Originally published in: International journal on digital libraries (2018), First Online 17.11.2018
DOI: 10.1007/s00799-018-0254-x

An empirically validated, onomasiologically structured, and linguistically motivated online terminology

Re-designing scientific resources on German grammar

Karolina Suchowolec[1] · **Christian Lang**[2] · **Roman Schneider**[2]

Abstract
Terminological resources play a central role in the organization and retrieval of scientific texts. Both simple keyword lists and advanced modelings of relationships between terminological concepts can make a most valuable contribution to the analysis, classification, and finding of appropriate digital documents, either on the web or within local repositories. This seems especially true for long-established scientific fields with elusive theoretical and historical branches, where the use of terminology within documents from different origins is often far from being consistent. In this paper, we report on the progress of a linguistically motivated project on the onomasiological re-modeling of the terminological resources for the grammatical information system *grammis*. We present the design principles and the results of their application. In particular, we focus on new features for the authoring backend and discuss how these innovations help to evaluate existing, loosely structured terminological content, as well as to efficiently deal with automatic term extraction. Furthermore, we introduce a transformation to a future SKOS representation. We conclude with a positioning of our resources with regard to the Knowledge Organization discourse and discuss how a highly complex information environment like *grammis* benefits from the re-designed terminological KOS.

Keywords Grammatical information system · Grammatical terminology · Grammatical KOS · Concept system visualization · SKOS · Example-based querying

1 Introduction—the grammatical information system *grammis*

Web-based information systems, which do not impart research results in the form of singular text publications, but as interlinked online content, play a special role in the sustainable provision of research results to both the scientific community, and to the interested public. They exploit the entire potential of hypertextual media, in such a way that related content is distributed over a large number of interlinked, coherent modules. One of the most successful examples for German is the XML-based *grammis* information system, hosted by the Institute for the German Language (IDS) in Mannheim.[1] It brings together terminological, lexicographical, and bibliographic information about German grammar. Initiated more than two decades ago, *grammis* is based on a comprehensive scientific grammar [44], and 'combines traditional description of grammatical structures with the results of corpus-based studies [...]' [7, p. 622]. From a technical point of view, all primary and meta data are coded within more than one thousand semi-structured XML instances. Since semantical markup element types (*title*, *header*, *example*, *link anchor*, etc.) are used, the machine-aided access and processing of this information are straightforward.

From the user perspective *grammis* consists of varying content levels. The core component uses comprehensive sci-

✉ Karolina Suchowolec
karolina.suchowolec@th-koeln.de

Christian Lang
lang@ids-mannheim.de

Roman Schneider
schneider@ids-mannheim.de

[1] Technische Hochschule Köln, Cologne, Germany

[2] Institut für Deutsche Sprache (IDS), Mannheim, Germany

[1] See [36]; the information system can be found online at http://www.ids-mannheim.de/grammis/.

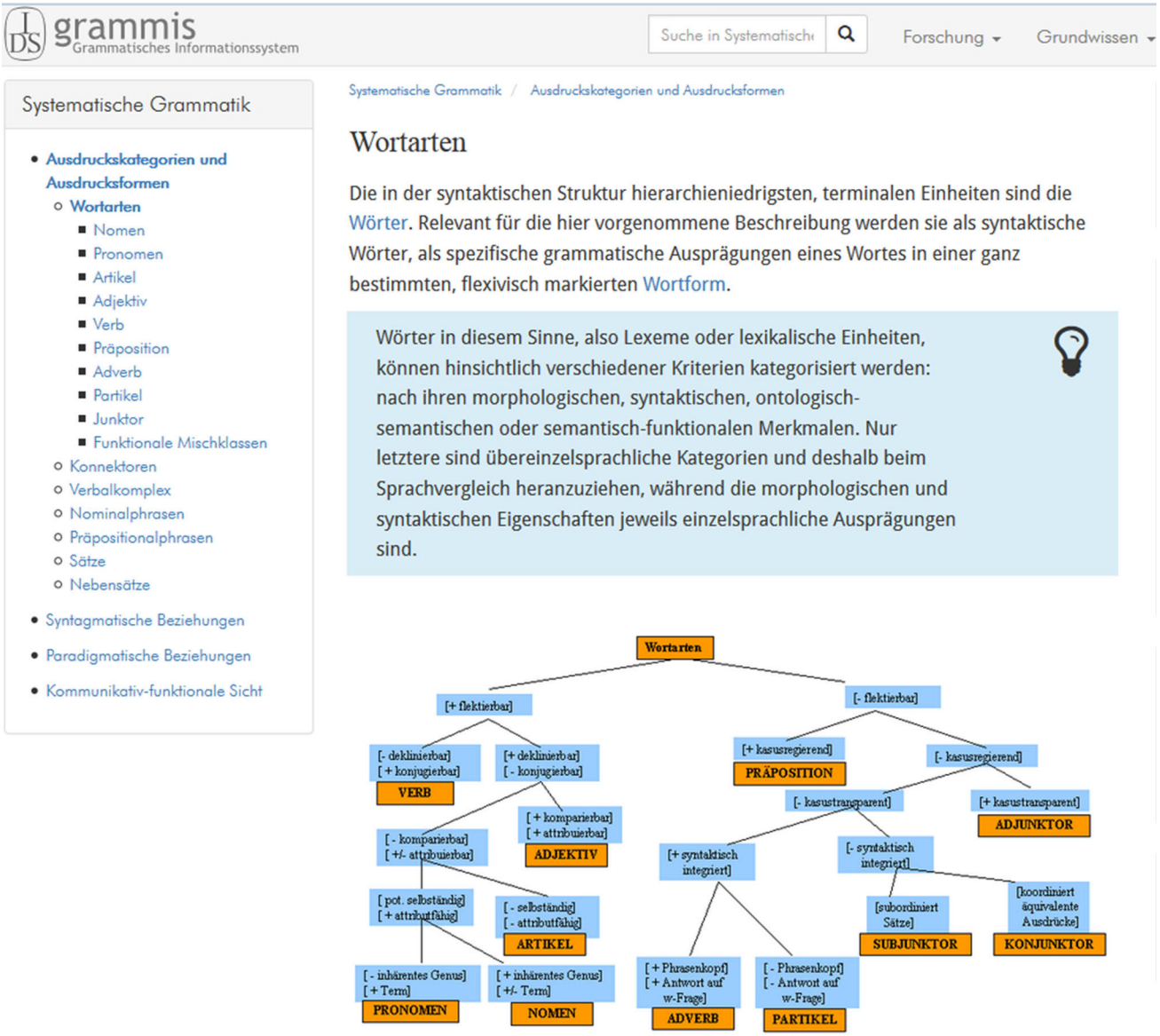

Fig. 1 Main Component of the Information System *grammis*

entific reference texts, which describe complex grammatical phenomena in great detail, and is aimed at a specialized public (in particular, academic linguists, see Fig. 1). Other modules explain cases of doubt for lay people appealing to a wider audience by avoiding specialist jargon. Additionally, the terminological component is meant to be a short reference (see Fig. 2), where each entry concisely describes a grammar concept and, therefore, gives just its basic notion. It also (statically) points the user to relevant entries in the main components for further reference.

Regardless, the intended user of *grammis* needs a sound knowledge in the field of grammar in order to find the desired content or explanations. This means that he or she should preferably be a professional linguist, most likely a grammar scholar, or at least have some background in linguistics.

To complete the big picture: the *grammis* information system is complemented by other online resources on German grammar that target at different user groups. For instance, the *ProGr@mm* spin-off is dedicated to teaching and explaining basic grammatical knowledge to non-scholars, such as students of linguistics.[2] It largely mirrors the structure of *grammis*, i.e., has similar components such as comprehensive references texts and terminological reference. These components, however, may or may not have the same textual content as corresponding *grammis* entries. Also, there are other, more specialized, scientific resources, for example on so-called *connectors*, which are results of different research projects at the IDS.

[2] http://www.ids-mannheim.de/progr@mm/.

