

Lexicographic Data Boxes Part 1. Lexicographic Data Boxes as Text Constituents in Dictionaries*

Rufus H. Gouws, *Department of Afrikaans and Dutch,
Stellenbosch University, Stellenbosch, South Africa (rhg@sun.ac.za)*
and
D.J. Prinsloo, *Department of African Languages, University of Pretoria,
Pretoria, South Africa (danie.prinsloo@up.ac.za)*

And they all get put in boxes, little boxes all the same
(Malvina Reynolds)

Abstract: This article, the first in a series of three on lexicographic data boxes, focuses primarily on the occurrence of lexicographic data boxes as text constituents in dictionaries. Following a brief look at what data boxes are, the focus shifts to the different venues where these boxes can be accommodated within the central list of a dictionary. Boxes containing items and/or item texts can be positioned within articles, or article-externally as phased-in inner texts within a partial article stretch of a dictionary. Data boxes are used to convey data that need to be highlighted and are therefore often formally marked (a coloured background or within a frame) and are put in an article slot that has a position of salience. As dictionary entries they can participate in procedures of both lemmatic and non-lemmatic addressing. It is shown that a box should preferably be inserted close to its address. In articles of polysemous words, the user should unambiguously know for which sense(s) the box is relevant. As phased-in inner texts data boxes can be addressed at a lemma within the same partial article stretch but also, in the case of synopsis boxes, at lemmata in other article stretches. This demands procedures of remote addressing.

Keywords: ADDRESSING, ARTICLE STRETCH, ARTICLE WINDOWS, ARTICLE-EXTERNAL DATA BOXES, ARTICLE-INTERNAL DATA BOXES, DATA BOXES, DATA DISTRIBUTION, EXPANDED WORD LIST, INSERTS, LEXICOGRAPHIC DATA BOX, PARALLEL MACROSTRUCTURE, PARTIAL ARTICLE STRETCH, PHASED-IN INNER TEXTS

Opsomming: Leksikografiese datakassies. Deel 1. Leksikografiese datakassies as tekskonstituente in woordeboeke. Hierdie artikel, die eerste in 'n reeks van drie oor leksikografiese datakassies, fokus veral op die voorkoms van leksikografiese datakassies as

* This is the first in a series of three articles dealing with various aspects of lexicographic data boxes.

tekstkonstituente in woordeboeke. Na 'n kort bespreking van wat datakassies is, verskuif die fokus na die verskillende plekke waar hierdie kassies in die sentrale woordelys van 'n woordeboek geakkommodeer kan word. Datakassies wat aanduiders en/of aanduidertekste bevat, kan binne-in artikels, of artikel-ekstern as ingefaseerde binnetekste in die deeltrajekte van 'n woordeboek geplaas word. Datakassies word gebruik om data oor te dra wat beklemtoon moet word en word daarom dikwels gemerk (met 'n gekleurde agtergrond of in 'n raam) en word in 'n artikelgleuf geplaas wat in 'n posisie is wat die aandag trek. As woordeboekinskrywings word datakassies betrek by prosedures van lemmatiese en nielemmatiese adressering. Dit word aangetoon dat 'n kassie liefso na as moontlik aan sy adres geplaas moet word. In artikels van polisemiese woorde moet die gebruiker ondubbelsinnig kan weet vir watter betekenisonderskeiding(e) die kassie relevant is. As ingefaseerde binnetekste kan datakassies aan 'n lemma in dieselfde deeltrajek geadresseer wees, maar ook, in die geval van sinoptiese kassies, aan lemmata in ander artikeltrajekte. Dit vereis prosedures van verwyderde adressering.

Sleutelwoorde: ADRESSERING, ARTIKELDEELTRAJEK, ARTIKEL-EKSTERNE DATAKASSIES, ARTIKEL-INTERNE DATAKASSIES, ARTIKELTRAJEK, ARTIKELVENSTERS, DATAKASSIES, DATAVERSPREIDING, INGEFASEERDE BINNETEKSTE, INVOEGING, LEKSIKOGRAFIESE DATAKASSIE, PARALLELE MAKROSTRUKTUUR, UITGEBREIDE WOORDELYS

