

**PREPRINT.** TO APPEAR IN: RUFUS H. GOUWS, ULRICH HEID, WOLFGANG SCHWEICKARD & HERBERT ERNST WIEGAND(EDS.): DICTIONARIES. AN INTERNATIONAL ENCYCLOPEDIA OF LEXICOGRAPHY. SUPPLEMENTARY VOLUME: RECENT DEVELOPMENTS WITH SPECIAL FOCUS ON COMPUTATIONAL LEXICOGRAPHY. BERLIN/NEW YORK: DE GRUYTER. (DATE OF SUBMISSION: AUGUST 2009)

# TEXTUAL STRUCTURES IN ELECTRONIC DICTIONARIES COMPARED WITH PRINTED DICTIONARIES: A SHORT GENERAL SURVEY

CAROLIN MÜLLER-SPITZER, INSTITUT FÜR DEUTSCHE SPRACHE MANNHEIM

- 1 Introduction
  - 2 Terminological and formal preliminary remarks
  - 3 Basic differences between textual structures in printed versus electronic dictionaries
  - 4 Creating order in 'terminologyland' for textual structures in EDs
  - 5 Describing and analysing actual EDs and planning new EDs
  - 6 Closing remarks
- References

## 1 INTRODUCTION

Textual structures in printed dictionaries are well known, adequately researched, and rather exhaustively described (cf. articles 3-10). This article investigates whether or not the models of textual structures in printed dictionaries can be applied to electronic dictionaries (EDs); or, more precisely, which parts of the order and terminology of textual structures in printed dictionaries are applicable to electronic ones and which differences one should be aware of. The focus will be on online dictionaries because these are the most important kind of digital dictionary, and will only become more important in the future. Furthermore, the emphasis will be more on potential future forms of online dictionaries than on current ones which are still sometimes produced as copies of their printed counterparts (cf. Tarp 2009: 21, also other articles in Bergenholtz/Nielsen/Tarp 2009).

To approach this question, firstly, basic differences between textual structures in electronic versus printed dictionaries will be discussed. Secondly, further terminological and formal preliminary remarks will be made. The main part of the article will then follow. To adapt de Schryver's idea of "Creating order in dreamland" in the article "Lexicographer's dreams in the electronic dictionary age" (de Schryver 2003), the aim here is to begin 'creating order in terminologyland' for textual structures in electronic dictionaries. A definitive order cannot be given here because electronic lexicography today involves constant change. In order to discuss the order of textual structures in EDs, not only theoretically, but also in concrete terms, their basic properties will be illustrated by means of a notional online dictionary. Following on from this fictitious scenario, a provisional survey of textual structures in EDs will be presented. Thereby, the focus is less on current online dictionaries than on the possibilities which the new medium provides. Finally, an explanation will be given as to how this view of structures in electronic dictionaries is useful for analysing actual EDs and for planning new ones. The overall aim here is not to introduce new kinds of textual structure in EDs and a corresponding terminology in detail, but to point out some constitutive differences between textual structures in printed dictionaries and those in electronic dictionaries.

## 2 TERMINOLOGICAL AND FORMAL PRELIMINARY REMARKS

### 2.1 THE TERM ED

In some works of metalexigraphic research, one area of discussion is whether the term "electronic dictionary" is still appropriate if the named object is a very innovative product that bears no resemblance to a printed dictionary. This problem is even more evident if the book metaphor is part of the term itself, as in the German "Wörterbuch", the Danish "ordbog", or the Dutch "woordenboek". Alternative terms are "leximat" (Tarp 2008: 121), "multifunctional lexicographic database" (Pajzs 2009: 327), or "vocabulary information system" (cf. Müller-Spitzer 2003: 147ff.). As far as content is concerned, it would be better to avoid the term "dictionary" in order to clarify terminologically the difference between printed dictionaries and digital lexicographic information systems, especially if one has very innovative products in mind. However, the term electronic dictionary (ED) is used here because it accords with the terminology of this handbook (cf. the table of contents), and it would be confusing to diverge from that. Here, EDs are regarded as products compiled by humans and produced for human use (cf. de Schryver 2003: 145f., Müller-Spitzer 2003: 142ff.). Dictionaries are "utility products or tools which are, or should be, designed to satisfy certain human needs" (Tarp 2009: 29). Further criteria which have to be taken into account in order to distinguish dictionaries from other reference works are not discussed here (cf. e.g., Wiegand 1998a, Wiegand 2005, Tarp 2008: 4ff.) Dictionaries which are compiled purely automatically or word databases as well as dictionaries compiled for machine-internal use are not the subject of this article. Rather, the focus will be on standardised (cf. Wiegand 1990) linguistic EDs for general purposes.

### 2.2 TEXTUAL STRUCTURES

Another terminological question is whether we can still speak of textual structures in electronic dictionaries if an essential criterion of innovative EDs is to connect data in different media? The term "lexicographic text" is based on a broad concept of text and is well established. Furthermore, the whole theory of lexicographic texts is based on the view that dictionary articles are texts (e.g., cf. Wiegand 1990). Therefore, in order to have a basis for comparing textual structures in printed dictionaries and structures in EDs, the basic concept of textual structures should be maintained. However, in the context of EDs, it is worth noting that "text" includes data represented in different media: as text, audio files, videos, graph-based views, etc.

## 2.3 DEFINITION OF TEXTUAL STRUCTURES IN EDS

Textual structures in EDs are much more difficult to define than textual structures in printed dictionaries due to the fact that it is not possible to look into the underlying database and the corresponding programme procedures (cf. section 3). In particular, it is difficult to determine the set of elements of a special structure in its entirety.

Therefore, in describing the textual structures of EDs, the aim is not to meet the requirements of definitions that define the elements of a set completely, as Wiegand does in his theory of lexicographic texts (cf. e.g., Wiegand 1989a, 1990, 1998a). Rather, a prototypical approach is chosen. The order of textual structures in EDs given under 4 is to be understood as a prototypical classification to which textual structures correspond to a greater or lesser degree (cf. the approach of typologisation of dictionary portals in article 75a). A more formal definition requires a more established practice and further metalexigraphic research focused on EDs. In the area of printed dictionaries, it took 20 years of research and several hundred pages of publications to gain the current level of theoretical clarifications.

## 3 FUNDAMENTAL DIFFERENCES BETWEEN TEXTUAL STRUCTURES IN PRINTED DICTIONARIES AND THOSE IN ELECTRONIC DICTIONARIES

Very generally speaking, printed dictionaries typically consist mainly of dictionary entries, accompanied by external texts such as user guidelines, indexes, and so on. These elements of a dictionary are partial lexicographic texts with special structural properties (cf. e.g., Wiegand 1990, Welker 2009: 9). For analysing and planning printed dictionaries, several kinds of structure are differentiated in printlexicographic research and are outlined below:

- (i) the data distribution structure (cf. Bergenholtz et al. 1999, Beißwenger 2002) which determines the distribution of lexicographic data to the different parts of a dictionary (article stretch, outer texts, etc.);
- (ii) the access structure (cf. Wiegand 2007, 2008a, article 5) which determines how users can access the lexicographic data;
- (iii) the macrostructure (cf. Wiegand 1989b, 1998b, article 4), consisting of the guiding elements which determine the order of the entries;
- (iv) the microstructure (cf. Wiegand 1989a, article 6), consisting of the lexicographic items in one article with their typographic and non-typographic indicators, and
- (v) the mediostructure (cf. Tarp 99, Wiegand 2002, article 7) between items.

In addition to this, the addressing structure (cf. Wiegand 2006, article 9), search zone structure (cf. Wiegand 2000), and textual architectures (cf. article 8) are identified as special properties of lexicographic texts.