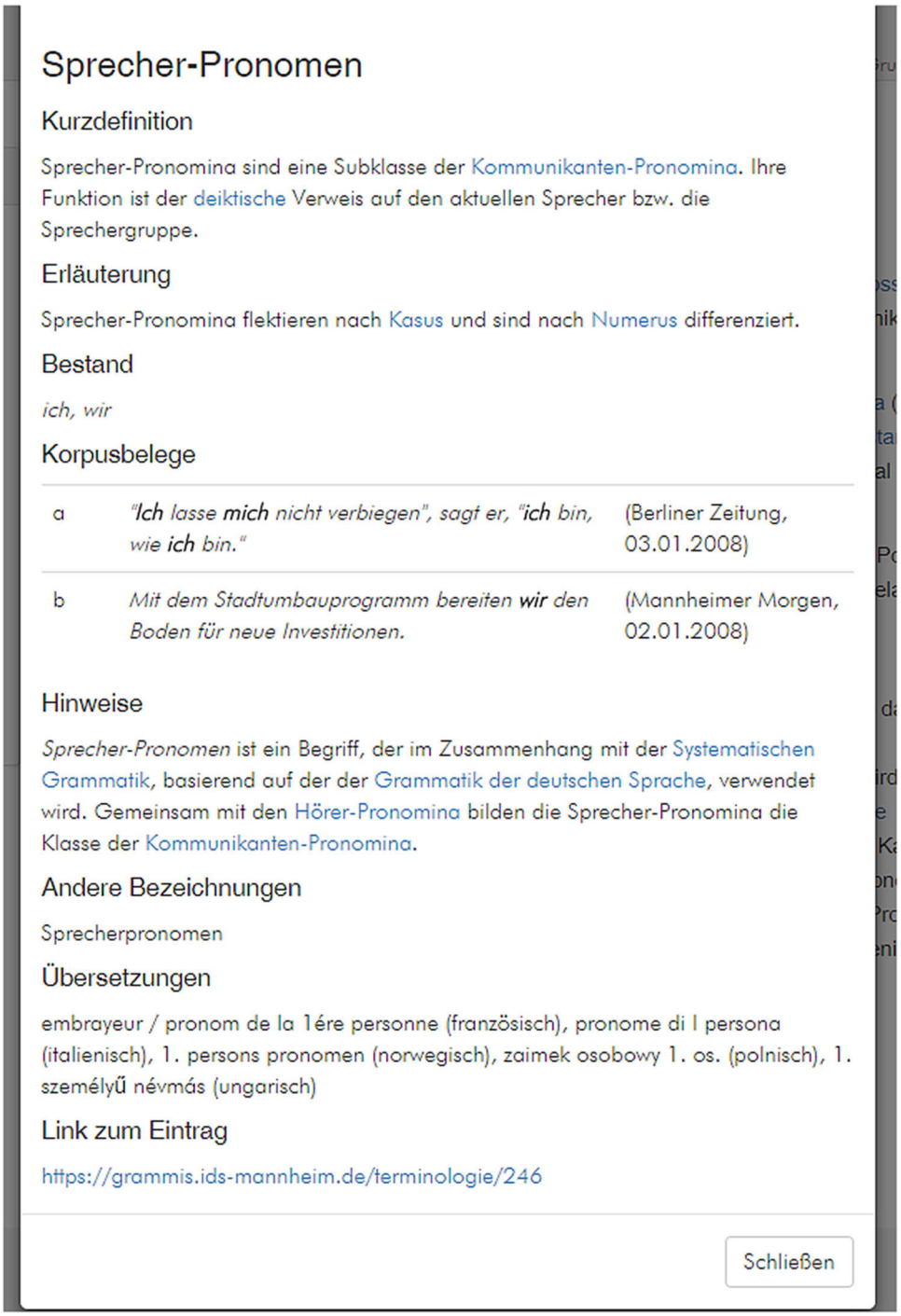

Fig. 2 Terminological Short Reference as a Modal Window, the writing of the description text is in progress

Finally, there is a separate terminological resource called *Grammatical Ontology* [37]. It is, in fact, a thesaurus of standardized hierarchical relations, such as broader/ narrower term (both generic and partitive) (BT/ NT), related term (RT) as well as synonym relation (cf. [23]). No textual descriptions of terms and concepts such as definitions are given. In addition, there is another conceptual relation called *concept ring* to link concepts that semantically overlap yet belong to different theoretical frameworks. Because all hierarchical and non-hierarchical relations of a concept within a *concept ring* apply to all other members of the *concept ring*, this data model is prone to misrepresenting grammatical theories.

This resource is implemented in an object-relational database management system (ORDBMS) and serves different purposes—it (dynamically) generates a list of references to the widespread terminological, lexicographical, and bibliographic resources by the IDS on a given topic, and it uses the

hierarchy in order to generate more relevant hits for full-text searches.

To sum up, the landscape of online grammar resources at the Institute for the German Language is heterogeneous. Above all, it is a result born out of different research projects, with different goals, scopes, scholarly traditions, and authors involved. Therefore, it covers different areas of grammar with different degrees of specialization. In such a heterogeneous landscape, a sound terminology system is indispensable in order to organize, manage, and access information. To this end, we see great potential in optimizing the terminological resources of *grammis*.

Terminology is, so far, managed within unconnected tools, depending on whether it is used in the *grammis* dictionaries, thesaurus, or bibliography. Terminology and register often reflect the needs of heterogeneous user groups. Additionally, different resources were designed with different functional purposes in mind—to serve as a concise specialist reference, or as a repository for enhancing information retrieval.

Finally, there is a broad spectrum of terminology, used within the hypertextual base, that is yet to be covered by terminology resources. To deal with these heterogeneities, distributions, and unsatisfactory coverage, the current terminology management needed to be re-designed.

2 Re-design principles

Following [11], we consider terminology management as the general process of how terminological inventory is set up, structured, maintained, and made available. In general, it comprises different aspects such as methodical, technical, but also economic and personal ones. In order to address the above-mentioned issues of heterogeneity, distribution, and coverage of the current terminology resources, we proposed the following work packages for the re-design of the terminology management, which we describe in more detail below.

- Combining distributed terminology resources into a single resource;
- Updating and expanding the content using automatic term extraction (ATE), i.e., the automatized identification and extraction of terms from domain-specific corpora;
- Ensuring interoperability of the new resource with other projects;
- Implementing a new backend for terminology management.

Hence, our re-design process primarily concerns the methodical and technical aspects of terminology management. Once completed, we will proceed to revising the terminological content.

2.1 Combining resources into a single resource

One way of improving the current terminology management for both, the user and the (terminology) author, is to unite the scattered terminology resources into one global resource. The new resource should first incorporate not only the descriptions of single concepts and terms, but also the relations between them. In other words, it should contain both—the hierarchy and the descriptive information about terms and concepts within this hierarchy.

Moreover, we decided to further unify this resource to globally serve different target groups and results from different research projects. These considerations have implications not only for the content of the entries, but also for their structure and, consequently the data model. Therefore, we evaluated the style and the structure of the existing terminology entries in order to find a common and hence more consistent way of authoring them.

2.2 Updating and expanding the content using ATE

To enhance the coverage of the new resource, we use ATE on the core component's entries.

2.3 Interoperability

The end-user of the new resource will be provided, as with the current resource, with an online interface. However, we also want to ensure that our data are transparent as well as easily accessible, exchangeable, and reusable within different (scientific) contexts and applications. In particular, we want to make it available to the communities that provide the scientific backbone of our project, i.e., the terminology community and the Knowledge Organization (KO) community. Therefore, we implement standard exchange formats for our data.

For the terminology community, such standard is TBX [24]. A resulting TBX representation will be more flat than our meta model (cf. [35]), because we need to reduce the number of available relation types to fit into the three-level TBX meta model.

For the KO community, we identified SKOS as a possible format.[3] We discuss the future SKOS implementation in Sect. 3.4. We prefer SKOS to other RDF(S) vocabularies for terminological resources such as Lemon,[4] because the scope of our project does not extend to adding linguistic expressiveness to the knowledge representation.

[3] https://www.w3.org/TR/2009/REC-skos-reference-20090818/.

[4] https://www.w3.org/community/ontolex/wiki/Final_Model_Specification.

2.4 New backend for terminology management

It follows then that the current terminology backend tools need to be reconsidered in order to account for the changes in the terminology management. Most importantly, the new tool should manage both the hierarchy and the descriptive information of concepts and terms within this hierarchy. Since no standardization for grammatical terminology is intended, it also needs to efficiently manage quasi-synonyms, i.e., partially equivalent terms, accounting for different schools of linguistics. Further features comprise the support of the above-mentioned exchange formats and the interoperability with other in-house applications. Finally, visualizing data as a graph is a desirable feature, as it facilitates the concept-based terminology management (see Sect. 3.2.1) (cf. [10]).

