcke's work is mainly targeted at German, and as, for example, German and English compounding patterns differ notably on the surface. One of Roelcke's (2010) LSP criteria is richness in compounds. In our German corpus we find examples of compounds like *Kantenzug*, *wohlquasigeordnet* or *Kantenfärbung*. A derivative is *Wohlquasigeordnetheit*. Abbreviations are also very common in mathematical texts in general: *f* stands for *function*, or *G* stands for *graph*.

Another criterion for a LSP according to Roelcke (2010) is the preference for the third person. To check this for the texts on graph theory we did an investigation on the part of the corpus which is already machine processable.[2] In German we searched for *ich, du, man, er, sie, es, wir, ihr, Sie, Leser, Leserin*. In English for *I, you, one, he, she, it, we, they, reader*. The results are given in Table 1.

We excluded the cases from the table in which *ihr* is used as a possessive pronoun as well as those in which *er* or *sie* refer as a pronoun to things, e.g. to a graph. As a result, the relevant subject pronouns in German are *man*, *es* and *wir* which together represent about 95 percent of all pronoun occurrences. Unlike in Roelcke's hypothesis, it is the first person plural, not the third person which dominates.[3] This is a special feature of the LSP in mathematics. In English, the difference is even stronger, as one third is the use of *it* and two thirds concern *we*. The results are independent of the text

[2] 20,938 types and 482,604 tokens in German; 10,245 types and 378,629 tokens in English.

[3] This investigation will be repeated as soon as the complete corpus is available.

type in the corpus. Obviously, this small investigation can only give a first overview of the usage of the person. Further investigation is necessary, but not part of this dictionary project.

	hits			**hits**	
ich	40	0.32%	I	6	0.09 %
du	0	0%	you	49	0.76%
man	2,177	17.19%	one	29	0.45%
er	2	0.02%	he	27	0.42%
sie	0	0%	she	0	0%
es	3,363	26.55%	it	2,078	32.17%
wir	6,550	51.71%	we	4,249	65.78%
ihr	0	0%	they	0	0
Sie	464	3.66%			
Leser|in	70	0.55%	reader	21	0.33%
Sum	12,666	100%	Sum	6,459	100%

Table 1: Usage of personal pronouns

Nevertheless, we suppose the language of graph theory to be an LSP because we find examples for the other criteria, including recurrence and isotopy. We will make use of the latter in the creation of the pattern list.

3. Planning the dictionary

According to the model by Roelcke (2010), we want to consider the *intrafachlichen* and parts of the *interfachlichen Fachsprachwortschatz* (intra/inter domain specialized vocabulary) for the dictionary which means all the terms from the domain of graph theory and some terms from related mathematical domains.

3.1 Target group

The function theory of lexicography distinguishes between dictionaries for communicative, cognitive and interpretative situations (Fuertes-Olivera & Tarp, 2014; Tarp, 2008). The following description is based on the terminology in the taxonomy

presented by Bothma et al. (2017). They divide communicative situations into text reception and production, where the usage can be either automated or interactive.

The planned e-dictionary is primarily aimed at providing information interactively to the user in communicative as well as cognitive situations. On the one hand, users have to prepare presentations and texts in German on the basis of English texts, which is regarded as a communicative (text production) situation. The equivalents should help with that. On the other hand, the users do not simply have to translate the texts but also have to completely understand their content, which constitutes a cognitive situation. The communicative needs will be addressed by the provision of LSP equivalents. Here, the dictionary goes far beyond what can be found in general bilingual dictionaries: The latter would give both *komplett* and *vollständig* as equivalents of *complete*, while in graph theory the only acceptable and collocational equivalent is *vollständig*. The cognitive needs will be addressed by the inclusion of an ontology, such that the dictionary will support both semasiological and onomasiological access.

According to Roelcke (2010), there are some decisions to make. The target group are students, so they are semi-experts with a basic but no deeper knowledge of the subject. Furthermore, the dictionary will have a descriptive as well as a prescriptive function. The first step in dictionary creation is only descriptive, but some of our lemma selection criteria will include prescriptive elements. This is particularly true for the decisions related to variants, as we have to choose one main term for each variant. The main term should later be the main lemma. This is further investigated in Section 4.3.