What all these structures have in common is that they consist of a finite set of elements. This is because printed dictionaries (like all other books) are tangible physical objects with a finite amount of content. If one were to look at all the texts in a printed dictionary and describe and analyse them exhaustively, there would be no doubt that the whole printed dictionary had been exhaustively described. There is another vital property that all lexicographic texts in printed dictionaries share, which is that data and form merge into one another; they are inextricably linked with one other. The data is in a special layout, physically represented on paper, and not modifiable.

In electronic dictionaries, especially online dictionaries, things look “radically different” (cf. Tarp: 2008, 110). If you look at an online dictionary, normally only a virtual part of the whole dictionary can be seen. The dictionary in its entirety is not tangible because it is not limited by a single data medium like a CD-Rom or a book. Thus, it is much more difficult to determine what *the* microstructure or *the* macrostructure of the dictionary is. The set of elements of one structure is in its finitude not determinable. Typically, every user search question on the presentation level, for example where a search term is entered in a search box, starts a search procedure on the database. As a result of the user request, a virtual part of the dictionary is presented. Thus, searching, navigating and representation of the content in an electronic dictionary is a circular process that is only then analysable and describable if both levels – the database and the presentation level – are the objects of enquiry. However, it is worth noting that this modification process from the database to the presentation level is a deterministic one. Lexicographers, together with programmers, must determine which views correspond to which requests. However, this process of modification is not part of the lexicographic process. In an electronic dictionary, the modification of data is typically a continual process that is started again and again as the electronic dictionary is used. This modification may consist of selecting lexicographic content or reformatting the content into a different data format, for example an XML-based data format to HTML. In contrast, the lexicographic data of printed dictionaries is modified during the lexicographic process. But at the end, there is a printed dictionary where the data is presented in one final form. This end-point may be missing in innovative electronic dictionaries (for the term “innovative dictionaries” cf. Lehr 1996). More precisely, this end-point may have very variable forms that are more difficult to survey. Thus, the traditional dictionary articles with static data and fixed structures may be replaced by articles with “dynamic data which are [...] unique for each search related to a specific type of user in a specific type of user situation” (Tarp 2009: 29, cf. also Lew 2009: 257). Therefore, in order to grasp the structures of an electronic dictionary, it is necessary to look not only on the presentation level but also on the database level, and it is also necessary to identify which structures belong to which level.

Generally, the Internet is considered to be a “pull-medium” rather than a “push-medium” like television, radio, or books (cf. Bucher 2004: 10). Therefore, users are both sender *and* receiver. They are active in ‘pulling’ data from the website, saving relevant parts, etc. Thus, the Internet provides a very new form of communication in general. It is communication in an innovative combination with new media (cf. Bucher 2004: 12). This general property of the Internet as communication medium obviously has consequences for the property of online dictionaries as one type of text on the Internet. The process of pulling and, thus, representing lexicographic data according to a user request is essential for EDs and must not be lost from view when the textual structures of EDs are being looked at.

It should become obvious that it is necessary to scrutinise the transferability of textual structures from printed dictionaries to electronic ones. This is necessary for planning and analysing EDs and for every context where the most precise description possible of EDs is important. Another important case where a precise awareness of the structural properties of EDs is required is in developing user-adapted and function-related views of a lexicographic database (cf. Müller-Spitzer 2008b, Bergenholtz/Gouws 2007: 256).

## 4 CREATING ORDER IN ‘TERMINOLOGYLAND’ FOR TEXTUAL STRUCTURES IN EDS

### 4.1 THE BASIC PUBLICATION MODEL

In order to illustrate the complexity of the levels where textual structures exist while developing printed and electronic dictionaries, Figure 1 shows a publication model for lexicographic processes which are designed to be medium-neutral (cf. Müller-Spitzer 2003:161). It is not relevant in this context if most EDs are produced in such an environment. The focus here must be on the *potential* complexity of textual structures in the context

of EDs. For this reason, the notional ED described in 4.2 also provides user-adapted views. At the moment, there are only a small number of dictionaries which are going in this direction. However, in the future, hopefully more and more dictionaries will seize this opportunity if this corresponds to the needs of the users.

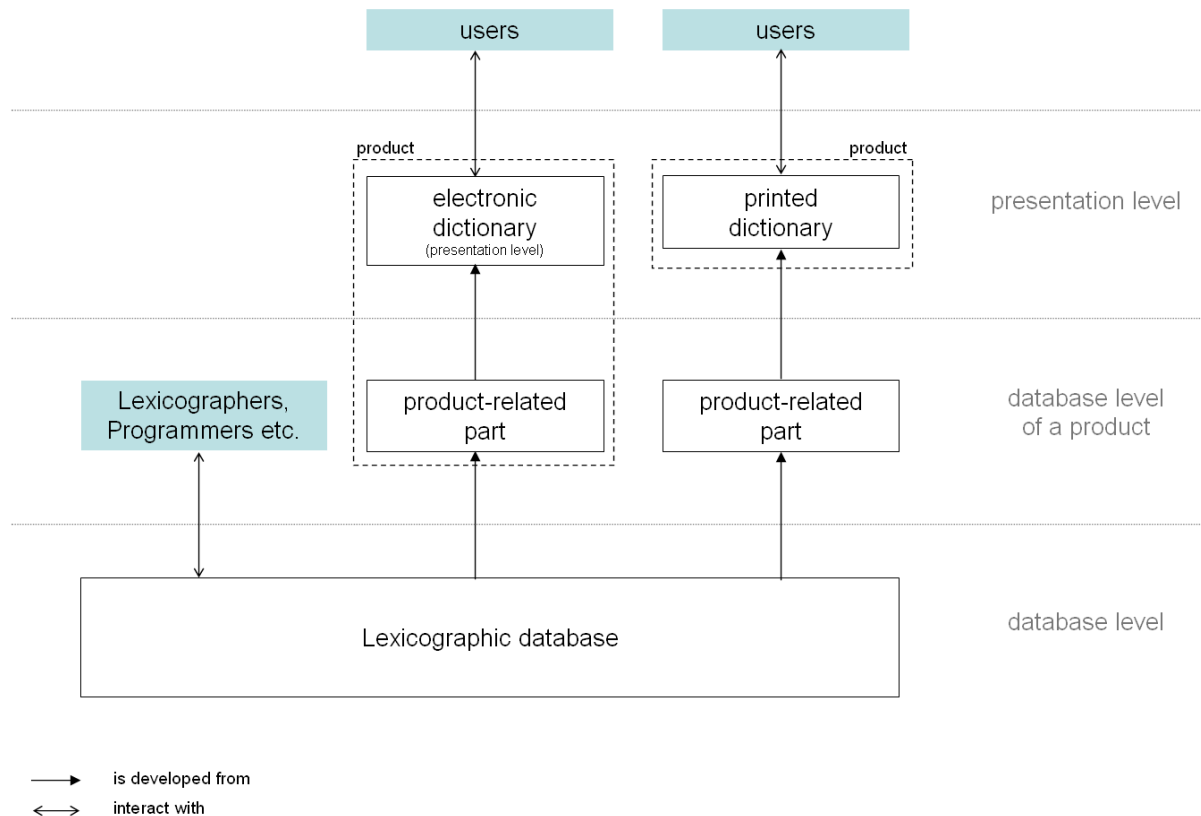


Figure 1: Basic publication model for developing printed and electronic dictionaries from a lexicographic database

The basic level is the level of the lexicographic database. Here, all lexicographic data, including outer texts, illustrations, etc., are stored. This is the level at which lexicographers work. Typical tasks are to check an article out of the database, edit it, check it in again, check for inconsistencies, and so on.

Only a subset of the data of the lexicographic database may be included in a published dictionary. For example, out of a lexicographic database which includes the entire vocabulary of German, a dictionary is published with only the most frequent words. Therefore, the level of a product-related part of the lexicographic database is introduced in the publication model. The product-related part contains the lexicographic data of one dictionary.

