

POSTPRINT

# The consistency of sense-related items in dictionaries

Current status, proposals for modelling  
and applications in lexicographic practice

Carolin Müller-Spitzer

Consistency of reference structures is an important issue in lexicography and dictionary research, especially with respect to information on sense-related items. In this paper, the systematic challenges of this area (e.g. 'non-reversed reference', bidirectional linking being realised as unidirectional structures) will be outlined, and the problems which can be caused by these challenges for both lexicographers and dictionary users will be discussed. The paper also discusses how text-technological solutions may help to provide support for the consistency of sense-related pairings during the process of compiling a dictionary.

## 1. Introduction

This paper deals with problems of consistency of sense-related items in dictionaries. It focuses on paradigmatic structures with regard to some lexicographic issues and consists of four main parts. Firstly, the term *consistency* will be defined in order to determine more precisely its meaning in this context. Secondly, a short overview of the current status of the consistency of sense-related items in some German dictionaries such as the *Duden Synonymwörterbuch* (2007) or *eleXiko* (2003ff.) will be given. Thirdly, in the main section, a concept of data-modelling for reference structures will be presented. This concept for modelling reference structures is a way of supporting consistency in general lexicographic practice. Finally, some possible applications of the proposed model will be outlined. These concern, for example, new kinds of presentation of sense-related items, provided the underlying data are structured consistently and precisely. In this paper, the emphasis is placed on lexicographic practice and text-technological questions rather than on theoretical lexicological considerations.

## 2. The term consistency

As a first step, the term **consistency** needs to be defined. **Consistency** is derived from the Latin word *consistere*, meaning 'to stand, hold or continue'. From this rather vague and broad meaning, a number of other senses have developed. One of them is a special meaning in the tradition of Aristotelian logic:

In logic, a consistent theory is one that does not contain a contradiction. The lack of contradiction can be defined in either semantic or syntactic terms. The semantic definition states that a theory is consistent if it has a model; this is the sense used in traditional Aristotelian logic, although in contemporary mathematical logic the term *satisfiable* is used instead. The syntactic definition states that a theory is consistent if there is no formula *P* such that both *P* and its negation are provable from the axioms of the theory under its associated deductive system.

(Quoted from [www.wikipedia.com](http://www.wikipedia.com))<sup>1</sup>

More generally, it can be summarised as 'something characterised by a lack of contradiction'. In the context of databases, consistency refers to the lack of contradiction of data, controlled by constraints. Therefore, consistency is well defined in this special field. In addition to this, **consistency** has different meanings in everyday language. One of these meanings is influenced by the technical meaning described above. Merriam-Webster paraphrases this sense as "agreement or harmony of parts or features to one another or a whole" [Merriam Webster Online 2009]. For the purposes of the present paper, I will refer to the term **consistency** in two different ways: in the context of data modelling, the more technical sense is adopted, whereas in the context of dictionary usage, it is the more general meaning of consistency which is intended.

## 3. Consistency of sense-related items in dictionaries: A short insight into the current status

In this section, I shall outline the situation we are currently facing in terms of consistency of sense-related items in existing German dictionaries. In the context of describing lexical-semantic relations in dictionaries, consistency has to mean that bidirectional relations, as existing in paradigmatic relations, are always given for both reference points between which a specific relation holds. For example, if *require* is given as a synonym of the entry *demand*, then *demand*

---

1. For a similar definition see also the entry "classical logic" in the Stanford Encyclopedia of Philosophy.

<p><b>arbeitsunfähig</b>          bettlägerig, krank, unpässlich; (<i>bildungsspr.</i>): indisponiert; (<i>oft emotional</i>): malade.</p>
<p><b>dienstunfähig</b>          arbeitsunfähig, erwerbsunfähig, invalide, krank, nicht arbeitsfähig, nicht dienstfähig, nicht einsetzbar, untauglich.</p>
<p><b>erwerbsunfähig</b>          arbeitsunfähig, behindert, dienstunfähig, invalide; (<i>Amtsspr.</i>): schwerbehindert, schwerbeschädigt</p>

Figure 1. Entries *arbeitsunfähig*, *erwerbsunfähig* and *dienstunfähig* from Duden: Das Synonymwörterbuch (2007)

should also be listed as a synonym of the entry *require*. This is a form of consistency which is important for the underlying lexicographic data model as well as for the dictionary user.

In the following, two conventional German dictionaries of synonyms (Duden 2007 and Wahrig 2006) and *elexiko*, a new corpus-based dictionary of contemporary German which is being compiled at the IDS, have been chosen in order to illustrate the current situation.