3 Re-design results

As of now, we conceptually re-designed the terminology management according to the principles specified above. Recently, we have proceeded to the revision of the terminological content. In the following sections, we summarize the results of the re-design process along the defined work packages: Backend, Visualization Tool, ATE and SKOS representation.[5]

3.1 Merged resources

We combined our scattered terminology resources into one powerful, state-of-the-art onomasiologically structured resource. It allows us to manage the hierarchy and the descriptive information as well. In the following, we refer to the former as macrostructure and to the latter as microstructure.[6] In the macrostructure, we keep the above-mentioned hierarchical and non-hierarchical as well as the *concept ring* relations. However, we are more cautious in using *concept rings*; instead, we address the problem of overlapping concepts by making use of theory tags more extensively. Parallel modeling of different grammatical theories becomes the key feature of the merged resource (cf. [41]).

3.2 Backend

Our decision to unify the terminological resources into one resource has necessitated a new backend. After an evaluation of different commercial and non-commercial tools according to the requirements mentioned in Sect. 2.4, we decided to redesign our own tools for the sake of the in-house integration (cf. [42]).

The main component of the new backend allows us to manage both the micro- and the macrostructure. The microstructure is encoded within an XML template, which contains the following semantic markups: *definition*, *explanation*, *example sentences*, *corpus examples*, *inventory*, and *note*. Concept descriptions can be uploaded and previewed in the backend tool. As for the macrostructure, the new backend allows us to manage the synonyms and the immediate relations of a given concept, i.e., exactly one level up or down in the hierarchical and associative structures. This macrostructure is displayed in a table. However, this type of display has its limits due to our methodical and practical approach. We follow the approaches in terminology management that take concept structure as a starting point and revise the terminological entries not alphabetically, but by considering a concept in its broader conceptual environment (cf. [40]). Thus, it is beneficial for the authors to access more than the immediate relational level of a concept.

We therefore looked into other options to give our authors access to a wider segment of concept relations and decided to develop a tool for data visualization. The idea of visualizing concept systems is a cornerstone of our traditional concept-oriented approach to terminology (cf. [10]).[7]

3.2.1 Visualization tool

The visualization tool—which is shown in Fig. 3—complements the main backend by generating an interactive graphical display of the macrostructure. It has grown organically alongside the authors' everyday work, i.e., development has been guided by the authors' needs and our general experience in terminology management. The tool is implemented in the programming language *R* [32] using i.a. the packages *shiny* [8] and *visNetwork* [2] and is accessible to authors on the IDS intranet. Authors are able to explore a concept in its structural environment via an intuitive graphical user interface (GUI). The GUI consists of five main modules (labeled A, B, C, D, and E in Fig. 3).

Upon entering a query term in the search field (A), the tool retrieves the associated concept(s) and their conceptual environment from the resource database. In case of ambiguity (i.e., if the query term is associated with multiple concepts), a suggestion list of concepts is displayed and the user picks the desired one. The retrieving options are highly customiz-

[5] Some of the results we already discussed in other publications in greater detail; we make reference to those in the corresponding sections.

[6] This distinction is inspired by lexicography where macrostructure is defined as the usually alphabetical order of entries and microstructure as the actual structure of one entry of a lexicon (cf. [18, p. 372]). Note that our macrostructure is ordered by concept relations instead of alphabetically.

[7] Note, however, that the application of visualization techniques is not limited to the traditional approach as is illustrated by EcoLexicon [28], a resource dealing with environmental terminology, based in the paradigm of frame-based terminology (cf. [14]).

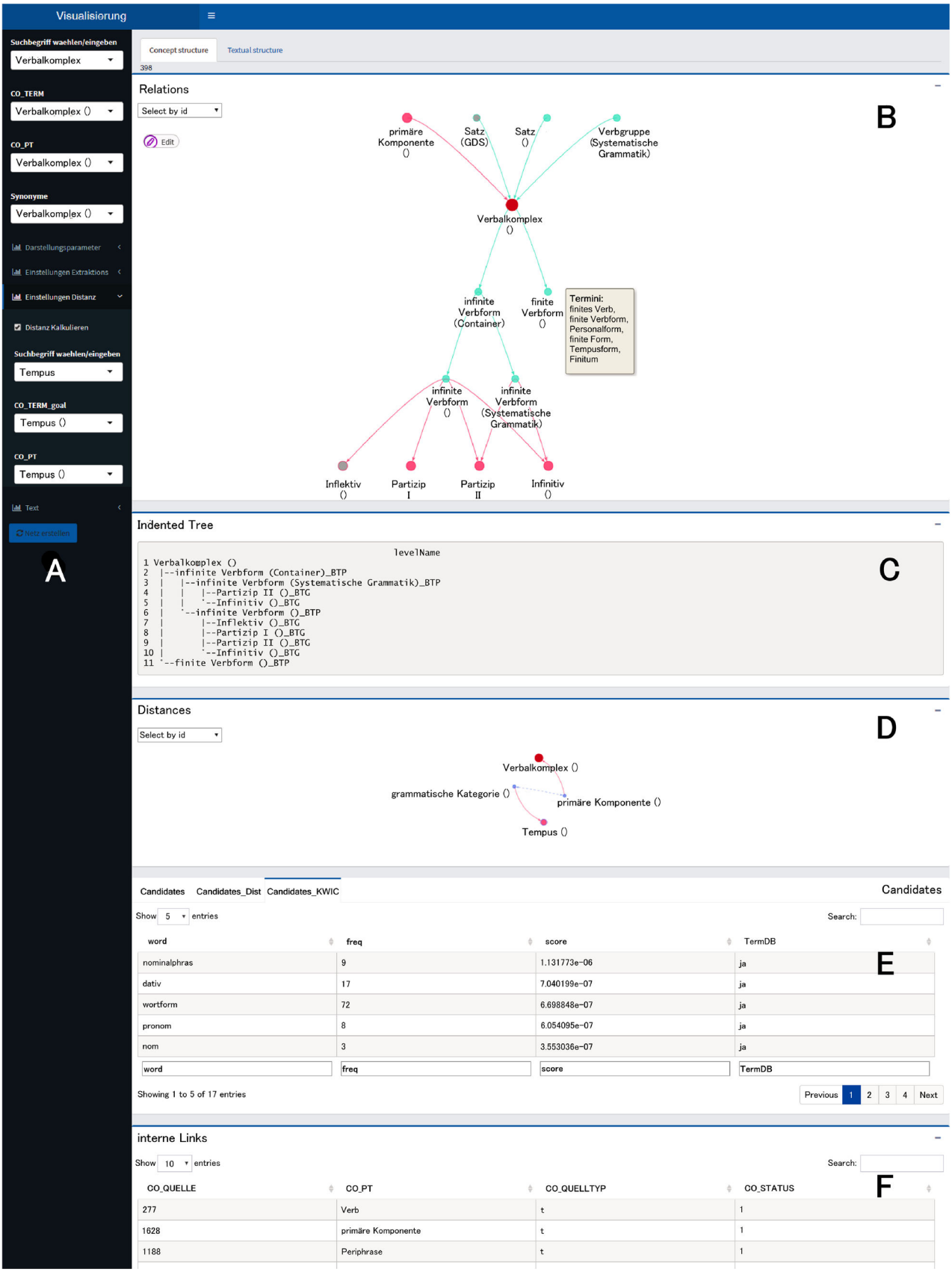

Fig. 3 Visualization Tool, showing four levels of hierarchical concept structure (RT are not included) of the concept *Verbalkomplex* (*verbal complex*) as graph in B, as indented tree in C. The graph has the display mode of a vertical tree

able with respect to the type, the number, and the display mode (hierarchical tree or non-hierarchical network)[8] of the conceptual relations. Subsequently, both a graph object and a data structure for an indented tree are created according to the user's parameter settings.

The graph is displayed in area B. Based on their relation type, both nodes and edges are color-coded (red: BT/ NT generic, green: BT/ NT partitive, blue: RT). The edges of the graph are arrow-shaped coding the direction of a relation (RT are always represented by bi-directional arrows), and its nodes are labeled with the preferred term of the concept it represents. However, as there is a high degree of synonymy in the linguistic domain, in most cases, more than one term is associated with a concept. In order to keep the graphs' readability and to provide the user with a comprehensive overview on a concept's synonyms beyond the preferred term, a single click on a node displays all the terms associated with a particular concept (as can be seen from the box next to the node labeled *finite Verbform*, '*finite verb*', in Fig. 3). By double-clicking on a node, the user triggers a new query such that a new graph is created and displayed. This way, the user can interactively navigate through the conceptual structure. Furthermore, the user can zoom in and out and rearrange the nodes with the mouse to ensure readability of the graph in case node labels overlap. Finally, nodes and edges can be added to the displayed graph. The latter feature has a mere mockup function and helps the user to visualize the effect of planned changes. These changes, however, do not update the database.