The next level is the presentation level. In the case of a printed dictionary, publication is a linear process. The distribution of data in the printed dictionary, specification of its layout, and so on, must be done once. Thus, the presentation level and the level of the product in its entirety are equal. As mentioned above, the process is different in the case of EDs. The presentation level of the ED must not have a specific fixed form but may be a continually evolving result of user requests and views according to these requests. These views do not occur by chance but are predetermined by lexicographers and programmers (cf. section 3). Therefore, the level of

presentation – the pure view – is not the same as the level of the whole ED product. This differentiation is also illustrated in Figure 1. Thus, an ED is composed of both a presentation level and a database level.

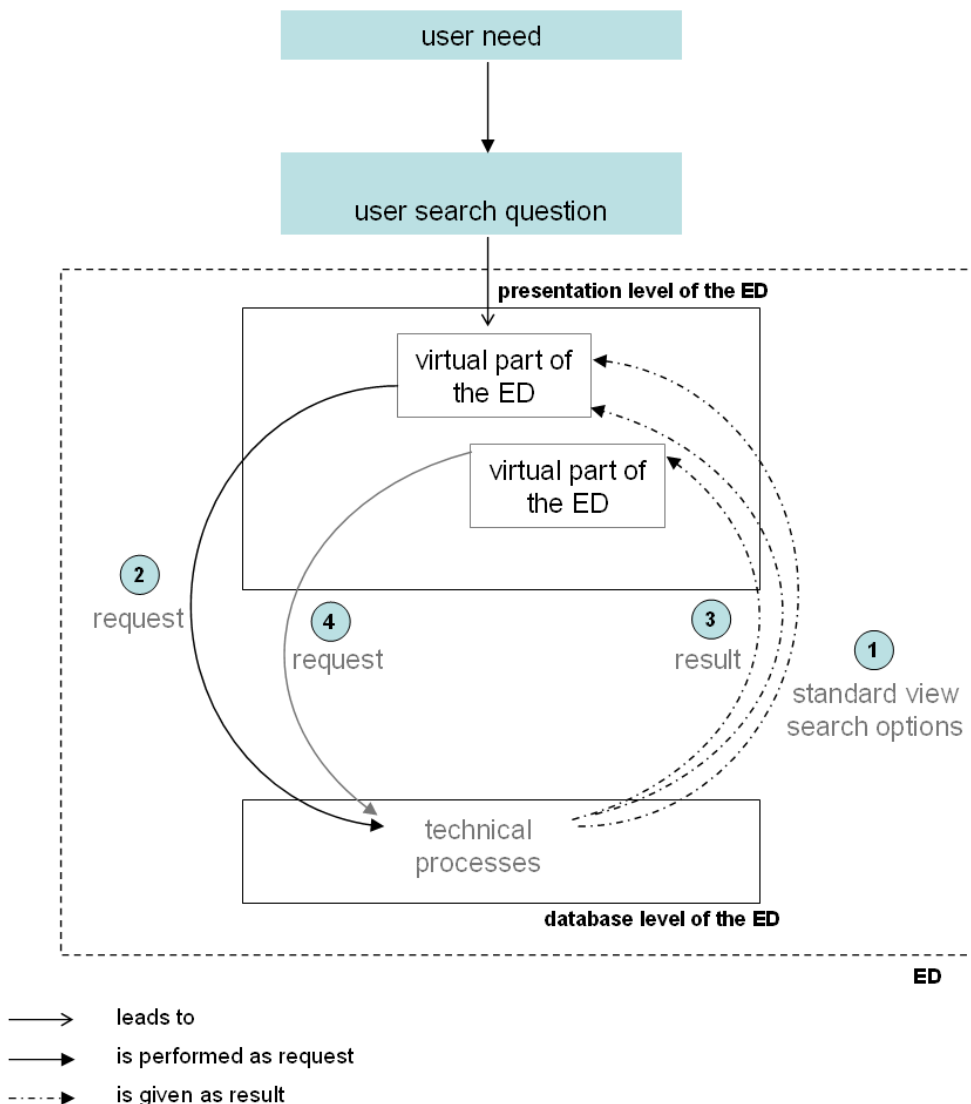


Figure 2: Focus on the ED product and the circular process of using it

The whole ED product must be viewed while ordering textual structures. Figure 2 focuses on this part of the entire publication model. Here, prominence is given to the process of interaction between users and EDs. The arrows illustrate the continual process of user requests and possible user-adapted views that are virtual parts of the ED. These virtual parts are not to be understood as distinct parts of the ED which share no common data, but as views of the ED which are not further defined. It is conceivable, for example, that a user might go back to the standard view before starting a new request.

As seen here, the use of an ED is a circular process, in which different levels play different roles, and in which one user action normally requires more than one technical process on the ED side. If the ED is considered here to be an interactive object, it must be borne in mind that the ED is acting as a representative of the humans who compiled the dictionary and its search options, possible views, etc. The computer itself is not an agent.

Therefore, any association between EDs and automatons should be avoided (cf. in contrast Tarp 2008: 121). Lexicographers and programmers must always pre-define possible requests and possible answers or views of the data. This begins with the concept of data modelling and structure which is the prerequisite for user-adapted views (cf. Müller-Spitzer 2008, Storrer 2001) and ends with the programming/coding of search options and the defining of possible displays of search results. Not one single link is displayed without a programme routine which tells the data to be displayed as such. And in the case of EDs compiled by humans and meant for human use, humans usually write these corresponding programme routines. One of the main tasks when creating an ED is to conceive and develop these routines.

## 4.2 A NOTIONAL ONLINE DICTIONARY AND ITS TEXTUAL STRUCTURES

Let us assume that there is a fictitious project on English monolingual lexicography. As a product, an English monolingual dictionary with special focus on semantics, especially on senses and synonyms of these senses, is to be derived from the lexicographic database, and presented in electronic form on the Internet.

In order to examine the textual structures of this dictionary more closely, the notional article “go” will be considered in more detail. The lexicographic items of this article are similar to the entry “go” in Wordsmyth ([www.wordsmyth.net](http://www.wordsmyth.net)), but this is not crucial here.

The data format chosen in this project is specified through an XML-DTD. The aim of the data model is to structure the lexicographic data so that it is as fine-grained as possible in order to have maximum flexibility in presenting the data. The tailor-made XML-DTD should elucidate the content-related value of the lexicographic data. Thus, the content-related value is machine-readable and interpretable by the computer (cf. Müller-Spitzer 2008a). Figure 3 shows the XML instance of the entry “go” in the form in which it is stored in the product-related part of the lexicographic database. All XML-structures are hierarchical tree structures. Therefore, the structure and the content could also be represented as a structure graph (cf. Figure 4).

Those who are familiar with Wiegand’s microstructural graphs (cf. Wiegand: 1989a, b) will notice the similarity. The graph in Figure 4 illustrates – like Wiegand’s graphs – the abstract structure together with the concrete structure. Also, all parts of the lexicographic data are exhaustively segmented (cf. Wiegand: 1989a, 440ff.). This kind of representation seems adequate for this structure rather than merging the concrete and the abstract structures together (cf. Ide/Kilgarriff: 2000). Related to the model of the whole dictionary and without the concrete data, the abstract structure grammar (cf. the term “Mikrostrukturengrammatik” – microstructure grammar – by Storrer: 1996, 247) could also be represented as a hierarchical tree. In the case of XML, an abstract structure grammar would be defined as an XML-DTD or XML-schema.