3.2 The corpus

The dictionary is based on two corpora, one in English and one in German, composed of textbooks and scientific papers from the field of graph theory. Text sources are chosen in two steps due to different aspects. First, we chose all texts used in the bibliography for the lectures on graph theory at University of Hildesheim, because students attending these courses are the (first) target group of the dictionary. These texts are the lecture notes and (parts of) seven German books. The English subcorpus from this first step contains five books and 21 scientific papers.[4]

Secondly, we did a survey with 40 students asking them which sources they had been using for the preparation of their talks and asked to rate them according to their importance for the preparation. The importance could be rated on a scale from 1 (=very important) to 5 (=not important at all). The scores were the following: Internet 1.7, papers 1.74, other students 2.12, consultation-hour 2.39, books 2.93, lecture notes 3.04. The survey also had the aim to find out if further online resources needed to be included

[4] Due to the amount of texts we will not give exact source references for the examples.

in the corpus. The Internet was used by 92% of the students and ranked highest with regard to importance compared to other resources. Wikipedia was the most common online resource, with 55% for the English and 47.5% for the German version. Other sources like forums were not relevant for the corpus due to quantitative and qualitative factors.

After a qualitative analysis, we included the relevant texts. Books with a general introduction to mathematics or algebra with no focus on graph theory were excluded. So we added two German and four English books as well as four scientific papers. Relevant scientific papers in German do not exist in this field. In total, the German corpus comprises the script of the lecture, five books on graph theory and four books of which only the parts about graph theory are chosen. At the moment not all components are fully digitized and accessible.

Using the typology of Gläser (1990), we deal with monographs and scientific articles (including abstracts) for the domain internal communication. For the domain external communication we have textbooks for academic purposes. The lecture notes shall be regarded as somewhere in between. In the English corpus, there are nine books and 26 papers. Both corpora contain approximately 500,000 tokens each, which is a relatively small but still acceptable size for an LSP-corpus.

3.3 The structure of the dictionary

For creating the dictionary, we have to consider aspects of micro- and macrostructure in the planning process. Furthermore, we will have a look at the planned access structure.

3.3.1 Microstructure

The dictionary will have a hierarchical microstructure (Wiegand, 1989). As already mentioned, many of the mathematical terms are also part of general language, so that information on pronunciation or part of speech is not needed by the target group.

The focus will be on semantic aspects. Therefore, the articles will contain definitions, abbreviations, equivalents, collocations as well as information on semantically related terms like, for example, synonyms, antonyms or hyponyms – basically all the relations which will be examined in Section 4.2 below. Additionally, there can be usage examples extracted from the corpus. An etymological indication might be interesting but depends on whether there are valid data available for the majority of the terms.

The decision about which grammatical information shall be included depends on a further analysis of the material. For example, the users have German as their L1, and therefore there is no need to include the gender of the nouns as many of the nouns are

also used in the general language. Only in the case of irregularities might it be worth including gender indications. Similarly, there is no need to include further information on morphological inflection forms.

3.3.2 Macrostructure and access structure

We use the term macrostructure in the way presented by Wiegand and Gouws (2013) and Bergenholtz et al. (2008). We strive to achieve a fully developed macrostructure which means that all elements of the macrostructure will be linked (Nielsen, 1994).

The main part of creating the macrostructure is the lemma selection. The dictionary should contain nouns, adjectives, verbs and the corresponding multi-word terms; additionally pronouns or adverbs if they appear in patterns with the mentioned items. The terms will be from the field of graph theory, in both German and English with their equivalents.

Nouns indicate, for example, parts of graphs, special kinds of graphs or graph groups having specific names, but also problems, algorithms and theorems with a proper name and terms you can associate with graphs. Adjectives mainly indicate qualities of a graph or of its parts. Verbs denote things a graph or its parts can do or things one can do with a graph.

In addition, common phrases shall be included. It will be discussed where to draw a line with regard to other parts of the mathematical language, because graph theory also includes aspects of linear algebra. This decision will be made on the basis of corpus evidence.

According to the terminology discussed in Giacomini (2015), the dictionary shall have a search interface, an alphabetical index with a list of the alphabet characters as well as a list of alphabetically ordered terminological lemma signs and a systematic index. The latter might be based on the ontology, as the user can browse it with this index. For example you can choose 'qualities of a graph' and find the subcategories *vertex*, *edge* and *other*. Potentially, there will be included a tool in which one can insert a graph and the corresponding qualities and articles are shown.