In Figure 1, three entries taken from Duden (2007), a dictionary of synonyms, are shown: *arbeitsunfähig* (*unfit or unable to work*), *erwerbsunfähig* (*unable to work, incapacitated*) and *dienstunfähig* (*disabled, unfit for service*). The descriptions of these three entries are semantically very close. The terms constitute a synonymous set or cluster. Nevertheless, there are striking inconsistencies. For example, in the entry *arbeitsunfähig*, the synonym *erwerbsunfähig* is missing, although *arbeitsunfähig* is given as a synonym of the headword *erwerbsunfähig*. In addition, *dienstunfähig* is not listed as a meaning equivalent to *arbeitsunfähig*, whereas in the entry *dienstunfähig* both *arbeitsunfähig* and *erwerbsunfähig* are listed as synonyms. It could be argued that consistency is not of particular importance here. Presumably, most lexicographers attempting to compile a reference work of synonymy aim to provide an abundance of words with similar meanings which can be substituted for each other. Their aim is not to depict theoretical lexical-semantic structures as lexicographic information (cf. also Lew 2007). However, it is argued here that, as the entry *arbeitsunfähig* in particular illustrates, a more consistent approach would help to provide the

<b>Brennpunkt</b> Mittelpunkt, Kern, Herz, Herzstück, Zentrum, Knotenpunkt, Schnittpunkt, Zentral- punkt, Sammelpunkt, Puls, Nabel (der Welt), Drehpunkt, Achse, Pol, Mitte, Hochburg, Tum- melplatz, Seele, Herd Ursprung, Nährboden	<b>Knotenpunkt</b> Mittelpunkt, Schnittpunkt, Kreuzungs- punkt, Drehpunkt, Kon- vergenzpunkt, Sammel- punkt, Brennpunkt, Kern, Schwerpunkt, Mitte, Zen- trum, Herz, Herzstück, Nabel (der Welt), Tum- melplatz, Center, Hoch- burg, Achse, Pol, Puls, Seele, Scheitel	<b>Mittelpunkt</b> Zentrum, Herz, Kern, Mitte, Herz- stück, Center, Hochburg, Nabel, Achse, Dreh- und Angelpunkt, Brenn- punkt, Knotenpunkt, Schnittpunkt, Sammel- punkt, Schwerpunkt, Zentralpunkt, Zentrale, Seele, Kerngebiet, Scheitel, Kreuzung, Kreuzungspunkt, Sammelbecken
--	---	--

Figure 2. Entries *Mittelpunkt* (centre), *Brennpunkt* (focus) and *Knotenpunkt* (junction) from Wahrig Synonymwörterbuch (2006)

dictionary user with better information. Presumably, any lexicographer would have added *erwerbsunfähig* as a synonym of *arbeitsunfähig* to this dictionary, had the incomplete listing been noticed.

Figure 2 shows an example taken from a different reference work, the “Wahrig Synonymwörterbuch” (2006). Again, as the example demonstrates, the three entries are semantically very similar to each other: *Mittelpunkt* (centre), *Brennpunkt* (focus) and *Knotenpunkt* (junction). As Cruse (2004: 156) points out, generally no one is puzzled by the content of dictionaries of synonyms. Certainly, at a closer look, most native speakers of German would be surprised by many of the synonyms which are listed. However, it is not synonymous relations as such which are the focus of attention here.

Turning again to the problem of consistency, the following questions arise: Why is only *Herzstück* (heart, core) listed as an equivalent of *Mittelpunkt* (centre), whereas in the entries *Brennpunkt* (focus) and *Knotenpunkt* (junction), both *Herz* (heart) and *Herzstück* are mentioned? Or why does the entry of *Mittelpunkt* only contain *Nabel* (navel, hub), whereas *Nabel der Welt* (hub of the world) is listed with *Brennpunkt* and *Knotenpunkt*? These are just two examples which pinpoint some of the issues concerning the problem of consistency. They are a good illustration of how more support in terms of general consistency would improve lexicographic results.

Elexiko, which is an online resource of contemporary German, serves as a last example to illustrate a more complex and less obvious case of inconsistency. As mentioned before, elexiko is still in the process of being written (elexiko 2003 et. seq.), which means that this dictionary is not a complete reference book following

**ehemalig** Lesart: **früher**  
 Lesartenübergreifende Angaben

Lesart: früher

Bedeutungserläuterung    Semantische Umgebung u. lexikalische Mitspieler    Typische Verwendungen    Sinnverwandte Wörter    Besonderheiten des Gebrauchs    Grammatik

**Sinnverwandte Wörter** ⓘ

Beziehung(en) der Bedeutungsähnlichkeit/-äquivalenz

ⓘ Synonym(e)

damalig	<a href="#">Beleg(e)</a>
einstig	<a href="#">Beleg(e)</a>
früher	<a href="#">Beleg(e)</a>