The indented tree is shown in area C.[9] The type of relation is coded by a suffix (BTP or BTG). Note that only the concept's hierarchical relations are included in the visualization. Moreover, due to the polyhierarchical nature of our concept structure, the query term's parent concepts are not considered.

In addition, the tool offers three features that go beyond mere data visualization. The first was implemented to support authors in expanding the system's content; based on the visualized concept, the tool proposes relevant term candidates from the ATE (see 3.3) in E. Authors can choose between three methods of candidate proposal; the first method proposes candidates that contain the entire string of the query term. This method helps to find candidates that are compounds and—due to characteristics of German word formation—potential narrower terms of the visualized concept.[10] The second method computes the Jaro–Winkler string distance [43] between the query term and candidates in the list. The system presents all relevant candidates within a string distance below a certain threshold (customizable by the user in A). This method is particularly useful for finding typos or spelling variants in the reference texts. The third method suggests candidates based on a Key Word in Context analysis (KWIC) and presents relevant candidates that are particularly frequent in a 10-word window before and after the query term.

Furthermore, the user can check the shortest path between the current concept and any other concept in the resource. The path is visualized in D; this way the user can check his expertise-based expectations on the conceptual structure against the actual structure in the database. In other words, if the graph in B or the indented tree in C is missing a concept the user expects to be there, they can find out how this concept is connected to the concept in question.

Finally, the app provides the authors with additional information on microstructure level. In section F, all hyperlinks linking into and out of the reference text associated with the visualized concept are displayed (including the respective sources or goals of the links).

Figure 4 shows a use case that illustrates the benefits of considering a concept within its broader conceptual environment in general and the application of visualization techniques in particular.

Looking at the graph in Fig. 4—even without entering a linguistic discussion of the concepts depicted—it is immediately apparent that there are inconsistencies in the way the concept structure is modeled. For instance, the integration of the concept *Satzadverbiale* ('*sentence adverbial*') into the concept structure is invalid, as it is simultaneously modeled as a hyponym and cohyponym of *Supplement* ('*supplement*'). Prerequisite for spotting this kind of inconsistencies is the inclusion of multiple hierarchy levels in a graph.

Despite the theoretical—and practical—importance of concept structure visualizations in terminology management and their practical benefits in everyday terminology work, there is still a lack of accessible tools for data visualization; Drewer et al. [12, p. 21] point out that, in particular, tools that enable a dynamical generation of visualizations from a database are still scarce, even though there has been a recent emergence of a new generation of software and tools

[8] Note that both modes—hierarchical and non-hierarchical—are available to all types of relations, i.e., what is often known as hierarchical relations BT/NT can also be displayed as non-hierarchical network in our tool. The inclusion of non-hierarchical relations (RT) in a hierarchical tree, however, renders the resulting graph overly complex and hardly interpretable.

[9] This feature was recently added as a result of a usability study by Fu et al. [17]. They find that depending on the area of application it can be more beneficial to use indented trees or node-link diagrams (graphs) as visualization technique; hence, Fu et al. conclude that in designing a visualization tool, multiple visualization techniques should be combined. Furthermore, they indicate the importance of customization to enable the user to adapt the visualization to their needs.

[10] If the current concept is labeled *Supplement* ('*supplement*'), this method will propose *Adverbialsupplement* ('*adverbial supplement*') as a candidate because the former is a substring of the latter.

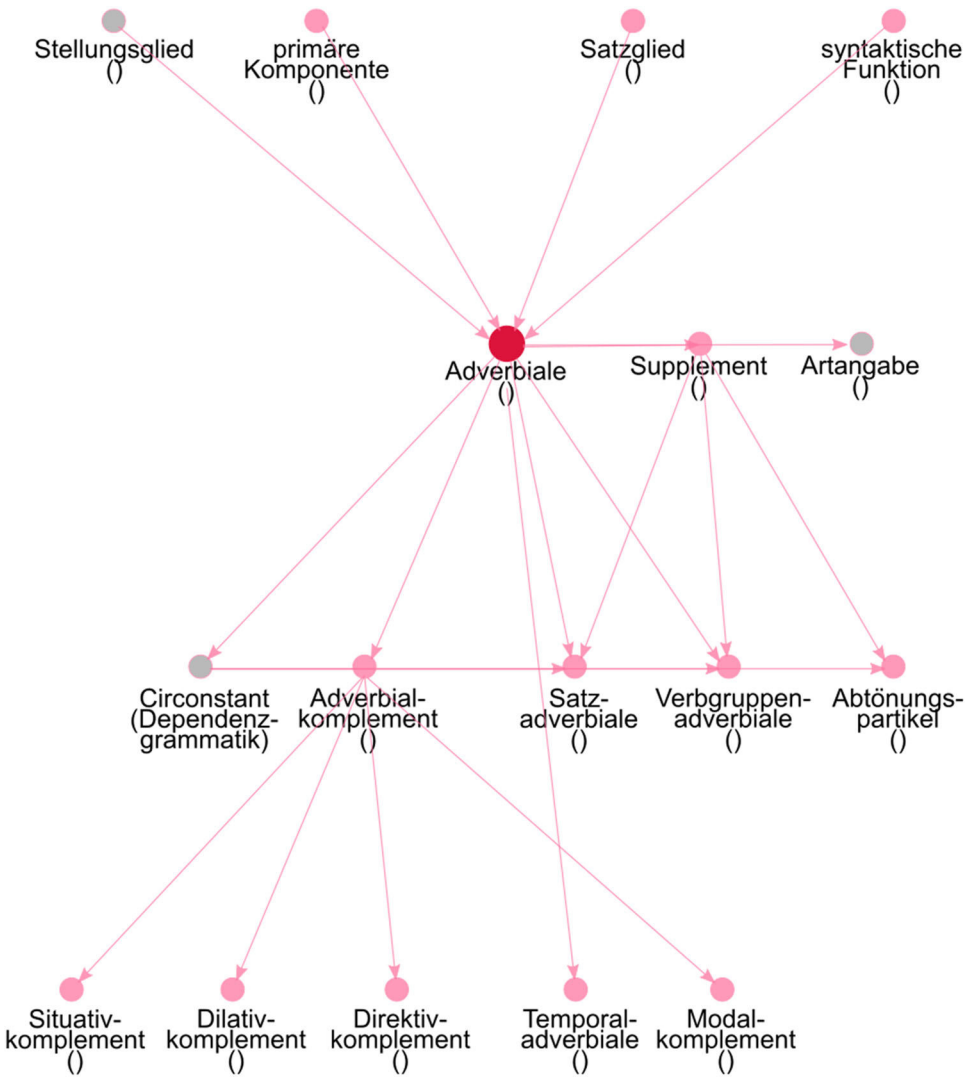

Fig. 4 Inconsistencies in the Hierarchical Concept Structure, *Adverbiale* (‘*adverbial*’)

(i.a. Coreon,[11] LookUp,[12] Quickterm,[13] and Termweb,[14] see [16]); we tested some of these tools during the evaluation mentioned above. Our tool meets the requirement of dynamically generating visualizations from the database; moreover, it offers a high degree of adaptability to the user’s needs and implements multiple visualization techniques (hierarchical or non-hierarchical graph and indented tree).

Currently, the visualization tool is part of our system’s backend only; we plan, however, to include a similar kind of visualization in our frontend in the future.

[11] https://www.coreon.com/.

[12] http://www.dog-gmbh.de/de/produkte/lookup/.

[13] https://www.kaleidoscope.at/de/terminologie/quickterm.

[14] http://www.interverbumtech.de/.

3.3 Updating the content using automatic term extraction (ATE)

As outlined in Fig. 1, we observed that the reference texts of both *grammis* and *ProGr@mm* contain terminology not covered by the existing terminological resources. To remedy this and to enhance the new resource’s coverage, we applied automatic terminology extraction (ATE), i.e., the automatized identification and extraction of terms from domain-specific corpora. In our case, the domain-specific corpus consisted of *grammis’* and *ProGr@mm’s* comprehensive reference texts; the size of the target corpus amounts to 2491 documents (all in the form of XML-files) with a total of 1.2 million tokens and 44,000 types.