The articles can be addressed either by semasiological or onomasiological access (Engelberg et al., 2016). For the first case, there will be a query form where after two or three letters a drop-down menu appears offering terms fitting the query. Thereby, the user might save some time during the search process. Speech recognition can be an option if appropriate software is available, but will not be a main focus. Furthermore, there will be the possibility of searching terms with an alphabetical index. Additionally, graphic elements can be included to show graphs and their corresponding lemmas.

4. Preparing the extraction of patterns, relations and variants

4.1 Finding definition patterns

We build on the methods used by Meyer (2001) and Barnbrook (2002). We identified typical patterns for definitions. They were found by looking closely at some of the texts, finding the patterns in the definitions in the first chapter, and using them as a random sample. In the next step, the detected patterns were applied to the corpus in order to verify if a pattern generalizes.

A further step was made by looking for all possible complements the patterns could have, and so resulting in the final patterns. The list is not fixed yet, but shall be extended during the project.

4.2 Semantic relations

Given the list of patterns, we tried to associate each pattern with a particular semantic relation. In some cases, the relations were ambiguous which resulted in an adjustment of the patterns. For example, we had the pattern *X is called Y* which was used for hypernyms, attributes and synonyms. A more detailed analysis allowed us to distinguish more refined patterns of *is called* as shown in Table 2. As Table 2 shows, it might be also possible to extract several relations from the same pattern as per (6), (7) and (8).

	Pattern	**Relation**
(1)	If-clause N1 is called N2	N1 hyp N2
(2)	N1 is called N2 If-clause	N1 hyp N2
(3)	N is called ADJ	ADJ attr N
(4)	N1 is called N2	N1 syn N2
(5)	N1 of N2 is called N3 If-clause	(N1 of N2) hyp N3
(6)	ADJ N1 is called N2	ADJ attr N1
(7)	ADJ N1 is called N2	ADJ N1 syn N2
(8)	ADJ N1 is called N2	N1 hyp N2

Table 2: Pattern *is called*. *hyp* stands for hyperonymy, *attr* for an attributive relation and *syn* for synonymy.

The chosen relations are based on GermaNet (Hamp & Feldweg, 1997; Heinrich & Hinrichs, 2010). Some adjustments were made as not all GermaNet relations are relevant for the domain of mathematics. At the same time some relations were added.

In GermaNet there are the following relations: synonymy, antonymy, hyperonymy / hyponymy, meronymy / holonymy, causation, association, pertonymy, participle and compound relations.

We use synonymy, antonymy, hyperonymy / hyponymy, meronymy / holonymy and pertonymy in the same way as GermaNet. Causation might be interesting, but most of the examples we found had a structure like *färben – gefärbt* which is a pertonymy relation.

For an association GermaNet gives the example *Schließvorrichtung – schließen*. We use the term association in a sense more typical for mathematics, in which it describes a kind of mapping, e.g. *weight – edge*. Compound relations might be investigated at a later point in time.

Furthermore, we use some new relations: an attributive relation between adjectives and nouns as not every noun term can be described by any attribute. For example a *Graph* can be *zusammenhängend* (engl. *connected*) but a *Kante* (engl. *edge*) cannot.

Additionally, with each algorithm or each mathematical process, we can associate its purpose: you use the *Hierholzer-Algorithmus* to find an *Eulertour*. We call the semantic relation between *Hierholzer-Algorithmus* and *Eulertour* 'purpose'. Eponyms shall also be indicated in the dictionary, cf. *Euler – Eulertour*.

Another domain-specific relation is given by alternatives, for example two different algorithms for the same purpose. Additionally, there are analogies, such as *Eckenfärbung* and *Kantenfärbung*. An open topic to investigate in this context are differences between German and English in the cases where the German language tends to use compounds which do not exist in a similar form in English. Therefore we not only have relations between single word terms, but between multi-word terms as well.

The last type of relation, e.g. combinations between verbs and nouns appearing together, cannot be found within patterns.