Beziehung(en) des Bedeutungsgegensatzes

ⓘ komplementäre(r) Partner:

aktuell	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
derzeitig	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
gegenwärtig	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
heutig	<a href="#">Beleg(e)</a>	<input type="checkbox"/> <a href="#">Beleg(e)</a>
jetztig	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
		<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
künftig	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
neu		<a href="#">Beleg(e)</a>
zukünftig	<a href="#">Beleg(e)</a>	<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>
		<input checked="" type="checkbox"/> <a href="#">Kommentar(e)</a>

Figure 3. Extract from the dictionary section “sense-related items” of the entry *ehemalig* in *exlexiko*

an alphabetical compiling procedure.<sup>2</sup> In the presentation of *exlexiko*, each entry is divided into two main parts. The first part gives sense-independent information of a lexeme (in German “lesartenübergreifende Angaben”), and the second part provides information which is bound to a specific sense of the search item (in German “lesartenbezogene Angaben”, cf. Storjohann 2005:62). Sense-independent details focus on information that applies to the entire search word and not to a specific sense, and it contains details of spelling, spelling variants and syllabication, morphological information, and so on. Information on a specific sense can be accessed by clicking on ‘signposts’, which are guidewords used for identifying senses (cf. Storjohann 2005:68) (see, for example, the entry *ehemalig* and within that entry the sense ‘früher’ in Figure 3). The sense-related items are presented in a special dictionary section (or tab) headed “sinnverwandte Wörter” (“sense-related items”).

If, for instance, the link to the synonym *einstig* is followed, the link target is the corresponding sense labelled ‘ehemalig’ which holds the synonymy relation (see Figure 4).

2. *Exlexiko* can be accessed free of charge and without registration.

**einstig** Lesart *ehemalig*  
 Lesartenübergreifende Angaben

Lesart *ehemalig*

Bedeutungserläuterung	Semantische Umgebung u. lexikalische Mitspieler	Typische Verwendungen	Sinnverwandte Wörter	Besonderheiten des Gebrauchs	Grammatik
-----------------------	---	-----------------------	----------------------	------------------------------	-----------

**Bedeutungserläuterung**

Mit **einstig** wird auf einen Zeitraum oder Zeitpunkt verwiesen, der zum Sprechzeitpunkt (in der Vergangenheit liegt) Auf diesen Zeitabschnitt wird verwiesen, wenn beispielsweise Personen(gruppen), Sachverhalte und Orte früher einen von heute abweichenden Status bzw. eine von heute abweichende Funktion besessen haben

 Wortklasse Temporales Deliktikum  Belegte

Figure 4. Extract of the entry *einstig* in *exlexiko* (Tab: paraphrase) (link target)

This means that it is not the dictionary part which contains general sense-independent information which is opened when a user follows a link, but it is the corresponding paraphrase of a specific sense to which the user is guided (as the first part of the sense-related information).

However, when we go back to the entry *ehemalig* and look up another synonym, for instance *früher*, we notice that the link target is the entry *früher* as a whole, but it does not refer to a specific sense as such.<sup>3</sup> In the context of *exlexiko*, this is incorrect because all sense-related items are to be given specifically for a semantic instantiation or context as displayed by a sense or subsense respectively. Presumably, the linking mistake occurred because the entry *ehemalig* was edited first, before *früher*, and the lexicographer had no means of systematically checking each item of the same semantic field for its link reference. For this reason, the link to the entry *früher* could only be given on the sense-independent level at that time because, unless a headword has been fully analysed and described, each link to another term is always given without a context and thus sense-independently. As a result, the lexicographer simply forgot to add the necessary details on link reference, from general to sense-bound information, a mistake that is bound to happen at some point when no computer support is available

3. After the colloquium, the entry *ehemalig* was revised so that the relation between *ehemalig* and *früher* is now related to specific senses in both directions.

(cf. Blumenthal et al. 1988: 365; Engelberg/Lemnitzer 2001: 211). It is simply impossible for lexicographers to bear all specific references in mind. This explains why it is essential to have computer assistance at hand.

Turning our attention to the entry *früher*, this is a case where in the reverse direction, the synonym relation is connected to one specific sense. Opening the sense *ehemalig* in the head word *früher* and following the link to *ehemalig*, the corresponding paraphrase is again already open to the user. This is also a form of inconsistency: whereas in one direction the synonym relation is recorded as sense-related, in the other case it is not.

These examples show clearly that, generally, entries in various German dictionaries are organised inconsistently with respect to reference structures. Both the practical working routines of lexicographers as well as dictionary users would benefit from better ways of organising reference structures. However, two questions are still open. Why is consistency important here? And who is really investigating the consistency of dictionary entries apart from metalexigraphic critics and someone who is writing a paper about consistency of sense-related items in dictionaries? The average user of a printed dictionary is unlikely to pay much attention to the consistency of given sense-related items. Nevertheless, keeping entries in special dictionaries like dictionaries of synonyms consistent with each other may help to avoid semantic confusion in terms of possibilities of lexical substitution in specific contexts.