1. Introduction

The emergence¹ and establishment of dictionary structures as a focal area of research in the field of metalexigraphy and dictionary research has had a significant influence on the theory and practice of lexicography. The scope of this influence included a critical analysis and discussion of the occurrence and positioning of a variety of text compound constituents, texts and textual segments in dictionaries and dictionary articles. Research into dictionary structures had both a contemplative and a transformative approach, cf. Tarp (2004: 224). Existing structures were studied and described but a whole range of new structures were identified, proposed and employed in the planning and compilation of dictionaries. Research into dictionary structures resulted in lexicographers realising all articles in a given dictionary do not necessarily have to display exactly the same structure. A sequence of articles in a single article stretch does not merely have to be a presentation of more of the same. Instead of a single and consistently applied homogeneous article structure the lexicographer has the liberty to opt for less rigid heterogeneous article structures in which the obligatory microstructure, the default microstructure of the specific dictionary, can be supplemented by items enabling an extended obligatory microstructure. Different types of word list, micro- and article structures give lexicographers the opportunity to present partial article stretches, microstructural items and item texts in a way that would best fit the needs and reference skills of their intended target users. Yet again, the user-perspective plays a dominant role in the planning and compilation of dictionaries. This perspective also co-determines the data distribution and data presentation in any given

dictionary as a search region, the word list as a search field, an article as a search area and the article slots as search zones, but also the structure allocated to any article as well as the potential inclusion of phased-in inner texts in a given article stretch. For the study of dictionary structures the occurrence and positioning of each partial article stretch, item text, item and indicator is of significance. This also applies to data boxes as lexicographic text constituents.

Lexicographic data boxes are lexicographic text constituents frequently employed in the presentation of data in various search venues of dictionaries. Although data boxes have become a common phenomenon in dictionaries, relatively little attention has been paid to their presentation, the motivation for their use and the type of data a dictionary displays in this specific way, cf. Gouws and Prinsloo (2010) and Taljard, Prinsloo and Gouws (2014). This article firstly motivates the need for a theoretical discussion of data boxes, focusing on the types of data boxes and the textual positions allocated to them before following a contemplative approach by taking a look at some aspects of the current use of data boxes. A transformative approach is also followed and some ideas deemed necessary for an improved use of this lexicographic text segment will be presented in the final article of this three-part series.

The focus in this article is on the occurrence of data boxes as text constituents in dictionaries. In the follow-up articles, part 2 and 3 in this series, the focus will be on the contents of data in boxes and a look at future improvements respectively.

2. What are lexicographic data boxes?

Before proceeding with a discussion of lexicographic data boxes it is important to have a clear understanding of what is meant in this article by the term *lexicographic data box* and its shortened form *data box*.

A lexicographic data box, or just data box, is a data-carrying constituent of the word list of a dictionary. It contains data presented

- as part of the treatment of a specific lemma that is the guiding element of the article in which the data box appears or that is in close proximity to the data box;
- as part of the treatment of various lemmata where the data box is an entry in an article or in a partial article stretch that accommodates one of the lemmata for which the treatment in the data box is relevant;
- in a phased-in inner text in an article stretch that contains a lemma for which the contents of the data box is relevant.

Data boxes contribute to the lexicographic treatment of a lemma in the same article stretch where they appear, specifically a lemma that is the guiding element of the article in which the data box is presented as entry or a lemma in close proximity of that data box. In addition, synoptic data boxes can also contain data relevant to the lexicographic treatment of lemmata in other article

stretches of the dictionary. Data boxes are usually distinguished from other data-carrying constituents in terms of the form in which they are presented. They often occur within a framed box, as a text constituent with a coloured background, as an article window (cf. Wiegand and Gouws 2011: 281), or in clearly identified phased-in inner texts that split a partial article stretch. Although data boxes often contain data presented within a frame or against a coloured background, this is not always the case. Some data boxes are not presented in boxed or coloured format.

The term *data box* is preferred to the frequently used term *text box* because texts are not the only type of data to be accommodated in these boxes.

Examples of different types of data boxes will be presented in later sections of this article.

3. Why data boxes

Lexicographers should refrain from procedures that could result in a situation of data overload in their dictionaries, cf. Gouws and Tarp (2017). In its default microstructural presentation a dictionary article should contain the data the lexicographer regards as sufficient to satisfy the lexicographic needs of the envisaged target user. In addition to the data presented in the word list of a dictionary the lexicographer may also use a data distribution structure that allocates certain data to texts in the front and/or back matter sections of the dictionary. Where a lexicographer is convinced that specific data entries are needed to improve the lexicographic presentation and treatment in the word list of a given dictionary but the default microstructure, article structure and word list do not offer an appropriate position for such data entries, the lexicographer may embark on using data boxes — either article-internally or as phased-in inner texts. The use of data boxes needs to ensure an added value to the default treatment on offer in a given dictionary and the contents of data boxes should not be more of the same with regard to the default data presentation in dictionaries.