4.3 A closer look at variants

4.3.1 The notion of synonymous variation

In this contribution, we would also like to address the topic of synonymous variation as a phenomenon in mathematical terminology. Just like other LSP, the language of mathematics is not free from synonymy. As already seen in the previous section, synonymy is one of the semantic relations that can be identified in definitional patterns.

This study deals with homogeneous text genres. This means that synonymous variants of a term can be found in texts with comparable content and structural characteristics.

Hence, synonymous variation is not embedded in different systemic levels (like in the case of chronological or geographical variation), but rather in the same textual system. In order to adequately cover this kind of non-diasystemic synonymy, we apply the definition and the classification model proposed by Giacomini (2019) and originally developed for technical language. In this model, variation is defined as the presence, within a domain discourse, of one or more synonymous and morphologically similar terms. Synonymy is understood as a semantic function shared by words in the same or in similar contexts. The notion of functional synonymy also allows for the inclusion of near synonyms.

Despite our focus on non-diasystemic variation, we cannot exclude the presence of some register variants *a priori*. In our future work, we will be able to provide more details on this.

Lexicographic resources supporting text production should include variation inside a specific microstructural position, providing dictionary users with necessary information about variant types available for a certain term and their distribution in the reference corpus (e.g. source type, source name, author, etc.).

4.3.2 Variant location and distribution

In our comparable corpora, synonymous variants can be found in

- definitions (definitional patterns) and
- other textual components (e.g. titles, text body).

The former type of variant description is particularly relevant for its substantial contribution to the explicit and normative building of mathematical terminology. Among variants are both single-word terms and multi-word terms. We will now give some examples of definitional patterns in which the available variant pairs or chains are highlighted:

(a) A **closed path** is called a **cycle**

(b) A **connected forest** is called a **tree**

(c) A **maximal independent set** is called a **basis**

(d) Die **Elemente von V** nennen wir **Ecken** (oder **Knoten**; engl. **vertices**) **von G**, die **Elemente {u, v} in E** heißen **Kanten** (engl. **edges**) **von G**

(e) Die **Elemente von V** nennen wir **Ecken von D**, die **Elemente (u, v) in A** heißen **Bögen** (oder **gerichtete Kanten) von D**

(f) Im folgenden bezeichnen wir mit **K = K(G)** immer die **Anzahl der Komponenten eines Graphen G**

Besides variation at the level of contents related to graph theory, definitional patterns also reveal 'functional' variants, i.e. variants of terms which are employed to build the definition itself, e.g. *X bezeichnen wir mit Y* and *X nennen wir Y* in German, as well as *X is called Y* and *X heißt Y* in the English-German language comparison. In these patterns, Y indicates the definiendum, X the definiens.

We consider the definiens to be per se a variant of the definiendum, independently of its form, which can be

1. the combination of a genus proximum and differentia specifica like in *closed path* (cf. example (a)), with the hypernym path specified by *closed*, or *maximal independent set* (cf. example (c)), with the hypernym *set* subsequently specified by *independent* and by *maximal*;

2. a proper synonym or paraphrase like in *Elemente von V* (cf. example (d)) or *Anzahl der Komponenten eines Graphen* (cf. example (f)).

Definitional patterns may include more than one variant. Among variants, we also count English equivalents provided by some German sources (cf. example (d)). Variation within definitions is sometimes expressed in more complex ways, for instance through the inclusion of conditional restrictions for synonymy (cf. example (g)), or cross-referencing to other passages (cf. example (h)):

(g) Eine **Menge M ⊆ E von Kanten in einem Graphen G = (V,E)** heißt **Matching** (oder **Paarung**), **wenn keine zwei Kanten aus** *M* **einen gemeinsamen Knoten besitzen**

(h) Der in der Graphentheorie übliche Name für eine **Tabelle, die einen Graphen in der oben angegebenen Weise beschreibt**, ist **Adjazenzmatrix**

We also observe the presence of concatenated definitions in successive sentences, with a term first used as a definiendum and then as the definiens of a new term, for instance in:

(i) Das lässt sich leicht durch einen weiteren Begriff beschreiben: Ein **Graph, der als ebener Graph gezeichnet werden kann**, d.h. **zu einem ebenen Graphen isomorph ist**, heißt **plättbar** (oder **planar**). Ein **Würfel** ist also ein **plättbarer Graph** und wie wir oben gesehen haben ebenso alle anderen **Polyeder**

In example (i), the following complex variation structure can be identified in discourse:

- *Polyeder* is a hypernym of *Würfel*

- variants of *Polyeder* and *Würfel* are *plättbarer Graph*, *planarer Graph*, *Graph, der als ebener Graph gezeichnet werden kann* and *Graph, der zu einem ebenen Graph isomorph ist*.