However, is the structure represented in this graph a microstructure? Microstructures are, according to Wiegand, the result of the method of (exhaustive) functional-positional segmentation of articles in printed dictionaries. They provide information about the content-related composition of the dictionary article as well as the order of the items in the entry. As well as lexicographic items, they also contain non-typographical indicators such as commas, numbers, etc. This is missing in this structure because such indicators would be generated automatically for the presentation of content. The data here is also in a special order, but this order says *nothing* about the arrangement of the data at the presentation level of the ED. The layout and, therefore, the order can be defined separately. In the first instance, this structure is the result of a purely functional segmentation of lexicographic content which presents the types of lexicographic content in a machine-readable way. Therefore, this structure should be introduced as *content structure* (for more details, cf. Müller-Spitzer 2006).

```

<entry>
  <lemma id="243599">go</lemma>
  <part-of-speech type="intransitive-verb"/>
  <inflected-forms>
    <inflected-form>went</inflected-form>
    <inflected-form>gone</inflected-form>
    <inflected-form>going</inflected-form>
    <inflected-form>goes</inflected-form>
  </inflected-forms>

  <sense id="1">
    <paraphrase>to proceed; travel</paraphrase>
    <synonyms>
      <synonym refid="12478.1">pass</synonym>
      <synonym refid="34598.3">move</synonym>
      <synonym refid="25346.1">advance</synonym>
      <synonym refid="65475.3">travel</synonym>
      <synonym refid="36546.1">proceed</synonym>
    </synonyms>
  </sense><sense id="2">
    <paraphrase>to move away from a place; leave</paraphrase>
    <synonyms>
      <synonym refid="453637.1">depart</synonym>
      <synonym refid="546474.1">leave</synonym>
    </synonyms>
  </sense><sense id="3">
    <paraphrase>to extend from one point to another</paraphrase>
    <synonyms>
      <synonym refid="453637.3">reach</synonym>
      <synonym refid="125674.1">extend</synonym>
      <synonym refid="987565.1">stretch</synonym>
      <synonym refid="456374.10">run</synonym>
    </synonyms>
  </sense>
</entry>

```

Figure 3: "go" as XML instance

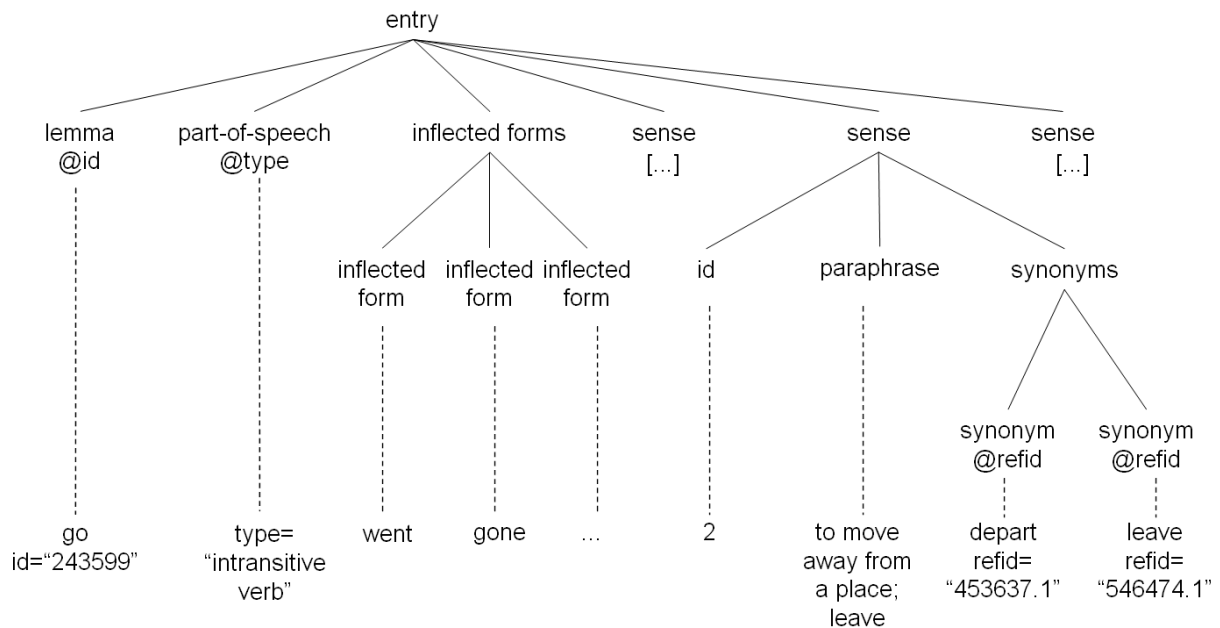


Figure 4: Content structure of "go" as graph



On closer inspection, it can be seen that some parts of the entry contain cross-references to other entries or parts of entries. The “idref”-attributes of the synonym-elements serve as connectors to the target entry and the appropriate sense. These are elements of the *cross-reference-structure*.

The main address element is the lemma, in this case “go”. All lexicographic items relate to this item and are addressed to it. These main address elements are able to build the main order criterion on the lexicographic database and build the *main order structure*. Typically, in linguistic dictionaries, these are the lemmas or headwords. The elements of this structure build the content-related junctions of the lexicographic content.

As noted before, the order of the content structure and its presentation properties can be defined separately from the content itself. In our scenario here, three parts of the ED presentation will now be regarded in more detail in order to exemplify the difference between the content structure, the cross-reference structure, the main order structure, and the corresponding textual structures at the presentation level.

This notional ED provides different views of the lexicographic data according to different user needs and user experience. Firstly, for users who need to browse the dictionary, it provides the opportunity to do so via a headword list, if, for example, the chosen user need is, “I want to know as much as possible about the words and their synonyms” (cf. Bergenholtz/Gouws 2007: 256, cf. also Langanke 2001: 249), and if, in addition to this, the given background of experience is “I normally use printed dictionaries”. For this user case, lexicographers decided to show a headword list similar to printed dictionaries as the main access structure. Inflected forms are also included in this headword list in order to provide access to one entry from different positions in the alphabet. Thus, from the inflected form “went”, one cross-reference entry is generated with the content:

**went** past tense to ->go.

This entry may be positioned in the article stretch between “wench” and “west”. At the presentation level, “went” appears as a lemma or headword, and “go” as a cross-reference item with the cross-reference mediating item (the arrow). The user has to click on “go” and, thus, receives the information s/he is looking for. In the case of a printed dictionary and in printlexicographic research, “went” would be regarded as an element of the macrostructure, and “-> go” as an element of the mediostructure. This example illustrates that the same data unit may play different roles on the different levels: “went” on the database level is part of the content structure; on the presentation level in the example above, it is part of the macrostructure.

Secondly, this ED provides the opportunity to browse through the dictionary entry in another form, namely in the form of a graph. Here, the chosen user need was, “I want to know as much as possible about one word and its synonyms”, and the given background of experience was “I normally use electronic dictionaries and thesauri”. In this particular user case, the entries are shown as schematically illustrated in Figure 5.

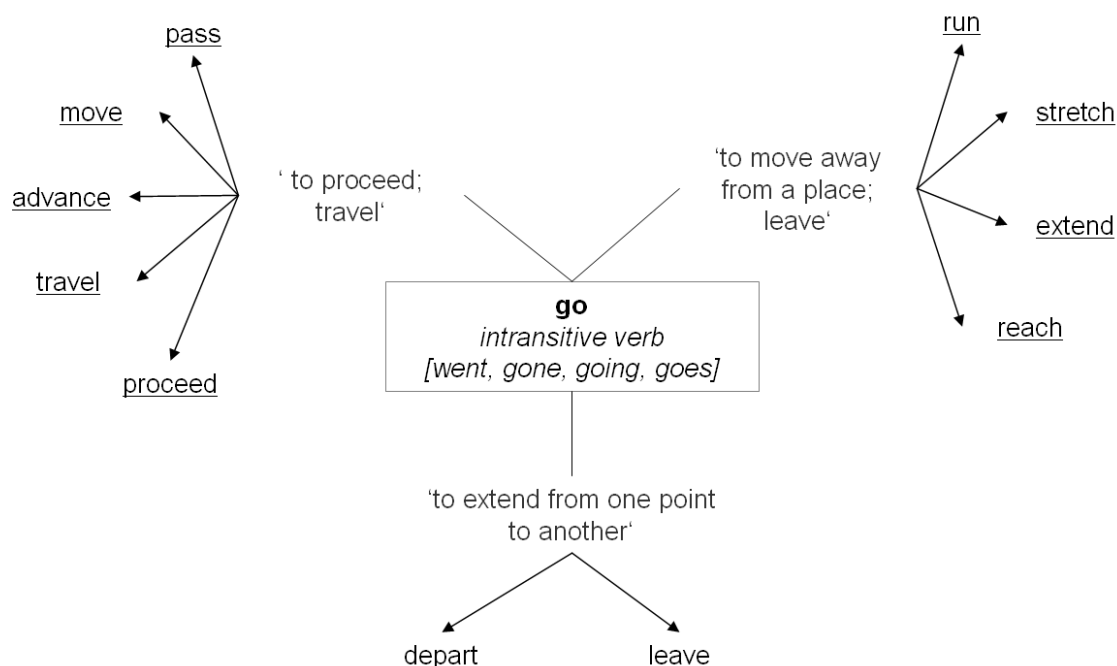


Figure 5: the entry “go” presented as a graphical view

“Went” is again a microstructural unit of the entry “go”. In contrast, all items of synonymy are mediostructural units because they are represented as links.