Figure 5 shows the key components of our approach to ATE (see [27] for a detailed description of our method and the results). After subjecting our corpus to the standard linguistic preprocessing procedures (assigning Part-of-speech (POS) tags with TreeTagger [34] and stemming the words),

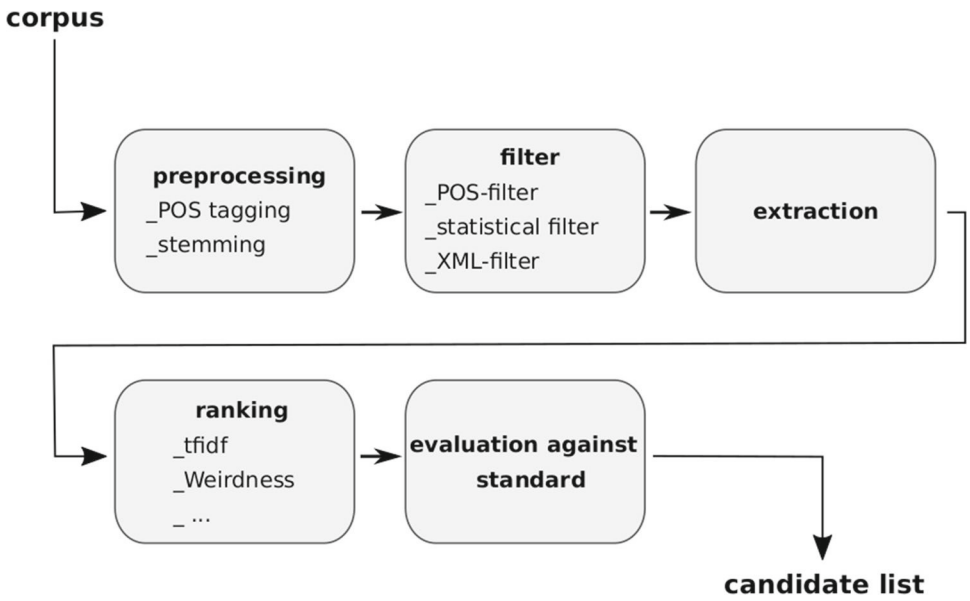

Fig. 5 Key Components of the Method Applied for ATE

we extracted all candidates from the corpus that matched certain POS patterns (see [25]), had a higher relative frequency in our target corpus than in a non-specialized language reference corpus[15] and were not marked as example sentence or bibliographical reference by the semantic markup of the XML files. We then ranked the extracted candidates by implementing an array of well-established ranking algorithms (i.a. Weirdness [1], Log-Likelihood-based distance [13], TF-IDF [39], PageRank [6] and *C*-value [15]) and evaluated the algorithms' performance by means of comparison to a standard retrieved from manual extraction by a terminology/ linguistics expert from a randomly chosen subset of 120 documents. We found that for our task Weirdness, the ratio of a candidate's relative frequency in the domain-specific target corpus and its relative frequency in the non-specialized language reference corpus,[16] showed the best overall performance considering precision and recall.

The end result of the ATE is a ranked list of approx. 31,000 candidates, an excerpt of which is shown in Fig. 6. We checked every candidate against our existing resources to evaluate if the candidate is already part of our system and—if so—if there is a short reference text associated with the candidate. The text check considers if a candidate is a non-preferred term of a concept, in which case the candidate inherits the text status of the corresponding preferred term. As the addition of candidates changes the existing system, the checking process is constantly reiterated. The sheer amount of candidates poses a challenge in terms of implementation; the automatic proposal of candidates in the environment of our visualization tool as described in Sect. 3.2.1 is one way of facilitating this process.

3.4 SKOS representation

In order to facilitate interoperability and integration of our resource with other resources, we want to make our resource available in different formats. As mentioned in Sect. 2.3, we identified SKOS as a possible format for the KO community. Our test implementation of SKOS is outlined in [41]. Here, we give some more details on the considerations regarding mapping between our data categories and SKOS vocabulary as well as its technical implementation using an RDB-to-RDF tool.

3.4.1 General mapping decisions

Most of our data categories can be directly represented using SKOS vocabulary. Only few categories need other existing RDF(S) vocabularies; also, we define one IDS-specific property. In the following, we discuss the use of these vocabularies in our scenario.

SKOS Because our macrostructure follows the general principles in the thesaurus design (cf. [37]), our resource fits well into a SKOS representation. Most importantly, because of its onomasiological design, the basic units of our resource,

[15] As reference corpus we used a sample from DeReKo corpus (cf. [26]) which covers various text types and genres.

[16] Ahmad et al.: [1, p. 720]:

$$\frac{w_s/t_s}{w_g/t_g}$$

Note that Ahmad et al. use slightly different labels for the corpora: w_s refers to the frequency of a word in the specialist language corpus (the domain-specific target corpus), w_g to the frequency of a word in the general language corpus (the non-specialized language reference corpus), t_s to the total count of words in the specialist language corpus, and t_g refers to the total count of words in the general language corpus.

Show 15 entries Search:

candidate	struc	gram	in.title	in.sub	freq	weird	in.db	txt
pras	n	1	yes	yes	1262	1443.6	yes	no
genitiv	n	1	yes	yes	1252	1432.2	yes	no
præteritum	n	1	yes	yes	1251	1431.1	yes	no
syntakt	n	1	yes	yes	2212	1265.2	no	no
partizip	n	1	yes	yes	993	1135.9	yes	no
verbform	n	1	yes	yes	914	1045.6	yes	no
akkusativ	n	1	yes	yes	834	954.0	yes	no
stammform	n	1	yes	yes	829	948.3	no	no
wortstell	n	1	yes	yes	778	890.0	yes	no
definit	n	1	yes	yes	778	890.0	yes	no
flektiert	n	1	yes	yes	763	872.8	yes	no
stark verb	an	2	yes	yes	704	810.2	yes	no
subklass	n	1	yes	yes	690	789.3	no	no
nominativ	n	1	yes	no	696	723.8	yes	no
lexikal	n	1	yes	yes	625	715.0	no	no
candidate	struc	gram	in.title	in.sub	freq	weird	in.db	no

Fig. 6 Top 15 Extracted Candidates, Ranked by Weirdness

i.e., concepts, can be mapped into the class `skos:Concept` without restrictions. Moreover, since we do not put strong transitivity constraints on the relations between concepts, we map both hierarchical relations to intransitive properties `skos:broader` and `skos:narrower` and use the symmetric, generally intransitive property `skos:related` for the associative relation. Note that the distinction between the partitive and generic hierarchical relation, which is made in the original resource, is lost in the SKOS representation (cf. [5]).

We also use basic SKOS vocabulary to map data categories for the description of single concepts that are located at the microlevel of our resource. We use `skos:scopeNote` to map so-called theory tags[17] for indicating that a concept is exclusively used within a specific theory of grammar. As mentioned in 3.2, the actual description of a single concept in our resource consists of several parts (definition, explanation, inventory, corpus examples, example sentences, note). In order to keep our Knowledge Organization System indeed Simple, we include only the definition part of the concept description. As a consequence, the original concept description is more rich than the resulting SKOS representation. However, it is generally possible to cover other parts as well by using different note types, e.g., `skos:note` for explanation, `skos:example` for examples and `skos:editorialNote` for notes.

Besides `skos:Concept`, SKOS does not define any other resource class. However, we follow the so-called term autonomy, which is one of best practices in terminology management (cf. [9], M2-6), and use, therefore, the SKOS-XL extension to map some of our term-related data categories; we use `skosxl:Label` to identify terms as resources as well as properties `skosxl:literalForm` and `skosxl:prefLabel` and `skosxl:altLabel` as defined in the SKOS reference.

Other RDF(S) Vocabularies SKOS natively relies on `rdfs:type` to state that a resource is an instance of class `skos:Concept`. Therefore, we need to include this RDFS vocabulary item in our representation. Also, we map the primary keys (IDs) from our database to Dublin Core `dc:identifier`. Finally, we introduce the prop-

[17] For a detailed description of our approach to these theory tags, see [41, pp. 207–211].

erty `ids:conceptRing` to state that a resource belongs to a particular concept ring. The notion of a concept ring is described in [42], (cf. [41]).