Data boxes are salient dictionary entries and as such they should be employed when there is a need to bring a non-default type of data to the attention of the user or to place more than the usual focus on a specific data item. As a result of lexicographic procedures used in an extended obligatory microstructure, care should be taken that data boxes do not become part of the obligatory microstructure and in so doing lose their significance and decrease the emphasis on the data included in these boxes. Lexicographers should make a clear distinction between using data boxes and using other lexicographic procedures, e.g. a system of labelling, to focus the attention of the user on a specific item.

4. The need for a discussion of data boxes

The broad field of lexicography is characterised by an interactive relation

between theory and practice. Because the theory of lexicography emerged much later than the practice there had to be a lot of catching up to ensure a comprehensive theoretical coverage of the endeavours of the lexicographic practice and to present future compilers of dictionaries with the necessary theoretical basis for their dictionaries. In lexicography theory and practice do not develop in a parallel way. In lexicographic theory new suggestions come to the fore and they are often only applied in practice at a much later stage, e.g. the proposal for semi-integrated microstructures (Wiegand 1996). Some suggestions resulting from research in the field of theoretical lexicography never even find an application in the lexicographic practice. Practical lexicographers often introduce innovative approaches in their dictionaries and the theory may eventually include a discussion of these approaches when they have already been firmly established in dictionary compilation processes. Yet, it remains important that there should be a relation of reciprocity between theory and practice in lexicography: theory should learn from and influence the practice and the practice should learn from and influence the theory.

The occurrence of data boxes in printed dictionaries, especially learner's dictionaries, as well as in electronic dictionaries, especially those which are based on paper dictionaries, has increased significantly in the last decades. Although (learner's) dictionaries continuously witness an enhancement in quality, also as a result of extensive research in the field of theoretical lexicography with regard to this dictionary type, the competitive market and the competition between publishing houses have also had an influence on changes in the structures and contents of dictionaries. Many of these changes have added value to the dictionaries but in some instances they were, according to Wiegand and Gouws (2011: 238), not much more than trends in lexicographic face-lifting. The lexicographic practice did, however, introduce innovative uses of data boxes coming to the fore in many dictionaries. In this regard practice took the lead and theory unfortunately failed to respond quickly enough. Consequently many of these approaches have not yet been sufficiently appreciated and discussed by metalexicographers.

In order for lexicographic theory to keep up with developments in the practice and in order to ensure that future dictionary compilers can use the theory for both a contemplative and a transformative commitment to the planning and compilation of their dictionaries, the use of data boxes needs to be included in theory-based discussions.

5. Data boxes as dictionary entries

5.1 Data boxes and addressing relations

The addressing structure and various addressing relations in dictionaries are important to identify the scope of each item in a dictionary article and the target of its treatment. According to Hausmann and Wiegand (1989: 328) a treat-

ment unit results "when a form mentioned and information relating to that form are brought together. The relation of form and information is that of topic and comment." The "way in which a form and information relating to that form are brought together is the *addressing procedure*." Each information item is addressed to a form that is known as its address.

Data boxes are functional constituents of dictionaries and should not be employed as mere procedures of lexicographic face-lifting. They are included in dictionaries as part of the lexicographic treatment procedures. When allocating a data box to a specific search zone cognizance needs to be taken of the specific treatment contributed by the data box. Dictionary users should know exactly what constitutes both the address and the addressing relation of each data box.

To ensure an optimal comprehension of the relevant addressing relation, it is important that the notion of addressing has to be employed in an unambiguous way. In this article the procedures of addressing as discussed in Gouws (2014) will be employed. Wiegand and Gouws (2013) restrict addressing to relations within condensed articles where items display the addressing relations. Wiegand and Gouws (2013: 273) say that there are no addressing relations in non-condensed dictionary articles or in other non-condensed accessible entries. This implies that an item like a lemma sign cannot be addressed by an item text, even if the item text contributes to the treatment of the lemma. In contrast to this approach, Gouws (2014: 183) expands the application of the notion of addressing. The term addressing is used to incorporate both primary addressing which is the traditional procedure prevailing in condensed articles, and secondary addressing to go beyond textual condensation and item status as prerequisites for addressed entries. In accordance with this approach both items and item texts can participate in procedures of addressing. The addressing element and its address do not have to be in the same dictionary article and addressing can occur between different types of constituents of a word list, e.g. also between a phased-in inner text and an item in a dictionary article.