These different views of the lexicographic data illustrate that the textual structures of an ED are ordered in different ways on the database and presentation level. Terminological considerations must refer to both levels and must provide terminological tools of the trade for both levels.

#### 4.3 TEXTUAL STRUCTURES OF EDS: A PROVISIONAL SURVEY

In research on EDs, the terminology of textual structures in printed dictionaries is used in an unrevised way. “[...] There are so many theoretical articles that freely discuss various types of structure in electronic dictionaries as if they were the same as those found in printed dictionaries.” (Tarp: 2008, 102) This is problematic if the electronic dictionary has an innovative form of presentation and, as a result, different structural properties. However, it is not useful to break with a long tradition of terminology established in research. Therefore, at the presentation level of EDs, these established terms are useful because they are also related to the presentation level in printed dictionaries. However, it is only useful if the basic properties of these textual structures can also be found in the digital environment. They may be based on a partially modified definition or concept but should share basic properties. Against this background and for the reason that the structures in their finitude are not determinable, use of “the macrostructure” for EDs will be avoided here; instead, these elements will be termed “macrostructural units” and “mediostructural units” respectively. Very generally speaking, macrostructural units will be regarded here as the main access elements (cf. Wiegand: 2008, 214ff.) of dictionary entries that constitute the (content-related) order of the entries and that build the main address node. Microstructural units are regarded as parts of a dictionary entry, addressed (in a last instance) to the main access element (typically the lemma) and without cross-reference function, whereas

mediostructural units are regarded as parts of a dictionary entry with cross-reference function. Thereby, mediostructural units cannot be “easily disentangled from the outer and inner access structures” (de Schryver 2003: 180). This is the prototypical schema which could also be applied to EDs.

Further differences between the various structures in printed versus electronic dictionaries and their specific peculiarities cannot be described here for every kind of structural unit. Generally, every structural unit in an ED should be seen as a part of a virtual view of the product-related database and has, therefore, no firm typographic figure.

Three kinds of textual structure which belong to the database level have already been mentioned briefly: the main order structure, the content structure, and the cross-reference structure. If these textual structures are coded as XML, they are strictly hierarchical structures. If they are not, they can often be imagined as if they were hierarchically structured (cf. Wiegand’s structure graphs). Therefore, the assignment of elements to these textual structures is illustrated by means of hierarchical trees.

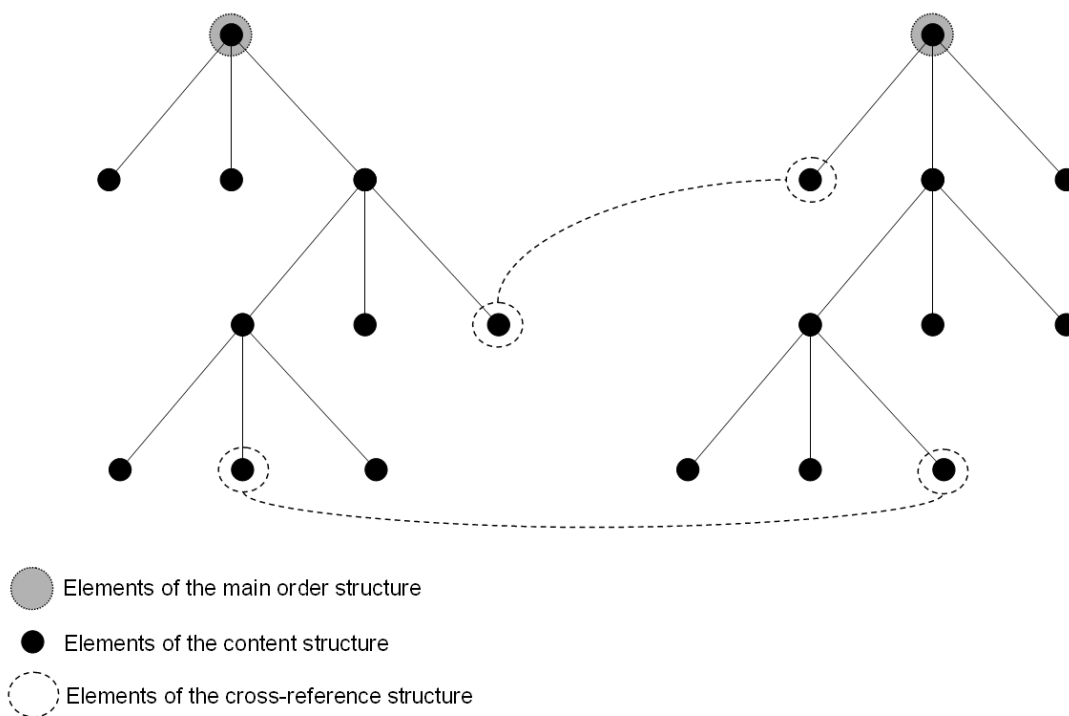


Figure 6: Assignment of textual structures on the database level illustrated by means of a hierarchical tree

On the database level, it is appropriate to speak of content *structure* because a database is typically a physically closed object; whereas on the presentation level, it seems to be more appropriate to speak of microstructural units because the *microstructure* might be difficult to grasp in its entirety. Typically, the elements of the main order structure function as macrostructural units on the presentation level, elements of the content structure as microstructural units, and elements of the cross-reference structure as mediostructural units. However, as demonstrated in the examples above, elements of the content structure may also function as mediostructural units in a special virtual view, elements of the cross-reference structure as microstructural units, and so on.

Textual structures which are crucial in the context of EDs and which have not yet been mentioned are the data distribution structure and the access structure. The distribution structure is according to Tarp “the central structure in both printed and electronic dictionaries [...], which determines the relations in terms of form and content [...]” (Tarp: 2008, 113) Against the background of the discussion above, the data distribution should be regarded more as a process than as a fixed structure in the context of EDs. The distribution of lexicographic data is a process which takes place on the way from the database level to the presentation level. This process starts again and again after every user request (cf. Figure 2). The distribution process determines the role which data units play on the presentation level (illustrated in Figure 7). Hence, defining these distribution processes according to user requests is the main task facing lexicographers and programmers working on the presentation concept of an ED.