This example also hints at a common feature of definitional texts: variants may be introduced for definitional purposes only (cf. *planar* as a synonymous variant for *plättbar*) without being further employed in the text. Tables 3, 4 and 5 display the corpus distribution of the synonymous variants collected so far, together with their absolute frequency.

Only a corpus-based diachronic study could provide relevant information for what concerns the origin of variation in the language of graph theory. Some cases, however, suggest the influence of the English language on German terminology, for instance for EN *adjacent* (which has a Latin origin) and DE *adjazent*, which coexists with the Germanic form *benachbart*, or EN *Chinese Postman Problem* and the loan translation DE *chinesisches Briefträgerproblem*, which coexists with some German adaptations such as *Problem des chinesischen Postboten*.

Motivation for the presence of one variant or another is also a complex aspect to handle, which would require a detailed analysis of textual structures and contents (cf. Freixa (2006) for a study on variation motivation).

4.3.3 Variant classification

In this study, we apply the classification devised by Giacomini (2017) and Giacomini (2019) for the technical language, with the following three variation types:

- orthographical variation (OV, mainly concerning changes in hyphenation and capitalisation),
- morphological variation (MV, concerning changes in lexical morphemes), and
- syntactic variation (SV, concerning changes in the order of compound elements, words, and syntagmatic structures).

According to this variation model, each pair *main term, variant* is analysed in terms of the combination of all three variation types, which can take the following values: OV / no OV; full MV / partial MV / no MV; SV / no SV. Among the criteria for determining which is the main term of a variant cluster, we choose frequency as the most suitable at the moment (for a discussion on the topic of main terms cf. Giacomini (2019)). We decided not to automatically choose a term introduced in a definition as the main term. This is due to the fact that distributions of variants in texts show that these terms are often not systematically employed in the argumentation following a definition.

Some of the previously listed variants will be classified in Table 3 in relation to the corresponding main term. The starting point are ten possible variation patterns resulting from the combination of the three variation types (cf. Table 3).

Information concerning the available variant patterns and the source in which they typically occur should be made available in the specialized dictionary to support users during text production.

Variation pattern			Language	Main term	Variant(s)
noOV	fullMV	SV	DE	TSP	Traveling Salesman Problem
			DE	Bogen	gerichtete Kante
OV	partMV	SV	DE	Dijkstra-Algorithmus	Dijkstras Kürzeste-Wege-Algorithmus
noOV	partMV	SV	EN	x and y are adjacent	y is a neighbour of x
OV	noMV	SV	DE	Eulerchar (S)	Euler-Charakteristik von S
noOV	noMV	SV	DE	Hamiltonkreis	Hamiltonscher Kreis
			DE	Eulertour	eulersche Tour
noOV	fullMV	noSV	DE	chordal	trianguliert
OV	partMV	noSV	EN	four colour theorem	four-color conjecture
noOV	partMV	noSV	EN	eulerian tour	Euler tour
			EN	plane graph	planar embedding
OV	noMV	noSV	DE	Eulerscher Kantenzug	eulerscher Kantenzug
			DE	Petersen-Graph	Petersen Graph
noOV	noMV	noSV	EN	Petersen graph	Petersen's graph

Table 3: Variant classification (OV: orthographical variation, MV: morphological variation, SV: syntactic variation).

4.3.4 Variant identification and extraction

Variants are either explicitly introduced in texts by means of formulations that usually put them in relation to a main term (this is mostly the case of definitions), or employed as alternatives to the main term.

As previously mentioned, variants can be also found in textual components other than definitions, for example in

(j) Zur geschickten Konstruktion von Eulertouren in Graphen, die diese Eigenschaften besitzen, gibt es zwei verschiedene Algorithmen, den **Zwiebelschalen-Algorithmus** (**Hierholzer-Algorithmus**) und Fleurys Algorithmus

At the present stage of the project, we cannot predict the level of heterogeneity of variation description in text bodies concerned with graph theory. Our assumption, however, is that heterogeneity poses particular problems for the automatic extraction of variants from a corpus.