5.2 The position for data boxes and the different types of data box entries

In accordance with the data distribution structure of a specific dictionary the lexicographer has to allocate data-carrying segments to the different search positions in the dictionary. Where a dictionary displays a frame structure the positioning of data in the articles of the word list will be complemented by text constituents or other entries presented in the front and/or back matter sections of that dictionary. As a carrier of text types a dictionary could also include outer texts in its middle matter section.

Dictionaries have different search positions, traditionally ranging from a search zone within a dictionary article to the article as search area and the word list, the largest search position, as search field, cf. Wiegand, Beer and Gouws (2013: 63). Gouws (2018: 228) argued in favour of a further and more compre-

hensive search position, i.e. the search region. This search position is constituted by all the textual components of a dictionary as a text compound. For a single volume dictionary this is the most comprehensive search position. Although the outer texts of the front and back matter sections of a dictionary fall outside the domain of the search field they are part of the search region of a dictionary. Data boxes typically occur within a word list, that is a search field, of a dictionary. This could be within the central and only word list which is the most frequent occurrence of data boxes, or within any other word list that is part of a word list series of a given dictionary. Looking at data boxes in this article, the focus will only be on those data boxes occurring in a word list, i.e. within the search field of dictionaries. Data boxes in outer texts will not be discussed.

As entries in a search field the status of all data boxes is not the same. Data boxes could be phased-in inner texts, article-internal item texts or mere items occupying a search zone in a dictionary article. These different types of data boxes will be discussed in the subsequent sections.

5.3 Expanded word lists

The data distribution structure of any dictionary determines the way in which the lemmata presenting the macrostructural coverage of the dictionary have to be ordered within the word list of that dictionary. Whether it is a straight alphabetical ordering or an ordering with sinuous lemma files that enables the use of niched and nested lemmata, a word list will consist of a number of article stretches that accommodate articles with the lemmata as their guiding elements.

When planning the data distribution structure of a dictionary lexicographers need to make a decision regarding the type of word list the dictionary will display. This is determined by specific features of the nature of data allocated to the word list. In this regard a distinction is made between a single word list and an expanded word list, cf. Bergenholtz, Tarp and Wiegand (1999: 1766). A single word list contains article stretches but no inserts or phased-in inner texts, whereas an expanded word list also contains article stretches as well as inserts and/or phased-in inner texts. These inserts and phased-in inner texts split sequences of articles within partial article stretches. Therefore an expanded word list is also known as a split word list.

The distinction between inserts and phased-in inner texts is significant for the identification of data boxes.

5.3.1 Distinguishing between inserts and phased-in inner texts

5.3.1.1 Inserts

According to the *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary*

of *Lexicography and Dictionary Research* (Wiegand et al. 2017), (henceforth abbreviated as WLWF) an insert is a text or text part that is inserted into the word list of a dictionary. It is an immediate constituent of the dictionary as a text compound. Inserts often are sections of photos, inserted between two article stretches or between two pages of an article stretch. In the *Merriam-Webster's Advanced Learner's English Dictionary* (Perrault 2008), (henceforth abbreviated as MWALD), the article stretch of the letter M is split between the lemmata *mascot* and *masculine* by an insert titled *Color Art*. This insert, with its own table of contents, contains pictures of themes like colours, vegetables, landscapes, gems and jewellery and clothing. These pictures do not adhere to the alphabetical ordering within the article stretch and although it splits the word list this insert is not an immediate constituent of either the specific article stretch or the word list. It is an immediate constituent of the dictionary as text compound and carrier of text types. Consequently inserts cannot be regarded as a type of data box or the type of data box discussed in this article.

The form of inserts sometimes resemble that of data boxes, e.g. the section on birds in an insert from the MWALD contains pictures of a variety of birds and the data box occurring in the article of the lemma *deer*, cf. figure 18 in paragraph 5.5.2.1, contains a number of pictures of different types of deer. Although both these text constituents are carriers of pictures, the insert is a different type of lexicographic text and will not be discussed any further in this article.