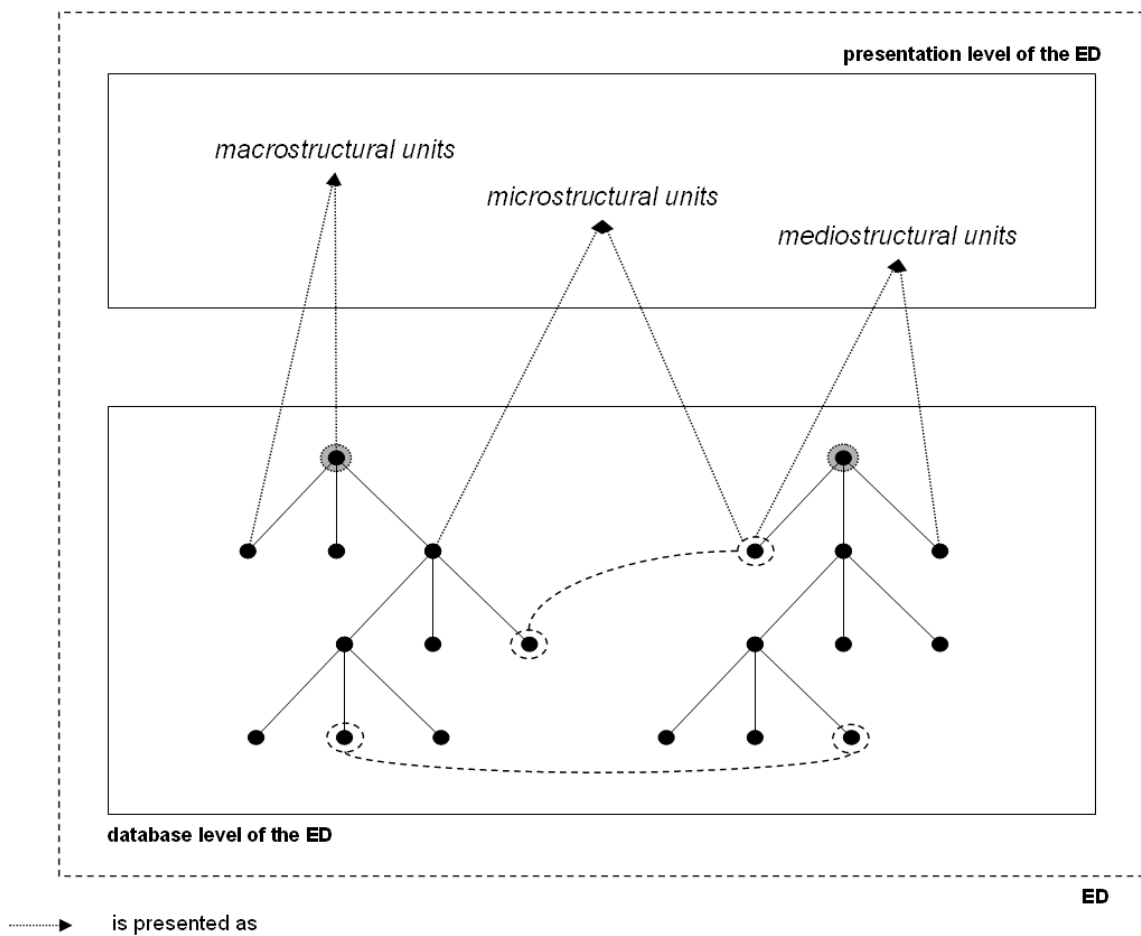


Figure 7: Illustration of the data distribution process according to user requests from the database level to the presentation level in EDs

“From a user’s perspective, the most innovative aspect of EDs is probably the retrieval system” (de Schryver: 2003, 146). Accordingly, access structures are very important for the presentation of EDs. In printlexicographic research, access structures are divided into outer and inner access structures (cf. article 5). Outer access structures provide users with access to the dictionary articles; inner access structures provide access to lexicographic items in the articles. There are many more properties or inner differentiations of access

structures in printed dictionaries which could be mentioned but play no central role here. In printed dictionaries, using access structures means turning pages. In EDs, there are typically two basic ways of accessing lexicographic data: the search options or the retrieval system and navigational elements, such as menus, headword lists, etc. Therefore, as a first step, access structures in EDs should be subdivided into those which are realised as search options and those which are realised in the form of navigational elements. The second step is to subdivide access structures in the form of search options into those which provide outer, outer and inner, or only inner access. One example of a search option which provides outer access is a search box in which a headword could be typed. An example of a search option which provides outer and inner access is a search for a headword with a special character string and a specific microstructural item, for example the item which classifies the lemma as an “intransitive verb”. An example of a search option which provides only inner access is a search only for specific microstructural items such as synonyms, information on word formation, etc. (for more examples, cf. de Schryver 2003: 173ff.)

The third step is to partition the search results into three groups, namely search results which contain only macrostructural units, search results which contain macrostructural and microstructural units, and finally, search results which contain only microstructural units. To illustrate this, if a search was carried out in the aforementioned dictionary for all synonyms of intransitive verbs, the search result for the entry “go” might have three different forms:

- (vi)            go
- (vii)          depart -> go
- (viii)         synonym: depart

In the first case, only the macrostructural unit, the lemma “go”, is given as a search result (although synonyms such as “pass”, “depart”, etc. are provided on condition that the article “go” is a relevant search result). In the second case, the microstructural unit “depart” as well as the macrostructural unit “go” is given. In the last case, only the microstructural unit “depart” appears.

Access structures in the form of navigation structures are not to be subdivided as strictly and formally as the search options, but rather according to their content. For example, navigation structures can provide access via a headword list, via semantic concepts such as word fields, etc. Thus, this subdivision has a more descriptive character. Currently, access structures in the form of search options are mostly presented as search boxes, pull-down-menus, etc. In the future, it is conceivable that access structures will become more and more suited to audio-input or touch screens (cf. de Schryver 2003: 173).

Typically, access structures are analysable in their entirety on the presentation level. As distinguished from the context of microstructural and macrostructural units, in the case of access structure, the term “structure” seems appropriate.

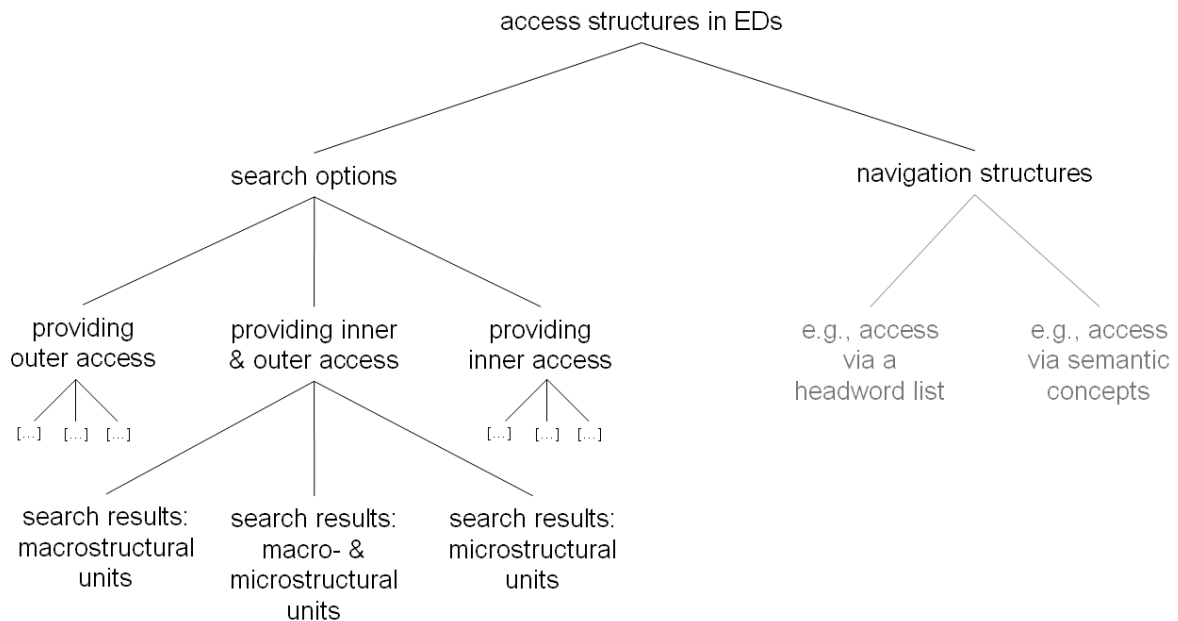


Figure 8: First overview of access structures in EDs

It makes sense to use established terms from printlexicographic research in the context of EDs if their basic properties apply to the textual structures in EDs. For identifying data units as microstructural units, macrostructural units, etc. with the prototypical properties mentioned above, the ED should be regarded as a kind of window to the relating database, the underlying data structure, and the distribution processes. Simply by scrutinising an ED through systematically performed user requests, it is possible to deduce which data units could be regarded as microstructural units, which as macrostructural units etc. Lexicographers and programmers determine which data is presented in which form in which cases. However, the user must either know the underlying database and corresponding programming procedures, or infer these through systematically performed requests. Tarp (2008: 111) shows what happens if the presentation level without the database is the only object of consideration. He takes as an example a search result in an electronic dictionary with the headline “nouns not occurring in the singular” with four nouns as search results. He interprets this search result as a “traditional article” where the headline takes the role as the “lemma” and the four nouns as “addressed data” building a “microstructure consisting of four alphabetically organised search fields” (Tarp 2008: 111). With this approach, the process of user requests, corresponding results, and the underlying ordering structures which determine these processes are not involved.