In electronic dictionaries, however, things look quite different. Users are more likely to follow links to sense-related items than to look them up by leafing through a printed dictionary. So if in an entry, a synonym is given for a specific sense, and in the link-targeted entry this headword is not mentioned as a synonym, users are probably surprised by the lack of reverse linking. And for the working practice of lexicographers, support of any kind resulting in consistent entries is not only useful but essential. For example, it would be very helpful if, when starting to write a dictionary entry, lexicographers were to receive a computerised message that the entry is already mentioned as a target in another entry. So providing consistency is only possible with extensive computer assistance, particularly for comprehensive data. This is why many lexicographic software tools name this as an important topic in their descriptions,<sup>4</sup> e.g.

TshwaneLex contains various innovative features designed to optimise the process of producing dictionaries, and to improve consistency and quality of the final dictionary product. (Joffe/De Schryver 2004: 1)

---

4. Cf. Lithkowski (2000).

And that leads us again to the meaning of consistency mentioned above, namely consistency with respect to the data model. A data model of reference structures which ensures that reference structures are encoded in a consistent way leads to consistent entries in dictionaries. In the next section, a way of achieving this goal will be proposed.

#### 4. Proposals for modelling sense-related items (and cross-reference structures in general)

One way in which computer assistance may be employed will now be described using the aforementioned electronic dictionary *ellexiko*. The intention is not to introduce specific software or a single content management system. Instead, a concept for modelling reference structures is presented which is applicable to any dictionary and which is not meant for modelling just sense-related items, but lexicographic reference structures in general.

The entries *ehemalig* and *früher* discussed earlier will serve as an example again. Figure 5 shows how the lexicographic data are realised on the underlying structure level below the presentation (on two very small parts of the whole entries). The data are inserted into a fine-grained XML-structure. Without going into detail, it is noticeable that the structure of the XML-based data is strictly content-based and not layout-based. The tags used here are termed `<synonymie>` or `<relpartner>` (relational partner) and not according to typographic presentation such as “*italic*” or “**bold**”. It is also the case that the

<pre>entry „ehemalig“  &lt;ewl-artikel id="272012" - &lt;ewl-lesart id="früher"&gt; - &lt;synonymie&gt;&lt;relpartnerA   artikel-refid="124465"ⓐ   lesart-refid="ehemalig"ⓑ   ltspez-refid="0"&gt;   früher &lt;/relpartnerA&gt; &lt;/synonymie&gt;  &lt;/ewl-artikel&gt;</pre>	<pre>entry „früher“  &lt;ewl-artikel id="124465"ⓐ - &lt;ewl-lesart id="ehemalig"ⓑ &gt; - &lt;synonymie &gt;.&lt;relpartnerA   artikel-refid="272012"   lesart-refid="0"ⓐ   ltspez-refid="0"&gt;   ehemalig &lt;/relpartnerA&gt; &lt;/synonymie &gt;  &lt;/ewl-artikel&gt;</pre>
---	---

Figure 5. Parts of the XML structure (*ellexiko* entries *ehemalig* and *früher*)

modelling is very fine-granular. Every single piece of lexicographic information has its corresponding XML-tag, so every unit is (individually) accessible by computer. For the presentation, the XML data are transformed by an XSLT stylesheet into a HTML-based browser view. Consequently, the presentation of the lexicographic information is defined separately from its content. So the main question concerns the problem of how necessary reference information is recorded in the entries. If the lexicographer is working on one entry (in this case *ehemalig*) and a synonym relation with another word (in this case *früher*) is detected, for example, in the corpus,<sup>5</sup> s/he has to list this term in its corresponding place within the XML structure of the entry. In this case, it is the tag `<relpartner>` within the parent element `<synonymie>`, where s/he has to insert a form of identification, a number assigned to the targeted entry (①) and in certain cases also the aforementioned signpost of an individual targeted sense (②). The ID of an entry is a numeric code greater than zero. The ID of a specifically linked sense of a synonym is the corresponding guide word. The attribute “0” is only given in cases where the linked synonym has not yet been lexicographically described in full, due to the fact that *ellexiko* is not yet a complete dictionary but an ongoing project.

The problem, however, is the lack of technical assistance. When the lexicographer edits an entry and the XML instance is being parsed against the DTDs, there is no technical support in *ellexiko* which signals whether the inserted item could be related to a specific sense or alternatively, if the reference needs to be set to the lexeme in general. If the latter is the case, the lexicographer chooses the attribute “0” instead of the right *lesart*-ID (the right guide word). So if a lexicographer is working on the entry *früher* at a later point than on its corresponding sense-related lexemes, s/he might forget to go back to a certain synonym, for example *ehemalig*. This step is, however, essential in order to add the necessary identification of the guide word and its corresponding sense to realise the link as a sense-related reference (③). Therefore, when starting the lexicographic work on one entry, there is a need for an automatic message with information on all other existing entries which contain data references with respect to the entry that is currently being edited. Again, this is nothing new for software developers working in (commercial) lexicographic context. However, in this context, the problem is discussed on another level, namely the modelling level.