5.3.1.2 Phased-in inner texts

Phased-in inner texts occur in expanded word lists where they split the article stretches, resulting in internally expanded article stretches (Bergenholtz, Tarp and Wiegand 1999: 1766). The phased-in inner texts typically contain data relevant to a lemma in close proximity within the same article stretch and often also to other lemmata occurring in remote articles. The data in the phased-in inner texts can also have a more general relevance to the specific lemma or might link thematically with it. Phased-in inner texts often have an inner text title that can concur, but not necessarily, with the lemma of an article in its proximity. This title could also identify the more general nature of the data in the specific text. These texts are usually typographically distinguishable from default articles as constituents of a word list by being framed or presented with a coloured background. Examples will be provided in a subsequent section of this article.

Article-internal data boxes and phased-in inner texts sometimes show strong resemblances but the nature of a specific box as a text constituent of the word list determines whether it has to be regarded as an article-internal box or a box presented as phased-in inner text.

5.4 Article-internal data boxes

5.4.1 Items and item texts

One of the most frequent occurrences of data boxes is within dictionary articles. As constituents of articles data boxes do not form part of the obligatory microstructure but rather enrich the obligatory microstructure, resulting in an extended obligatory microstructure. Data boxes in dictionary articles mostly contain text data but they can also display non-textual data. Data boxes are not default entries in any article but their use can be regarded as an extended treatment procedure. They can either contain items or item texts. Wiegand and Smit (2013: 153) distinguish as follows between items and item texts:

An item is a functional text segment without the status of a sentence but with the status of a text segment which is given as a discernible item form assigned with at least one genuine function, the latter being precisely such that a user can obtain knowledge about the item's subject as lexicographical information.

An *item text* is a functional text segment with item function and text constituent status exhibiting a complete and distinct natural-language syntactical structure and consisting of at least one sentence.

As constituents of dictionary articles data boxes may contain data presented in either condensed or non-condensed format. The articles in figure 1, a partial article stretch from the monolingual Afrikaans learner's dictionary *Basiswoordeboek van Afrikaans* (Gouws et al. 1994), (henceforth abbreviated as BW) contain two data boxes with items giving the pronunciation of the word represented by the lemma of the specific article. This dictionary only gives pronunciation guidance for a limited number of selected lemmata. It is not part of the obligatory microstructure and when pronunciation guidance is regarded as necessary, it is presented in a data box accommodated in the final slot of the article. The entries "Uitspraak 'gemie'" (= Pronunciation 'gemie') and "Uitspraak 'sjirurg'" (= Pronunciation 'sjirurg') are condensed forms and thus items.

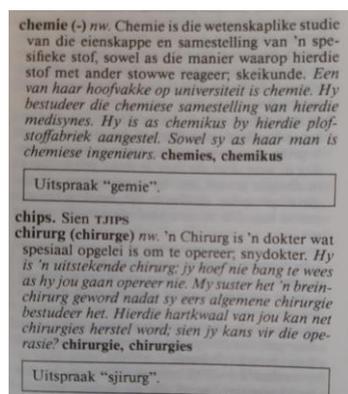


Figure 1: Partial article stretch from BW

The data box in figure 2 from the *Junior Tweekalige Skoolwoordeboek/Junior Bilingual School Dictionary*, (Stoman et al. 2018), (henceforth abbreviated as JBS) contains a sentence which is an entry given in non-condensed form. This entry therefore is an item text.

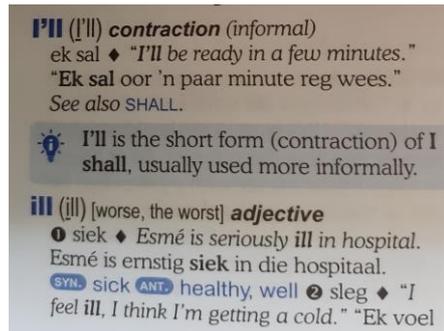


Figure 2: Partial article stretch from JBS

Whether entries are items or item texts may have implications for the addressing relations in an article. This will be discussed in a further section of this article.