So far, we have identified variants by manually analysing definitional patterns and by relying on our own specialized expertise. As soon as corpus pre-processing and annotation will be completed and textual data and structures analysed more closely, rule-based and statistical approaches will be applied to detect further synonymous variants in texts (cf. Giacomini, 2019) for the model of variant extraction from technical texts).

5. Conclusion and further work

We have proven that the language of graph theory is an LSP according to Roelcke, although there are some exceptions to his criteria definitions. Therefore we have the possibility of creating an electronic LSP dictionary. This process can be automated to a considerable degree, as there are pattern structures in the mathematical language which are used to express certain semantic relations. Another aspect we have to consider in the creation process of the dictionary are orthographical, morphological and syntactic variations. They can be extracted as well.

For our future work we have to come up with an approach that allows us to decide which variant should be regarded as the main term. For this decision, we will use linguistic and technical factors. In addition, it is still necessary to investigate how to guarantee that all patterns and all variants for a term are found.

6. References

Atayan, V., Metten, T. & Schmidt, V.A. (2015). Sprache in Mathematik, Naturwissenschaften und Technik. In *Handbuch Sprache und Wissen*. Berlin/Boston: De Gruyter, pp. 411–434.

Barnbrook, G. (2002). *Defining Language: A local grammar of definition sentences*. Amsterdam: John Benjamins.

Bergenholtz, H., Tarp, S. & Wiegand, H. E. (2008). Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In *Fachsprachen: Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft*. Berlin/Boston/New York: De Gruyter, pp. 1762–1832.

Bothma, T. J. D., Prinsloo, D. J. & Heid, U. (2017). A taxonomy of user guidance devices for e-lexicography. *Lexicographica*, 33, pp. 391–422.

Engelberg, S., Müller-Spitzer, C. & Schmidt, T. (2016). Vernetzungs- und Zugriffsstrukturen. In *Internetlexikografie. Ein Kompendium*. Berlin/Boston: De Gruyter, pp. 153– 195.

Euler, L. (2009 (1736)). Lösung eines Problems, das zum Bereich der Geometrie der Lage gehört (Solutio problematis ad geometriam situs pertinentis). In W. Velminksi (ed.) *Die Geburt der Graphentheorie: Ausgewählte Schriften von der Topologie zum Sudoku*. Berlin: Kulturverlag Kadmos, pp. 11–27.

Freixa, J. (2006). Causes of denominative variation in terminology. A typology proposal. *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*, 12(1), pp. 51–77.

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

Giacomini, L. (2015). Macrostructural properties and access structures of LSP edictionaries for translation: the technical domain. *Lexicographica*, 31, pp. 90–117.

Giacomini, L. (2017). An Ontology-terminology Model for Designing Technical edictionaries: Formalisation and Presentation of Variational Data. In *Proceedings of eLex*. Leiden, Netherlands, pp. 110–123. URL https://elex.link/elex2017/wp-content/uploads/2017/09/paper06.pdf.

Giacomini, L. (2019). Ontology - Frame - Terminology. A method for extracting and modelling variants of technical terms. Habilitationsschrift, forthcoming.

Gläser, R. (1990). *Fachtextsorten im Englischen*. Tübingen: Gunter Narr.

Hamp, B. & Feldweg, H. (1997). GermaNet - a Lexical-Semantic Net for German. In *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*. Madrid, Spain, pp. 9–15. URL https://www.aclweb.org/anthology/W97-0802.

Heinrich, V. & Hinrichs, E. (2010). GernEdiT - The GermaNet Editing Tool. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 2010)*. Valletta, Malta, pp. 2228–2235. URL http://www.lrec-conf.org/proceedings/lrec2010/pdf/264_Paper.pdf.

Hischer, H. (2010). *Was sind und was sollen Medien, Netze und Vernetzungen? - Vernetzung als Medium zur Weltaneignung*. Hildesheim/Berlin: Franzbecker.