Solving this problem for *ellexiko* is difficult at the moment because from a strictly formal perspective, all reference structures in *ellexiko* go from x to y and – at best – from y to x. This bidirectionality is not currently being checked

---

5. *Ellexiko* is a corpus-based dictionary.

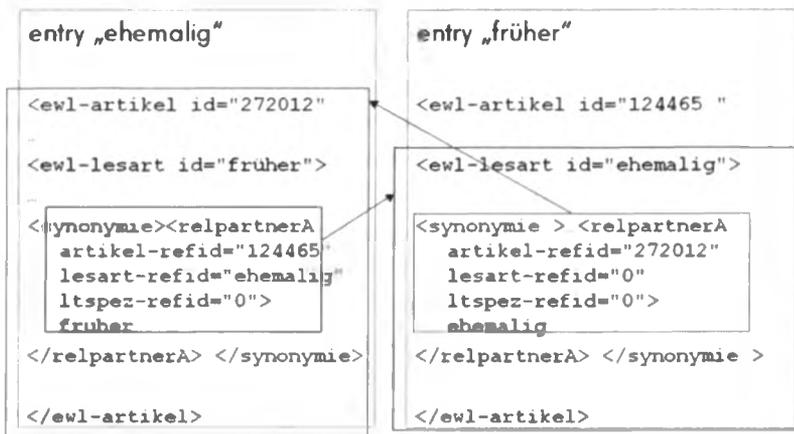


Figure 6. Bidirectional references in elexiko: current status

or tested for systematically by any tool. At the moment, two reference points between which a bidirectional relation exists are just two references without really being in connection with one another, as illustrated in Figure 6. This is the problem most dictionaries face. As things stand, bidirectional references are two independent references from one point out of the hierarchical tree (which is an XML document) to another. Additionally, the two poles between which a relation holds have a different scope and they are technically independent of each other. Improving this situation is quite straightforward, as demonstrated in Figure 6.

Provided that we want to base the modelling of reference structures on a standard, the XML-connected standard XLink (XML Linking Language) is one possible option. Regarding the application of such a standard, XLink is not currently implemented in many tools, so it could be argued that it is not necessary to map the modelling onto this standard. However, in my view, it is very useful to look at this standard, simply because the standard reflects a large number of considerations about reference structures in general, which can only be of benefit (cf. Nordström 2002). However, the following proposals for modelling reference structures could also be implemented in tailor-made XML DTDs or schemas.

XLink has been established to allow “elements to be inserted into XML documents in order to create and describe links between resources” (XLink: 2).<sup>6</sup> The view is taken here that it is a slim as well as an adequately complex format, enabling users to model simple but also more complex linking structures with XLink. In the introduction of the XLink specification, it is stated that: „XLink provides a

6. For the description of the XLink Standard see <http://www.w3.org/TR/xlink/>

framework for creating both basic unidirectional links and more complex linking structures. It allows XML documents to:

- assert linking relationships among more than two resources,
- associate metadata with a link,
- express links that reside in a location separate from the linked resources”.

It is especially important to be able to associate metadata with a link and to build a link database or (abbreviated) linkbase (which is meant for storing links separately from the linked resources).

Before going into detail as to why this is important, some notes on reference structures in general are necessary. Reference structures can be classified into unidirectional and bidirectional relations. For instance, a reference from a dictionary entry to an illustration, a corpus sample or an external encyclopaedia is a unidirectional link. It points in one direction only, i.e. from the target resource there is no reverse reference, which means that referring back to the original source is neither intended nor useful. In the majority of cases, the target resource is outside the lexicographer’s responsibility, as it is outside the lexicographic database. In the context of paradigmatic structures, it is bidirectional references which are of particular interest. Unlike the aforementioned unidirectional reference, creating bidirectional references is part of the lexicographer’s compiling responsibilities. It is only then that two resources may function as a source on the one hand and as the target resource on the other. For example, references to Wikipedia are always given in one direction only (cf. Müller-Spitzer 2007a and b: 169).

In this section, the modelling concept is looked at more closely. The first and most general guideline of the concept for modelling cross-reference structures is to model bidirectional references as **extended links** (in the terminology of XLink) and to store them in a linkbase. This approach has the following advantages:

- someone can work on the links or change them externally without touching the dictionary entries themselves and
- a link database supports the management of the cross-reference structures.

Looking at the current model of references in *elexiko* again, the difference should become obvious. The model idea which is pursued here is one where all links are stored on an individual level, separated from the entries (a draft of this is shown in Figure 7).