3.4.2 Implementation

As mentioned above, our re-designed resource is stored and maintained in an object-relational database. As an added value, we want to make it available in the RDF format. The so-called RDB-to-RDF approach allows to present the relational data as RDF triples. The RDF triples can be created on the fly or accessed as a dump. A comprehensive overview of this approach, including a classification and a survey of the tools, is given in [30]. Accordingly, RDB-to-RDF approach can be applied in various scenarios in which 'the data should remain hosted and delivered by the legacy RDBs' [30]. Michel et al. [30] distinguish three mutually non-excluding goals for the approach:

- to make the opaque relational data available to the standard web-browsing tools
- to make the data available to the Linked Open Data (LOD) paradigm
- to integrate data from heterogeneous applications in order to benefit from the data synergy

Our case predominantly involves the latter two goals, i.e., data integration within, but not exclusively, the LOD paradigm.

In order to test the application of a RDB-to-RDF tool for our purposes, we used the D2RQ platform[18] in a test environment. D2RQ platform is an academic, open-source RDB-to-RDF suite (cf. [30]). It consists of the D2RQ mapping language and different tools for data publishing and retrieval.

D2RQ mapping language is a declarative language. Its constructs are used to describe how to translate the specification and the content of the relational database (tables, columns, their names, the data itself, etc.) into RDF triples. These translation rules are written as RDF triples in Turtle notation and stored in a file. The D2RQ language does not comply with the W3C's recommendation for a language for 'expressing customized mappings from relational databases to RDF datasets'[19]—R2RML (cf. [30]).

Although the mapping file can be created manually, the D2RQ offers a tool—Generate Mapping—for an automatic generation of a default mapping file (mapping.ttl). This tool uses the so-called direct mapping approach, which 'automatically creates an ad hoc RDFS or OWL vocabulary reflecting the structure of the relational schema.' [30]. However, it is possible to manually edit the automatically generated file to achieve a domain semantics-driven mapping (cf. [30]).

Further, the D2RQ platform offers various tools for triples' publication and retrieval, supporting Data Materialization with an RDF dump (tool: *dump-rdf*) as well as On-Demand Mapping with a Query-based Access (tools: *D2R Server* and *d2r-query*) (see [30]). Finally, the platform can also be used with the Jena API, supporting the creation of Semantic Web applications.

In our local test environment, we used the tool for auto-generation of the default mapping file as well as D2R server for publishing and retrieving data.

As mentioned in [41], the customization of the default mapping file was straightforward. It entailed replacing the default D2RQ class and property names with the SKOS-specific vocabulary as described in Sect. 3.4.1. Because D2RQ language does not include specification of inverse properties,[20] we also limited the number of the automatically generated inverse relations of `skos:broader` to 0 and manually added the mapping for its inverse (`skos:narrower`). Finally, we defined a translation table for the values of the language attribute (deutsch, englisch, ungarisch, etc.) in order to comply with the literals of the `xml:lang` property (de, en, hu).

For publication and retrieval, we used the default setting of the D2R server. As shown in Fig. 7, the D2RQ platform gives access to our up-to-date resource as SKOS triples on the fly and there is no need to replicate the data into a triple store (cf. [41, p. 622]).

3.4.3 Further considerations

Because we model our resource independently of SKOS, we expect that a transfer to SKOS might reveal some structures in our model that are not consistent with SKOS definition.[21] Therefore, we plan to use tools for SKOS validation prior to the initial publishing. The results of the validation will feed back to the original model, allowing us to make some changes to the source data. Hence, we expect that using SKOS will improve the quality of the semi-formal representation of our resource by making it more clear and well-defined.

4 General classification, discussion and further application

So far, we focused on project progress; from a scholarly perspective, we predominantly referred to the discipline

[18] http://d2rq.org/.

[19] https://www.w3.org/TR/r2rml/.

[20] https://github.com/d2rq/d2rq/issues/45.

[21] For instance, clashes of associative and hierarchical links, see example 27 in the SKOS reference, https://www.w3.org/TR/2009/REC-skos-reference-20090818.

Description of http://localhost:2020/resource/TB_KONZEPT/123:

property	hasValue	isValueOf
ids:conceptRing	db:TB_KONZEPTRING/123	-
dc:identifier	123	-
rdf:type	skos:Concept	-
skos:broader	db:TB_KONZEPT/1897	-
skos:definition	"Kommunikanten-Pronomina sind eine Subklasse der Pronomina. Ihre Funktion ist der Verweis auf den aktuellen Sprecher bzw. die Sprechergruppe oder den aktuellen Hörer bzw. die Hörergruppe."	-
skos:narrower	db:TB_KONZEPT/103	-
skos:narrower	db:TB_KONZEPT/246	-
skos:related	db:TB_KONZEPT/317	-
skos:scopeNote	"Systematische Grammatik"	-
skosxl:altLabel	db:TB_TERM/619	-
skosxl:altLabel	db:TB_TERM/8587	-
skosxl:altLabel	db:TB_TERM/8710	-
skosxl:altLabel	db:TB_TERM/8711	-
skosxl:altLabel	db:TB_TERM/8712	-
skosxl:altLabel	db:TB_TERM/8713	-
skosxl:altLabel	db:TB_TERM/8714	-
skosxl:altLabel	db:TB_TERM/8715	-
skosxl:altLabel	db:TB_TERM/8716	-
skosxl:altLabel	db:TB_TERM/9836	-
skosxl:altLabel	db:TB_TERM/9837	-
skosxl:prefLabel	db:TB_TERM/615	-
skos:related	-	db:TB_KONZEPT/185
skos:related	-	db:TB_KONZEPT/394
skos:related	-	db:TB_KONZEPT/577
skos:related	-	db:TB_KONZEPT/717

Powered by D2R Server

Fig. 7 Screenshot of the SKOS Representation of our Resource Using D2RQ Platform [41, p. 205]

of terminology, but we also touched upon some practical ideas in Knowledge Organization. Both perspectives are non-excluding, as can be seen in the ISKO Encyclopedia of Knowledge Organization (IEKO), which lists Terminology as an adjacent discipline to Knowledge Organizations [22, Homepage]. In this concluding section, we discuss the progress and results of the system's re-design from the Knowledge Organization point view in greater detail. In particular, we investigate how the transition from the previous resources to the new unified terminology marks the emergence of a new, coherent—and more powerful than before—system.

4.1 KOS classifications

A comprehensive view of Knowledge Organization Systems (KOSs) is given in [29]. Referring to [21], Mazzocchi distinguishes a broader and a narrower reading of KOS:

- 'According to the broad reading, the notion of KOS refers, for instance, to encyclopedias, libraries, bibliographic databases, and, even in a more general sense, to conceptual systems, theories, disciplines, cultures, as well as to the social division of labor in society and models of activity and process systems in different domains.'
- The narrower reading comprises ' [...] functional items designed for organizing knowledge and information, and making their management and retrieval easier. [...] Since they are basically made of terms/concepts and, many of them, semantic relations, KOSs are also depicted as semantic tools (e.g., [20]).'

Furthermore, there are different typologies to subdivide KOSs in a narrower sense. Some of these typologies are unstructured (cf. [38]); however, many of them assume a hierarchical complexity gain so that the KOSs form a so-called *semantic staircase*.

Applying these conceptualizations to our setting, we can, in the first instance, identify *grammis* as a KOS in a broader sense, as it is a comprehensive knowledge resource of an encyclopaedic type. We also see that the initial terminological resources are different types of KOSs in a narrower sense, occupying different positions on the semantic staircase; most importantly, we can place the less structured grammatical dictionaries at the bottom, whereas the old thesaurus/ontology occupies the position in the middle of the staircase. Our re-design decisions combine these (narrower) KOSs into one KOS. Putting the thesaurus' macrostructure at the center of this new KOS keeps the highest possible position on the staircase within our resources and avoids, in turn, any complexity loss. What is more, it means a valuable complexity gain for the content of the original dictionaries, elevating them to the middle of the staircase. However, it could also be argued that the merged resource gains in complexity altogether, exceeding the staircase position of a regular thesaurus; while the structure, i.a. data model, stays the same, the content becomes more complex by adding multiple theoretical perspectives. This way we address the issue of *classificatory perspectivism* as discussed in [29].

4.2 Discussion

The conceptual re-design of *grammis'* terminological resources is beneficial from both methodological and practical perspective. The re-designed resources comply with best practice in thesaurus and terminology management (cf. [42]) and—as outlined in 3.4.1—can be mapped into a SKOS representation. This opens up possibilities to link our resource with similarly structured resources in the context of Linguistic Linked Open Data.