Meyer, I. (2001). Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework. In *Recent Advances in Computational Terminology*. Amsterdam/Philadelphia: John Benjamins, pp. 279–302.

Mulder, H. M. (1992). Die Entstehung der Graphentheorie. In K. Wagner & R. Bodendiek (eds.) *Graphentheorie: Zahlen, Gruppen, Einbettungen von Graphen und Geschichte der Graphentheorie*. Mannheim/Leipzig/Wien/Zürich: Wissenschaftsverlag, pp. 296–313.

Nielsen, S. (1994). *The Bilingual LSP Dictionary - Principles and Practice for Legal Language*. Tübingen: Gunter Narr.

Roelcke, T. (2010). *Fachsprachen*. Berlin: Erich Schmidt.

Sylvester, J. J. (1878). Chemistry and Algebra. *Nature*, 17, p. 284.

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.

Wiegand, H. E. (1989). Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexicographie*. Berlin/New York: De Gruyter, pp. 462–501.

Wiegand, H.E. & Gouws, R.H. (2013). Macrostructures in printed dictionaries. In *Dictionaries: An International Encyclopedia of Lexicography*. Berlin/Boston/New York: De Gruyter, pp. 73–110.

Synonymous variants in German	**number of texts**	**number of hits**
adjazent	3	212
Benachbart	7	207
Bogen	5	226
Gerichtete Kante	4	39
Chinesisches Brieftrgerproblem	2	10
Brieftrgerproblem	1	2
Chinesisches-Postboten-Problem	1	1
Problem des chinesischen Postboten	1	1
Chinese Postman Problem	1	1
chordal	2	18
trianguliert	4	11
Dijkstra-Algorithmus	3	6
Dijkstras-Algorithmus	1	4
Algorithmus von Dijkstra	2	3
Dijkstras Krzeste-Wege-Algorithmus	1	1
Euler-Charakteristik von S	1	1
Eulerchar (S)	1	2

Eulerscher Kantenzug	2	4
Eulerweg	1	3
offener Euler-Zug	1	1
eulerscher Kantenzug	1	1
Eulertour	3	48
eulersche Tour	1	33
Eulersche Tour	1	8
Euler-Kreis	1	8
Eulerkreis	1	5
geschlossener Euler-Zug	1	1
Hamiltonkreis	3	127
hamiltonscher Kreis	1	17
Hamiltonscher Kreis	2	7
Traveling Salesman-Tour	1	1
Königsberger Brückenproblem	4	29
Brückenproblem	2	3
Matching	7	538
Paarung	3	90
Petersen-Graph	5	37
Petersen Graph	1	2
plättbar	2	51
planar	6	73
TSP	1	18
Serien-Parallel-Graph	1	5
sp-Graph	1	4
Traveling Salesman Problem	2	14
Traveling Salesman-Problem	1	7
Rundreiseproblem	1	5
Problem des Handlungsreisenden	1	1
Vierfarbenproblem	5	15
Vier-Farben-Problem	4	12
Vier-Farben-Satz	2	7
4-Farbenproblem	1	2
Zwiebelschalen-Algorithmus	1	6
Algorithmus von Hierholzer	1	2
Zwiebelschalenalgorithmus	1	1
Hierholzer-Algorithmus	1	1
Bestimmung einer Eulertour nach	1	1
Algorithmus nach Hierholzer	1	1

Table 4: Examples for corpus distribution of the synonymous variants in German.

Synonymous variants in English	**number of texts**	**number of hits**
arc	6	301
directed edge	5	9
Chinese remainder theorem	1	17
Chinese Remainder Theorem	2	5
Euler totient function	2	5
Euler's totient function	1	1
Euler's Phi function	1	1
eulerian tour	1	53
Euler tour	1	15
Euler circuit	1	1
four colour theorem	2	20
four colour problem	2	4
four-color conjecture	1	1
Hamilton cycle	2	71
Hamiltonian cycle	3	13
if and only if	18	510
iff	1	1
Petersen graph	3	152
Petersen‘s graph	1	3
plane graph	3	115
planar embedding	3	12
embedding in the plane	1	1
x and y are adjacent	14	440
y is a neighbour of x	3	87
neighbor	5	10
X ˜ y	1	1

Table 5: Examples for corpus distribution of the synonymous variants in English.