What does this mean for the *elexiko* example? The lexicographic information relating to references should be imported automatically into the linkbase file. In Figure 8, we can see that every piece of important information relating to the crosslinking of the given sense-related items is transferred into XLink and its specific elements and attributes. The process is as follows: any information relat-

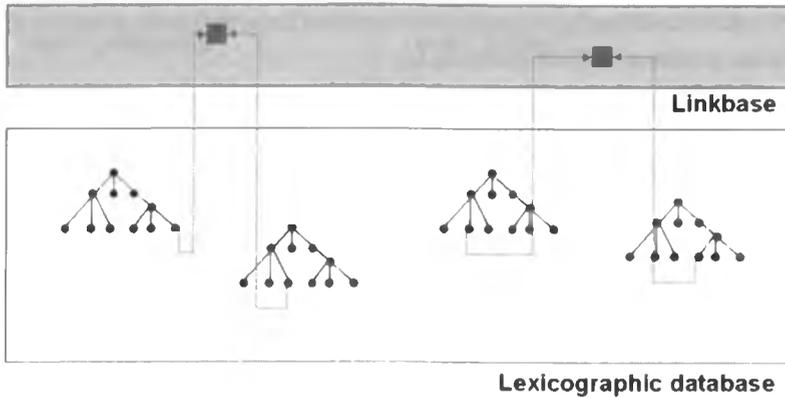


Figure 7. Bidirectional references in elexiko: use of a linkbase

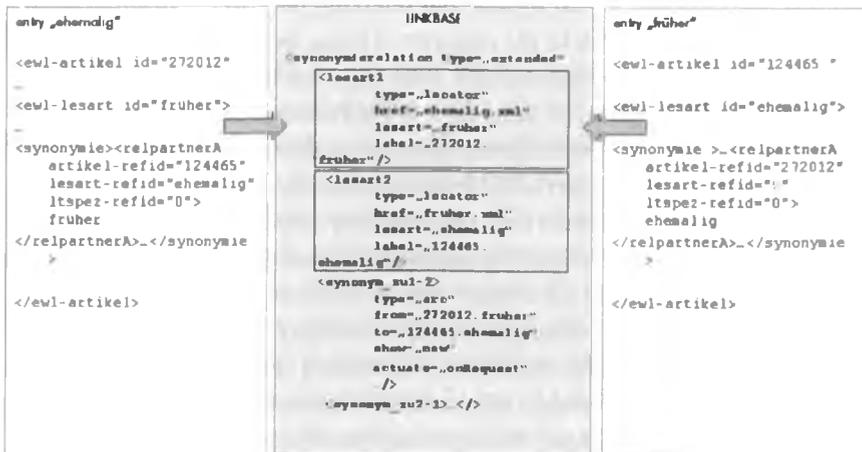


Figure 8. Part of the XLink-linkbase

ing to reference structures is stored in this linkbase. This procedure is performed automatically during the process of checking the entry file into the underlying dictionary database system.

In this model, the connections from one synonym to another and vice versa are not two independent references any longer, but form one complex link object. This has the following advantages:

Firstly, precise addressing of source and target resource is possible. In the data model which is currently being used, one reference goes from the <relpartner> (so from one lexicographic piece of information in the entry) to the target entry or, alternatively, to a specific sense as a whole. Coming from the

related item and looking in the other direction, it is just the other way around. So we have an incorrect addressing structure because the synonym relation is – with regard to the content – valid between two contextual instantiations as represented by senses. Now the storage in a linkbase enables us to address the starting point and the target resource independent of the position where the lexicographic information on these references is given in the entries. We can also see this in the part of the linkbase in Figure 8. The attribute `@label` addresses the sense as a whole as the starting or finishing point of the relation. We might ask whether this is really important. It is assumed here that it is, because if we think of quite different ways of presenting lexicographic data (for example a network of all semantically related senses), it is very important to identify the resources precisely. This is also of particular importance for accessing lexicographic data, for example, if one wishes to present a search word together with its sense-related items.

The second main advantage of using a linkbase is the ability to associate metadata with a link. This is explained further by introducing some general features of XLink.