From an author's perspective, content is now easier to manage and to revise. Moreover, the application of visualization techniques renders the macrostructure more accessible.[22] Apart from the obvious benefit of simplifying the revision and expansion of the macrostructure, terminology authors can now use the (revised) macrostructure as a guideline for writing or revising the reference texts of the microstructure. This results in the above-mentioned complexity gain of the reference texts as compared to the original dictionary entries; now, a reference text mentions more macrostructurally related concepts. We regard this as an increase in textual complexity which benefits end-users, as they are presented with more exhaustive concept descriptions.

We also expect an improved interplay between the terminological resource and *grammis*. As mentioned in Sect. 1, the macrostructure is used in information retrieval, e.g., full-text search and reference-generation. We believe that reviewing and expanding the macrostructure will lead to a qualitative improvement of information retrieval results.

Moreover, the revised macrostructure also plays a role in a prospective application that will further improve information retrieval and provides the end-users an alternative way of accessing information. We will discuss this application in the following section.

4.3 A prospective application

A notorious problem of complex information systems such as *grammis* is the finding of appropriate content that suits the concrete question of the current user—often a search for a needle in a haystack. Apart from traditional retrieval utilities—(semantically enriched) full-text search, keyword lists, table of contents, etc.—we believe that natural language processing combined with a powerful terminological KOS could play an important role in the exploration process, inasmuch as it allows users to gain accurate access to appropriate pieces of information without the need of learning specialized query languages, or without the time-consuming task of filling out complex search forms. Moreover, it could offer a way out in situations where users, due to terminological uncertainties, are unable to name a certain problem or the phenomenon they look for.

Related work exists for the underlying idea: The Linguist's Search Engine (LSE) tried to offer an 'intuitive, linguistically sophisticated but user-friendly way' [33] by adopting a strategy called *Query By Example* for web searches based on POS tagging. TIGER Corpus Navigator allows users to navigate a corpus, based on example sentences that represent abstract linguistic concepts [19]. Most recently, [3,4] introduced example-based treebank querying as a way to search within annotated corpus resources. They allow users to enter natural language sentences or phrase segments as a basis to look up similar syntactic constructions.

We will expand this methodologically attractive idea in several ways: First, we apply it not on annotated treebank corpora, but on heterogenous structured resources assembled within a linguistic information system. Since these systems provide information about natural language phenomena (object of investigation) with the help of natural language (means of communication), they consequently should be explorable with the same means of natural language. Second, we see example-based querying as an ideal way to open up scientific resources to a broader public. Our focus is not restricted to users who lack experience in specialized query languages, but also on users with different terminological background, or even without explicit knowledge of linguistic terminology. The objective is to combine a language-oriented retrieval approach, which is supposed to be suitable for both

[22] Prospectively, the end-users of grammis might also benefit from visualization techniques, similar to those described in 3.2.1.

linguists and linguistic laymen, with a data-oriented foundation.

As already mentioned, the *grammis* information system comprises nearly XML-coded hypertext documents on grammatical topics that establish a corpus of specialized language. Furthermore, dictionaries of selected linguistic phenomena (like verb valency, genitive formation, connectors, affixes, prepositions) contribute textual entries with customized XML microstructures. Both information types are valuable foundations for example-based querying, in the sense that they contain large quantities of natural language examples for illustration purposes and that they are completely categorized by terminological keywords. Keywords are organized as controlled vocabularies—covering and interconnecting different linguistic theories and schools—within a the re-designed terminology management system.

Out of the *grammis* specialist hypertexts and lexical entries, all XML-coded everyday language sentences are added to a corpus database of tagged examples. To enrich these approximately samples with POS and morphological annotations about case, number, gender etc., we use the statistical tagger *MarMot* [31], built upon conditional random fields (CRF). For each sentence, the corpus database stores a back reference to the source document and corresponding keywords.

The example-based retrieval algorithm now takes over in cases where a user does not formulate his search inquiry terminologically, and where simple word-based lookups yield no satisfactory result. Instead, each user input is regarded as prototypical example sentence or phrase and undergoes syntax-based processing. After morpho-syntactic annotations are added, the layer operates on the enriched input dataset and tries to identify similar POS-constructions. If this attempt is successful, it can either link directly to the corresponding hypertext documents or dictionary entries, or identify related information units by exploiting the attached keywords. These keywords used to be independent from the terminology resources. Recently, we have aligned the keyword list with the terminological KOS. Hence, in addition to the benefits discussed in Sect. 4.2 *Query By Example* might be a further application for our re-designed terminological resource.

4.3.1 Use of case for date specifications

Table 1 shows the tagged input of the first everyday language input example (‘am Freitag, den 13.’; English equivalent: ‘on Friday, the 13th’). Obviously, the underlying—but not explicitly expressed—question concerns the correct use of case within a German date specification: is the combination of dative and accusative acceptable, or should dative be maintained for the whole phrase (that would then be: ‘am Freitag, dem 13.’)?

Table 1 First Query Example as CRF input

Token	POS	Case	Num	Gen
am	APPRART	dat	sg	masc
Freitag	NN	dat	sg	masc
den	ART	acc	sg	masc
13.	ADV	acc	0	0

Table 2 Second Query Example as CRF input

Token	POS	Case	Num	Gen
das	ART	nom	sg	neut
Auto	NN	nom	sg	neut
von	APPR	0	0	0
meinem	PPOSAT	dat	sg	masc
Vater	NN	dat	sg	masc

And indeed, when applying the back reference model as described above, the algorithm references a suitable explanatory corpus text containing similar example sentences.[23] The corresponding keywords of this document are *Akkusativ (accusative)*, *Dativ (dative)*, *Datum (date)*, *Deklination (declension)*, *Flexion (inflection)*, *Kasus (case)*.

4.3.2 Use of genitive constructions

As a second example (‘das Auto von meinem Vater’; English: ‘the car of my father’), we choose an authentic user query that a human native speaker would probably classify as somehow related to the use of genitive constructions, although it does not contain any genitives at all (see Table 2). A possible genitive construction would be ‘meines Vaters Auto’; English: ‘my father’s car’.

A syntactically similar example is found within a *grammis* hypertext on the use of the preposition ‘von’ and dative case, compared to the ‘high-order’ style of genitive attributes. Consequently, our classification algorithm generates an expedient link to this document.[24] Its classifying keywords are *Attribut (attribute)* and *Genitiv (genitive)*.

5 Summary

In this paper, we discussed the re-design details of a terminological resource that integrate best practices in Terminology and KOS management. We showed how our integrated tool results in an improved usability for terminology authors in various management steps such as retrieval, processing, and

[23] https://grammis.ids-mannheim.de/fragen/3185.

[24] https://grammis.ids-mannheim.de/fragen/4550.

publication. We also outlined current and prospective benefits for the human terminology user as well as for various machine applications.

Currently, we are focusing on content revision, but we are also exploring further options for innovative access to our resource and its further integration with other KOSs both in broader and narrower sense.