The headline of a search result which is regarded as a lemma in this example is far removed from a typical macrostructural element in printed dictionaries. Properties which are constitutive for regarding data units as macrostructural units with an ordering function related to the lexicographic content are non-existent here. Therefore, such a search result should be described as it occurs (a headline followed by four search results). Then, the underlying order of lexicographic data should be inferred from this and similar results as well as from the corresponding presentation of the dictionary articles. With reference to this example, it would be inferred that there are items on the use of singular/plural of nouns as microstructural units and lemmas as macrostructural units to which these microstructural units are addressed. The description of an ED purely on the presentation level neglects the ED in its entirety and provides insufficient insights into the structural composition of the ED. The aim must be to realise which order structures the lexicographic data have and how the process of user requests and results is controlled. Although the database-internal processes are invisible for external users, they can be inferred to a certain degree.

## 5 DESCRIBING AND ANALYSING ACTUAL EDs AND PLANNING NEW EDs

Textual structures play a significant role when describing actual EDs. The key aspects of analysing textual structures in EDs have already been discussed. Further relevant points which are useful for describing and analysing actual EDs and for planning new EDs are outlined below.

Firstly, the ED must be closed off from any other content of the website which is referenced via a URL/URI etc. On the Internet, many EDs are surrounded by other lexical resources, such as tools for exploring corpus data, parts of a higher-level/superordinated dictionary portal, etc. Beyond that, most websites contain a large number of advertisements. Consequently, while describing an ED, the ED-part of the website has to be identified.

Secondly, the access structures should be scrutinised in addition to the general properties of the dictionary. By using the search options as systematically as possible, analysing the navigation structure and following the search path up to the dictionary article, the best insight into the construction of the textual structures in the ED can be gained. As a result, the textual structures of the ED for both the database and the presentation level can be explained. It is worth noting that in order to make the description as precise and comprehensible as possible, a note must always be made of which user action was performed. This means that the individual search or access path (cf. Wiegand 2008a: 211) must always be referenced. The search path is – unlike the access structures – an individual empirical phenomenon. Search paths are topics in research into printed dictionaries, but they are not as essential as in EDs. In printed dictionaries, different search paths lead to the same lexicographic data in the end. In EDs, potentially the virtual view might differ from one search path to another. However, any individual search path should always involve the same virtual part of the dictionary; otherwise, the ED is badly preprogrammed. To sum up, the crucial point is to bear in mind that the presentation level of the ED is only one visible layer of a product which consists of two levels with many processes interacting between them.

The same applies for planning new EDs. Here, one must consider that the construction of textual structures on the database level builds the essential conditions for developing and presenting textual structures on the presentation level, for example, search options. The more granular the modelling and structuring of the content structure, the more flexible user-adapted views can be, and the more advanced search options with special tailor-made search results can be provided for the user. A related observation is that, interestingly, the two “opposing theories” (cf. Bergenholtz/Tarp: 2003) – the modern function theory by Bergenholtz/Tarp et al. and the theory of lexicographic texts by Wiegand – may both play an essential role in the planning of new EDs. On the one hand, the theory of lexicographic texts provides very detailed insights into lexicographic structures. For modelling lexicographic data in as fine-granular a way as possible, it is fruitful to look at these works, even though they have to be stripped of all elements which belong to the presentation of a printed dictionary. However, in order to underpin the guidelines for modelling lexicographic data, it is useful to take into account these exact examinations and to use aspects of them (cf. for the modelling of content and cross-reference structures Müller-Spitzer 2006, 2007a and b). In the context of modelling lexicographic data, the focus must lie on multifunctional and user-independent structuring. The aim should be to build the basis for developing different views from one lexicographic database according to different user needs and usage situations. Notwithstanding this, lexicographic projects must adapt themselves to certain functions. However, in the context of modelling lexicographic data, the line of argument that the theory of lexicographic texts is too linguistic is unsound. Here, the key aspect is to have a clear and standard description basis for the database level, (initially) independent of the role which the lexicographic data may play on the presentation level. On the other hand, function theory contains essential aspects for planning and developing user-adapted views. This division is emphasised by Tarp also: „Naturally, lexicographical data can and should also be divided phenomenologically according to its specific content. If this is done, such data is related directly to the information categories [...]. However, it is necessary to underline that the category to which specific data

belongs says nothing (and can say nothing) about the specific purpose of this data in a given dictionary, since this purpose depends on the function or functions that this dictionary is designed to have.” (Tarp 2008: 100) For EDs, it would be better to replace “the specific purpose of this data in a given dictionary” with ‘specific purpose of this data in a specific usage situation and in a specific virtual view of the electronic dictionary’. Thus, the multifunctional modelled data on database level fulfils different functions in the process of using an ED. This division should be borne in mind while planning new EDs. Another crucial point about function theory is that it shifts the emphasis onto potential users and the social situations in which they participate. This is important for developing very innovative EDs “which will target the specific needs of specific users in specific social situations” (Fuertes-Olivera 2009: 99).

## 6 CLOSING REMARKS

The aim of this article is to provide a short general survey of textual structures in electronic dictionaries, especially in relation to textual structures in printed dictionaries. Thereby, the focus is more on the potential complexity which textual structures in EDs may have than on the properties of textual structures of actual EDs. The key aspect for analysing textual structures in EDs is the introduction of two levels: the database level and the presentation level. Textual structures in EDs “exist” on both levels. Therefore, the according terminology and order must be assigned to these levels. Another crucial point is that the act of using an ED should be regarded as a circular process which creates virtual views according to user requests. This model requires a intensive analysis of the relationship between textual structures in printed dictionaries and textual structures in electronic dictionaries. The main order structure, content structure, and cross-reference structure are proposed as partial textual structures on the database level of EDs; access structures, macrostructural, microstructural or mediostructural units are the chosen terms for the presentation level, based on established terms of printlexicographic research. Hopefully, more and more EDs will provide different user-adapted virtual views of one database if this is appropriate for the needs of the users. If more EDs make the possibility of presenting user-adapted views a reality, it is even more obvious that both the database level and the presentation level are crucial for analysing textual structures in EDs.

## REFERENCES

- Andersen, B./Nielsen, S. (2009): Ten Key Issues in Lexicography for the Future. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.), 355-366.
- Beißwenger, M. (2002): DIE DATENDISTRIBUTION IM DE GRUYTER WÖRTERBUCH DEUTSCH ALS FREMDSPRACHE. In: Wiegand, H.E. (ed.): Perspektiven der pädagogischen Lexikographie des Deutschen II. Untersuchungen anhand des "de Gruyter Wörterbuchs Deutsch als Fremdsprache". Tübingen (Lexicographica. Series Maior 110), 445-470.
- Bergenholtz, H. /Tarp, S./Wiegand, H. E. (1999): Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. In: Hoffman, L./Kalverkämper, H./ Wiegand, H.E. (eds.): Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft. 2. Halbband. Berlin/New York, 1762-1832.
- Bergenholtz, H./Tarp, S. (2003): Two opposing theories: On H.E. Wiegand's recent discovery of lexicographic functions. In: Hermes 31, 171-196.