One crucial point of extended links in XLink is that “the `extended-type` element may contain a mixture of the following elements in any order, possibly along with other content and markup:

- `locator-type` elements that address the remote resources participating in the link
- `arc-type` elements that provide traversal rules among the link’s participating resources
- `title-type` elements that provide human-readable labels for the link
- `resource-type` elements that supply local resources that participate in the link.” [XLink: 11]

The option to add a human-readable title attribute and to specify traversal rules may seem interesting features in this context. Traversal rules are rules which define how to follow a link and can be specified in extended links as follows:

An extended link may indicate rules for traversing among its participating resources by means of a series of optional arc elements. The XLink element for arc is any element with an attribute in the XLink namespace called `type` with the value ‘arc’. [XLink: 16]

[...] The `arc-type` element may have the traversal attributes `from` and `to` [...], the behavior attributes `show` and `actuate` [...] and the semantic attributes `arcrole` and `title` [...]. The traversal attributes define the desired traversal between pairs of resources that participate in the same link, where the resources are identified by their `label` attribute values. The `from` attribute defines

resources from which traversal may be initiated, that is, starting resources, while the to attribute defines resources that may be traversed to, that is, ending resources. The behavior attributes specify the desired behavior for XLink applications to use when traversing to the ending resource. [XLink: 17]

Although it seems to be very interesting to define traversal rules and titles here in the linkbase together with the data model, it has to be noticed that in present-day technology one would prefer to separate the data structure from the application logic. This is an important point of criticism of the XLink-Standard.

Looking again at the example (cf. Figure 8), we can see that in the presented part of the linkbase there are two objects modelled as resources, which are elements with the XLink attribute “locator”. They are identified by their labels, in this case the IDs, together with the guide words which are used for identifying senses in *elexiko*. At the bottom of the middle column in Figure 8, the first *arc* (an element with the XLink-specific attribute *@arc*) with the name `<synonym_zu1-2>` appears. The second *arc* is termed `<synonym_zu2-1>`. All these elements are parts of the complex link object *synonymierelation*. The behaviour of attributes of XLink can generally speaking be characterised as follows: these attributes refer to an application while traversing the links. As has been pointed out before, these attributes should be defined separately from the linkbase.

What impact do these general explanations have when they are related to the concrete examples of sense-related items in *elexiko*? By specifying a model for a linkbase, it is possible to define a fixed type of extended link for the relation type synonymy. This model may be presented as follows. The synonym link object always connects two remote resources with one another. These resources are entries or individual senses respectively. The corresponding arcs between these resources could be specified by traversal rules. In the case of synonymy, it would be useful to specify the value of *show* into *new*. That is, when a user clicks on a synonym, the targeted entry has to be opened in a new frame. In this way, both windows can be arranged next to each other on the computer screen and the two entries can then be received simultaneously. The value of *actuate* is probably to be assigned to *onRequest*. This means that the user has to click on the given relation partner (i.e. sense-related item) in the entry in order to follow the link. This abstract model is then applicable to each concrete synonym relation. This kind of model allows us to specify different presentations of different kinds of lexical semantic relation. For example, in the case of lexemes which are connected by hyperonymy (i.e. superordination) or hyponymy (subordination), users might like to look these up one after the other, whereas in the case of lexemes which are

connected by synonymy or antonymy, users might compare both by simultaneously looking up both entries.

Being able to add metadata to a link object can also be used in an extended way. It may be that a relationship between two words might be more significant for lexeme *a* than it is for lexeme *b*. When a collocation analysis of both terms individually is performed, synonym *a* might rank higher in its significance to *b* than the other way around. This is a regular observation with polysemous lexical items. Such an observation could also be given as metadata, so the XLink-feature of adding metadata could be used here as well. It is clear that this new modelling concept is much more powerful than the one that is currently being used, for example in *elexiko* and in other dictionaries as well.

The final question that remains open is what the consequences might be of the employment and integration of such a linkbase into an editing system for the practical lexicographic work in *elexiko* (and similar projects). The course of the working process can be roughly sketched as follows: a lexicographer checks an entry out of the database system in order to work on it. Then, as a first step, the linkbase verifies whether the chosen entry is registered as a target resource. If it is, the lexicographer receives an automatic message informing him/her which resource (that is which entry) is the starting point of the relation and which type of relation exists between the items. So the lexicographer can have this in mind while studying the corpus results. Then the entry is lexicographically edited in an XML editor, including all references. Next, another check routine is added. While checking the edited entry-file back into the database system, the corresponding information is imported into the linkbase and – and this is the most important thing – a query is initiated as to whether the information about one reference structure is *consistent* with the information about the other. If one connection runs from the start resource to the target resource but not back again, the lexicographer receives an error message. In *elexiko*, it is important from a pragmatic point of view to keep the writing practice in the XML editor as it is. Alternatively, in other projects, it may be better practice to source all linking information out of the entries and store them in the linkbase only. Not only does this form of modelling reference structures allow for typing and assigning attributes as well as for new ways of presenting reference structures, but, most importantly, it also allows lexicographers to make entries more consistent. At this point, it should be stressed again that the modelling concept presented in this paper does not necessarily imply the use of XLink. A tailor-made XML-structure, based on the same guidelines, is able to perform the same tasks.



headword itself is positioned in the middle, and around it are its senses, labelled “LA” for “Lesart”, followed by the types of sense relation such as synonymy or incompatibility.