References

1. Ahmad, K., Gillam, L., Tostevin, L.: University of surrey participation in TREC8: weirdness indexing for logical document extrapolation, retrieval (WILDER). In: Voorhees, E., Harman, D. (eds.) NIST Special Publication 500–246: The Eighth Text Retrieval Conference (TREC-8), Gaithersburg, MA, pp. 717–724 (1999)
2. Almende, B.V., Thieurmel, B.: visNetwork: network visualization using 'vis.js' library. R package version 1.0.3. https://CRAN.R-project.org/package=visNetwork (2016). Accessed 1 June 2017
3. Augustinus, L., Vandeghinste, V., Van Eynde, F.: Example-based treebank querying. In: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012), European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2012/pdf/756_Paper.pdf (2012). Accessed 8 Nov 2017
4. Augustinus, L., Vandeghinste, V., Vanallemeersch, T.: Poly-gretel: cross-lingual example-based querying of syntactic constructions. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2016/pdf/486_Paper.pdf (2016). Accessed 8 Nov 2017
5. Baker, T., Bechhofer, S., Isaac, A., Miles, A., Schreiber, G., Summers, E.: Key choices in the design of Simple Knowledge Organization System (SKOS). Web Semant. Sci. Serv. Agents World Wide Web **20**, 35–49 (2013)
6. Brin, S., Page, L.: The anatomy of a large-scale hypertextual search engine. Comput. Netw. ISDN Syst. **30**(1–7), 107–117 (1998)
7. Bubenhofer, N., Schneider, R.: Using a domain ontology for the semantic-statistical classification of specialist hypertexts. In: Papers from the Annual International Conference on Computational Linguistics 'Dialogue'. Moscow, 26 May 2010/30 May 2010, pp. 622–628 (2010)
8. Chang, W., Cheng, J., Allaire, J., Xie, Y., McPherson, J.: Shiny: Web Application Framework for R. R package version 1.0.0. https://CRAN.R-project.org/package=shiny (2017). Accessed 1 June 2017
9. Deutscher Terminologie-Tag eV: Terminologiearbeit—Best Practices, 2nd edn (2014)
10. DIN 2331: Begriffssysteme und ihre Darstellung (1980)
11. DIN 2342:2011-08: Begriffe der Terminologielehre (2011)
12. Drewer, P., Massion, F., Pulitano, D.: Was haben Wissensmodellierung, Wissensstruktur, künstliche Intelligenz und Terminologie miteinander zu tun? Technical Report, Deutscher Terminologie-Tag e.V (2017)
13. Dunning, T.: Accurate methods for the statistics of surprise and coincidence. J. Comput. Linguist. Spec. Issue Using Large Corpora **19**(1), 61–74 (1993)
14. Faber, P.: Frames as a framework for terminology. In: Kockaert, H.J., Steurs, F. (eds.) Handbook of Terminology, vol. 1. John Benjamins Publishing Company, Amsterdam (2015)
15. Frantzi, K., Ananiadou, S., Mima, H.: Automatic recognition of multi-word terms: the C-value/NC-value method. Int. J. Digit. Libr. **3**(2), 115–130 (2000)
16. Früh, B., Deubzer, F.: Von der Terminologieverwaltung zur Wissensorganisation. Edition **16**(1), 27–32 (2016)
17. Fu, B., Noy, N.F., Storey, M.A.: Indented tree or graph? A usability study of ontology visualization techniques in the context of class mapping evaluation. In: Alani, H., Kagal, L., Fokoue, A., Groth, P., Biemann, C., Parreira, J.X., Aroyo, L., Noy, N., Welty, C., Janowicz, K. (eds.) The Semantic Web—ISWC 2013: 12th International Semantic Web Conference, Sydney, NSW, Australia, 21–25 Oct 2013, Proceedings, Part I, vol. 8218. Springer, Berlin, pp. 117–134 (2013). https://doi.org/10.1007/978-3-642-41335-3_8
18. Hausmann, F.J.: Lexikographie. In: Schwarze, C., Wunderlich, D. (eds.) Handbuch der Lexikologie, pp. 367–398. Athenäum, Königstein/Ts (1985)
19. Hellmann, S., Unbehauen, J., Chiarcos, C., Ngonga Ngomo, A.C.: The tiger corpus navigator. In: Proceedings of the Ninth International Workshop on Treebanks and Linguistic Theories (TLT-9), Northern European Association for Language Technology (NEALT), pp. 91–102 (2010)
20. Hjørland, B.: Semantics and knowledge organization. Annu. Rev. Inf. Sci. Technol. **41**(1), 367–405 (2007)
21. Hjørland, B.: What is Knowledge Organization (KO)? Knowl. Organ. **35**(2/3), 86–102 (2008)
22. Hjørland, B. (ed.): ISKO Encyclopedia of Knowledge Organization (IEKO), online edn. http://www.isko.org/cyclo/ (2016). Accessed 30 Sept 2017
23. ISO 25964-1:2011: Information and documentation—thesauri and interoperability with other vocabularies—Part 1: thesauri for information retrieval (2011)
24. ISO 30042: Systems to manage terminology, knowledge and content—TermBase eXchange TBX, 1st edn (2008)
25. Justeson, J.S., Katz, S.M.: Technical terminology: some linguistic properties and an algorithm for identification in text. Nat. Lang. Eng. **1**(1), 9–27 (1995)
26. Kupietz, M., Keibel, H.: The Mannheim German Reference Corpus (DeReKo) as a basis for empirical linguistic research. In: Minegishi, M., Kawaguchi, Y. (eds.) Working Papers in Corpus-based Linguistics and Language Education, 3, Tokyo University of Foreign Studies, Tokyo, pp. 53–59 (2009)
27. Lang, C., Suchowolec, K., Schneider, R.: Extracting technical terminology from linguistic corpora. In: Proceedings of Grammar and Corpora 2016, Mannheim, Heidelberg University Publishing (heiUP), Heidelberg (2018)
28. León Araúz, P., Magaña Redondo, P.J.: EcoLexicon: contextualizing an environmental ontology. In: Proceedings of the Terminology and Knowledge Engineering (TKE) Conference, pp. 341–355 (2010)
29. Mazzocchi, F.: Knowledge organization system (KOS). In: [22], version 1.1. http://www.isko.org/cyclo/kos (2017). Accessed 30 Sept 2017
30. Michel, F., Montagnat, J., Faron-Zucker, C.: A survey of RDB to RDF translation approaches and tools. Technical Report, Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis (I3S) (2014)
31. Mueller, T., Schmid, H., Schütze, H.: Efficient higher-order CRFs for morphological tagging. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics, Seattle, Washington, USA, pp. 322–332. http://www.aclweb.org/anthology/D13-1032 (2013). Accessed 8 Nov 2017
32. R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/ (2016). Accessed 8 Nov 2017

33. Resnik, P., Elkiss, A.: The linguist's search engine: An overview. In: Proceedings of the ACL 2005 Interactive Poster and Demonstration Session, Association for Computational Linguistics (ACL), pp. 33–36 (2005). https://doi.org/10.3115/1225753.1225762
34. Schmid, H.: Improvements in part-of-speech tagging with an application to German. In: Proceedings of the ACL SIGDAT-Workshop, Dublin, Ireland, pp. 1–9 (1995)
35. Schneider, R., Gottron, T.: A hybrid approach to statistical and semantical analysis of Web documents. In: Proceedings of the IASTED International Conference Internet and Multimedia Systems and Applications (EuroImsa), pp. 115–120 (2009)
36. Schneider, R., Schwinn, H.: Hypertext, Wissensnetz und Datenbank: Die Web-Informationssysteme grammis und Progr@mm. In: Berens, F.J., Steinle, M. (eds.) Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache, IDS Eigenverlag, Mannheim, pp. 337–346. http://nbn-resolving.de/urn:nbn:de:bsz:mh39-24719 (2014). Accessed 7 Nov 2017
37. Sejane, I.: Wissensrepräsentation Linguistik. Modellierung, Potenzial und Grenzen am Beispiel der Ontologie zur deutschen Grammatik im GRAMMIS-Informationssystem des IDS, Mannheim. Ph.D. Thesis, Ruprecht-Karls-Universität Heidelberg (2010)
38. Souza, R.R., Tudhope, D., Almeida, M.B.: Towards a taxonomy of KOS: dimensions for classifying knowledge organisation systems. Knowl. Organ. **39**(3), 172–179 (2012)
39. Spärck Jones, K.: A statistical interpretation of term specifity and its application in retrieval. J. Doc. **28**(1), 11–21 (1972)
40. Suchowolec, K.: Sprachlenkung—Aspekte einer übergreifenden Theorie. Frank & Timme, Berlin, dissertation, Stiftung Universität Hildesheim (2018)
41. Suchowolec, K., Lang, C., Schneider, R., Schwinn, H.: Shifting complexity from text to data model. In: Gracia, J., Bond, F., McCrae, J.P., Buitelaar, P., Chiarcos, C., Hellmann, S. (eds.) Language, Data, and Knowledge. Proceedings of the First International Conference, LDK 2017, 19 June 2017/20 June 2017, Galway, Ireland, Springer, Cham, no. 10318 in Lecture Notes in Artificial Intelligence, pp. 203–212 (2017)
42. Suchowolec, K., Lang, C., Schneider, R.: Grammar and its terminology. Re-designing terminology management system according to best practices (forthcoming)
43. Winkler, W.: String comparator metrics and enhanced decision rules in the Fellegi-Sunter model of record linkage. In: Proceedings of the Section on Survey Research Methods (American Statistical Association), pp. 354–359 (1990)
44. Zifonun, G., Hoffmann, L., Strecker, B.: Grammatik der deutschen Sprache: Bd. 1–3. de Gruyter, Berlin (1997)