- Bergenholtz, H./Gouws, R. (2007): The Access Process in Dictionaries for Fixed Expressions. In: *Lexicographica* 23, 237-260.
- Bergenholtz, H./Nielsen, S./Tarp, S. (eds.) (2009): *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Frankfurt a.M./Berlin/Bern/Bruxelles/NewYork/Oxford/Wien.
- Bucher, H.J. (2004): Das Netzwerk-Medium: Wie das Internet die Kommunikation verändert. In: *Geographie und Schule* 147, 2-7.
- De Schryver, G.-M. (2003): Lexicographers' Dreams in the Electronic-Dictionary Age. In: *International Journal of Lexicography* 16.2, 143-199.
- Fuertes-Olivera, P. A. (2009): The Function Theory of Lexicography and Electronic Dictionaries: Wiktionary as a Prototype of Collective Free Multiple-Language Internet Dictionary. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.), 99-134.
- Geeraerts, D. (2000): Adding Electronic Value. The electronic version of the Grote Van Dale. In: Heid, U./Evert, S./Lehmann, E./Rohrer, C. (eds.): *Proceedings of the Ninth EURALEX International Congress, Stuttgart, Germany, August 8th - 12th, 2000*. Stuttgart: Universität Stuttgart, Institut für Maschinelle Sprachverarbeitung, 75-84
- Ide, N./Kilgariff, A. (2000): A Formal Model of Dictionary Structure and Content. <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.40.9724>>.
- Langanke, U. (2001): Dialogische Mikrostrukturen in hypermedialen Online-Wörterbüchern: Ein neues Verhältnis zwischen Lexikograph, Wörterbuch, Benutzer. In: *Jahrbuch der ungarischen Germanistik*, 243-255.
- Lehr, A. (1996): Zur neuen *Lexicographica*-Rubrik "Electronic Dictionaries". In: *Lexicographica* 12, 310-317.
- Lew, R. (2009): Towards Variable Function-Dependent Sense Ordering in Future Dictionaries. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.), 237-264. Müller-Spitzer, C. (2003): Ordnende Betrachtungen zu elektronischen Wörterbüchern und lexikographischen Prozessen. In: *Lexicographica* 19, 140-168.
- Müller-Spitzer, C. (2006): Das Konzept der Inhaltsstruktur: Eine Auseinandersetzung mit dem Konzept der Mikrostrukturen im Kontext der Modellierung einer lexikografischen Datenbasis. In: *Lexicographica* 22, 72-94.
- Müller-Spitzer, Carolin (2007): Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung. In: *Hermes* 38, 137-171.
- Müller-Spitzer, C. (2007): *Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis*. Tübingen.
- Müller-Spitzer, C. (2008a): The Lexicographic Portal of the IDS: Connecting Heterogeneous Lexicographic Resources by a Consistent Concept of Data Modelling. In: Bernal, El./DeCesaris, J. (eds.): *Proceedings of the Thirteenth EURALEX International Congress, Barcelona, Spain, July 15th - 19th, 2008*. Barcelona: Universitat Pompeu Fabra, Institut Universitari de Lingüística Aplicada (CD/DVD), 457-461.
- Müller-Spitzer, C. (2008b): Research on Dictionary Use and the Development of User-Adapted Views. In: Storrer, A./Geyken, A./Siebert, A./Würzner, K.-M. (eds.): *Text Resources and Lexical Knowledge. Selected Papers from the ninth Conference on Natural Language Processing KONVENS 2008, Berlin*. Berlin/New York, 223-238.
- Nielsen, S. (1999): Mediostructures in Bilingual LSP Dictionaries. In: *Lexicographica* 15, 90-113.

- Pajzs, J. (2009): On the Possibility of Creating Multifunctional Lexicographical Databases. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.), 327-354.
- Storrer, A. (1996): Metalexikographische Methoden in der Computerlexikographie. In: Wiegand, H.E. (ed.): Wörterbücher in der Diskussion II, Tübingen (Lexicographica. Series Maior 70), 239-255.
- Storrer, A. (2001): Digitale Wörterbücher als Hypertexte: Zur Nutzung des Hypertextkonzepts in der Lexikographie. In: Lemberg, I./Schröder, B./Storrer, A. (eds.): Chancen und Perspektiven computergestützter Lexikographie: Hypertext, Internet und SGML/XML für die Produktion und Publikation digitaler Wörterbücher. Tübingen (Lexicographica. Series Maior 107), 53-69.
- Tarp, S. (1999): Theoretical foundations of the so-called cross-reference structures. In: Lexicographica 15, 114-137.
- Tarp, S. (2008): Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner's Lexicography. Tübingen: (Lexicographica. Series Maior 134).
- Tarp, S. (2009): Beyond Lexicography: New Visions and Challenges in the Information Age. In: Bergenholtz, H./Nielsen, S./Tarp, S. (eds.), 17-32.
- Welker, H.A. (2009): An overview of Wiegand's metalexicographic works.  
<<http://www.let.unb.br/hawelker/Wiegand.pdf>>.
- Wiegand, H.E. (1989a): Der Begriff der Mikrostruktur: Geschichte, Probleme, Perspektiven. In: Hausmann, F.-J./Reichmann, O./Wiegand, H.E./Zgusta, L.: Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1). Berlin/New York 1989, 409-462.
- Wiegand, H.E. (1989b): Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. In: Hausmann, F.-J./Reichmann, O./Wiegand, H.E./Zgusta, L.: Wörterbücher. Ein internationales Handbuch zur Lexikographie. 1. Teilband (= Handbücher zur Sprach- und Kommunikationswissenschaft 5.1). Berlin/New York 1989, 371-409
- Wiegand, H.E. (1990): Printed dictionaries and Their Parts as Text. An Overview of More Recent Research as an Introduction. In: Lexicographica 6 (1990), pp. 1-126.
- Wiegand, H.E. (1998a): Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie, 1. Teilband, Berlin/New York.
- Wiegand, H.E. (1998b): Altes und Neues zur Makrostruktur alphabetischer Printwörterbücher. In: Wiegand, H.E. (ed.): Wörterbücher in der Diskussion III. Vorträge aus dem Heidelberger Lexikographischen Kolloquium. Tübingen (Lexicographica. Series Maior 84), 348-372.
- Wiegand, H.E. (2000): ÜBER SUCHBEREICHE, SUCHZONEN UND IHRE TEXTUELLEN STRUKTUREN IN PRINTWÖRTERBÜCHERN. In: Wiegand, H.E. (ed.): WÖRTERBÜCHER IN DER DISKUSSION. IV. VORTRÄGE AUS DEM HEIDELBERGER LEXIKOGRAPHISCHEN KOLLOQUIUM. Tübingen (Lexicographica. Series Maior 100), 233-301.
- Wiegand, H.E. (2002): Altes und Neues zur Mediostruktur in Printwörterbüchern. In: Lexicographica 17, 217-248.

Wiegand, H.E. (2005): Über die Datenakzessivität in Printwörterbüchern. Einblicke in neuere Entwicklungen einer Theorie der Wörterbuchform. In: *Lexikos* 15, 1-35.

Wiegand, H.E. (2006): Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. In: *Lexicographica* 22, 187-261. [appeared 2007]

Wiegand, H.E. (2007): Über Zugriffspfade in Printwörterbüchern. Ein Beitrag zur Schnittstelle von Benutzungshandlungen und Wörterbuchform. In: *Lexikos* 17, 1-32.

Wiegand, H.E. (2008): Zugriffsstrukturen in Printwörterbüchern. Ein zusammenfassender Beitrag zu einem zentralen Ausschnitt einer Theorie der Wörterbuchform. In: *Lexicographica* 24, 209-315.