The purpose of the graph is to show that as well as more traditional lexicographic presentations, there are other ways of presenting sense-related connections in a dictionary. What needs emphasising at this point is the fact that different ways of presentation can be made without any change to the underlying data. The preferred way of structuring data enables us to do this at the touch of a button, if the appropriate tool is available. Moreover, if the necessary data are structured according to the new modelling concept, much more elaborate ways of presentation are possible.

To conclude, providing and supporting consistency of sense-related items and reference structures in dictionaries in general should not be seen as an irrelevant hobby of dictionary reviewers and text-technologists. Providing consistency is crucial for lexicographic practice and it can have an effect on qualitative enhancements. At the same time, a consistent way of data modelling and structuring is a prerequisite for developing innovative forms of presenting lexicographic data in an electronic medium such as the Internet. The benefits of consistency in general lie therefore on both sides, for the lexicographer who is compiling the dictionary and for the user who needs to make successful look up procedures.

## References

- Blumenthal, Andreas, Lemnitzer, Lothar and Storrer, Angelika. 1988. “Was ist eigentlich ein Verweis? Konzeptionelle Datenmodellierung als Voraussetzung computergestützter Verweisbehandlung.” In *Das Wörterbuch. Artikel und Verweisstrukturen* (Jahrbuch 1987 des Instituts für deutsche Sprache), Gisela Harras (ed.), 351–373. Düsseldorf/Bielefeld: Pädagogischer Verlag Schwann-Bagel and Cornelsen-Velhagen u. Klasing.
- Cruse, D. Alan. 2004. *Meaning in Language*. (2nd ed.) Oxford: Oxford University Press.
- Engelberg, Stefan and Lemnitzer, Lothar. 2001. *Lexikographie und Wörterbuchbenutzung* (Stauffenburg Einführungen, vol. 14). Tübingen: Stauffenburg.
- Joffe, David and De Schryver, Gilles-Maurice. 2004. “TshwaneLex – Professional off-the-shelf lexicography software.” In *DWS 2004 – Third International Workshop on Dictionary Writing Systems*, 17–20. <http://tshwanedje.com/publications/dws2004-TL.pdf> (last visited on 2009/07/03).
- Lew, Robert. 2007. “Linguistic semantics and lexicography: A troubled relationship.” In *Language and meaning. Cognitive and functional perspectives*, Malgorzata Fabiszak (ed.), 217–224. Frankfurt am Main: Peter Lang.
- Litkowski, Ken C. 2000. “The Synergy of NLP and Computational Lexicography Tasks”. *Technical Report 00-01*. Damascus, MD: CL Research.
- Müller-Spitzer, Carolin. 2007a. “Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung.” *Hermes* 38/2007: 137–171.

- Müller-Spitzer, Carolin. 2007b. *Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis* (Studien zur deutschen Sprache 42). Tübingen: Gunter Narr.
- Nordström, Ari. 2002. "Practical XLink." In *XML 2002, Proceedings by deepX*. [www.idealliance.org/papers/xml02/dx\\_xml02/papers/06-00-11/06-00-11.html](http://www.idealliance.org/papers/xml02/dx_xml02/papers/06-00-11/06-00-11.html) (last visited June 2008/06/01).
- Storjohann, Petra. 2005. "elexiko: A Corpus-Based Monolingual German Dictionary." *Hermes* 34/2005: 55–73.
- XML Linking Language (XLink) Version 1.0 (2001). W3C Recommendation 27 June 2001 <http://www.w3.org/TR/xlink/> (last visited on 2009/07/03).

## Reference works

- Duden. 2007. *Das Synonymwörterbuch*, 4th edition. Mannheim/Leipzig/Wien/Zürich: Dudenverlag.
- elexiko. 2003ff. In *OWID – Online Wortschatz-Informationssystem Deutsch*, Mannheim: Institut für Deutsche Sprache, [www.owid.de/elexiko\\_/index.html](http://www.owid.de/elexiko_/index.html) (last visited on 2009/07/03).
- Merriam Webster Online. 2009. ([www.Merriam-Webster.com](http://www.Merriam-Webster.com)). entry *Consistency*. <http://www.merriam-webster.com/dictionary/consistency> (last visited at 2009/07/02).
- Wahrig. 2006. *Synonymwörterbuch*, 5th edition. München/Gütersloh: Wissen Media Verlag.
- Wikipedia.com. 2009. ([www.wikipedia.com](http://www.wikipedia.com)). entry *Consistency*: <http://en.wikipedia.org/wiki/Consistency> (last visited on 2009/07/03).
- Stanford Encyclopedia of Philosophy. (<http://plato.stanford.edu/entries/logic-classical/>). entry *Consistency* <http://plato.stanford.edu/search/searcher.py?query=consistency> (last visited at 2009/07/02).