5.4.2 Positioning data boxes within dictionary articles

Data boxes are functional dictionary entries because they contribute to achieving the genuine purpose of the dictionary in which they occur. That implies that from the data on offer in a data box users must be able to retrieve information that can assist them in finding a solution to the problem that initiated the specific dictionary consultation procedure. Because data boxes result from extended treatment procedures and are usually not entries belonging to the obligatory microstructure of a dictionary, users will not know beforehand whether an article contains a data box. That is why it is important that these boxes need to be clearly marked as framed or coloured text constituents. Dictionary articles have no fixed slot reserved for data boxes. When planning the data distribution structure of a dictionary lexicographer's need to decide on the slots in dictionary articles that could accommodate search zones populated by data boxes. Data boxes could be placed in different text positions in dictionaries, cf. Taljard et al. (2014: 698). They could be included within the comment on form or comment on semantics or in an alternative position, e.g. as precomment or postcomment.

5.4.2.1 Data boxes at the end of an article

One of the most typical positions allocated to data boxes, cf. Gouws and Prinsloo (2010), is a slot at the end of the article. In such a position the data box often

falls beyond the scope of either the comment on form or the comment on semantics. As postcomment the data box is in a position of salience — further accentuated by its frame or colouring. Figure 3 shows the article of the lemma sign *especially* in the *Oxford Advanced Learner's Dictionary of Current English* (Turnbull 2010), (henceforth abbreviated as OALD):

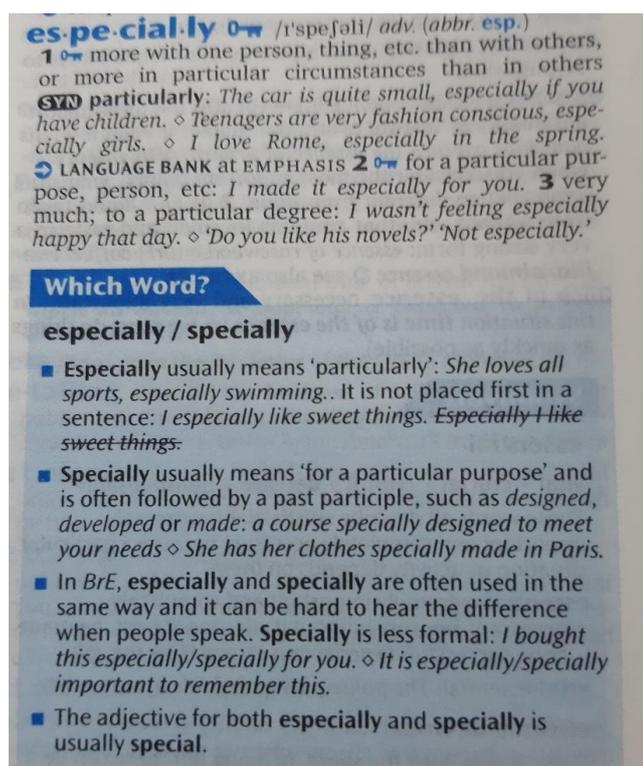


Figure 3: Article in the OALD

The obligatory microstructure ends with the items in the third subcomment on semantics. The lexicographers have felt the need to present the users with additional guidance regarding the use of the words *especially* and *specially*. The data box containing this data has the title "Which Word?" and is presented after the comment on semantics as a postcomment to the article. Another example of a box presented as postcomment in the same dictionary can be found in the article of the lemma sign *restaurant*, figure 4. A type of data box frequently given in this dictionary contains collocations, but not necessarily collocations in which the word represented by the lemma functions as either base or collocator, but rather collocations applicable in the semantic domain of the word represented by the lemma sign. The user gets assistance regarding typical collocations used when dining in a restaurant:

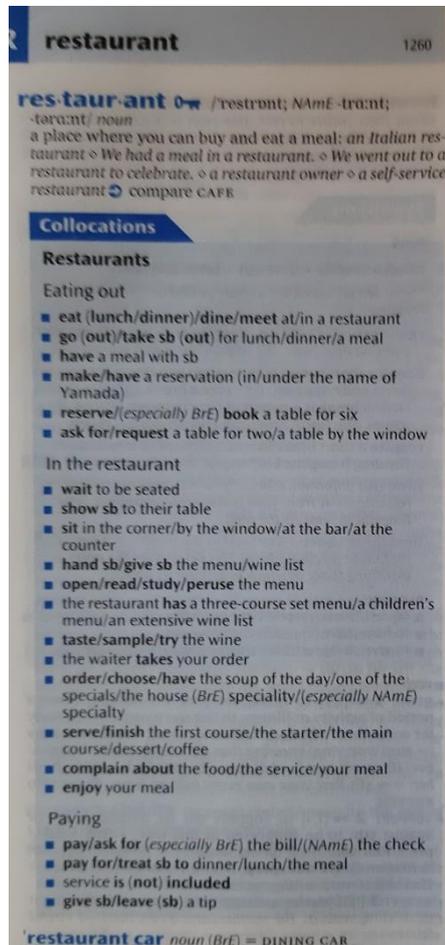


Figure 4: Collocations in the OALD

The OALD presents the entries in data boxes against a blue background. These data boxes can easily be identified by the users of this dictionary.

There is no restriction on the number of data boxes per article, but lexicographers need to be careful not to have a data box overload that could diminish the emphasis on these boxes. An inflation of data boxes should therefore be avoided, but even more than one clearly marked data box per article should still be in order, as seen in figure 5, the article of the lemma sign *school* in the OALD. The two boxes "British/American" and "Grammar Point" are clearly identifiable and can enhance the nature and extent of the lexicographic treatment in this article.

school **0**→ /sku:l/ noun, verb

noun

- **WHERE CHILDREN LEARN** **1** **0**→ [C] a place where children go to be educated: *My brother and I went to the same school.* ◊ (formal) *Which school do they attend?* ◊ *I'm going to the school today to talk to Kim's teacher.* ◊ *We need more money for roads, hospitals and schools.* ◊ *school buildings* **2** **0**→ [U] (used without *the* or *a*) the process of learning in a school; (BrE) the time during your life when you go to a school; (BrE) to *start/leave school* ◊ (NAME) to *start/quit school* ◊ *Where did you go to school?* ◊ (BrE) *All my kids are still at school.* ◊ (NAME) *All my kids are still in school.* ◊ (NAME) to *teach school* (= teach in a school) ◊ *The transition from school to work can be difficult.* ◊ **COLLOCATIONS** at EDUCATION **3** **0**→ [U] (used without *the* or *a*) the time during the day when children are working in a school: *Shall I meet you after school today?* ◊ *School begins at 9.* ◊ *The kids are at/in school until 3.30.* ◊ *after-school* activities
- **STUDENTS AND TEACHERS** **4** **0**→ **the school** [sing.] all the children or students and the teachers in a school: *I had to stand up in front of the whole school.*
- **FOR PARTICULAR SKILL** **5** **0**→ [C] (often in compounds) a place where people go to learn a particular subject or skill: *a drama/language/riding, etc. school*
- **COLLEGE/UNIVERSITY** **6** **0**→ [C, U] (NAME, informal) a college or university; the time that you spend there: *famous schools like Yale and Harvard* ◊ *Where did you go to school?* ◊ see also **GRADUATE SCHOOL** **7** **0**→ [C] a department of a college or university that teaches a particular subject: *the business/medical/law school* ◊ *the School of Dentistry*
- **OF WRITERS/ARTISTS** **8** [C] a group of writers, artists, etc. whose style of work or opinions have been influenced by the same person or ideas: *the Dutch school of painting*
- **OF FISH** **9** [C] a large number of fish or other sea animals, swimming together: *a school of dolphins* ◊ compare **SHOAL** **HEAD** There are many compounds ending in **SCHOOL**. You will find them at their place in the alphabet.

IDM **school(s) of thought** a way of thinking that a number of people share: *There are two schools of thought about how this illness should be treated.* ◊ more at **OLD**

verb

- **YOURSELF/ANIMAL** **1** (formal) to train sb/yourself/an animal to do sth: ~ **sb/sth/yourself (in sth)** to *school a horse* ◊ *She had schooled herself in patience.* ◊ ~ **sb/sth/yourself to do sth** *I have schooled myself to remain calm under pressure.*
- **CHILD** **2** ~ **sb** (formal) to educate a child: *She should be schooled with her peers.*

British/American

at / in school

- In BrE somebody who is attending school is **at school**: *I was at school with her sister.* In NAME **in school** is used: *I have a ten-year-old in school.* **In school** in NAME can also mean 'attending a university'.

Grammar Point

school

- When a school is being referred to as an institution, you do not need to use *the*: *When do the children finish school?* When you are talking about a particular building, *the* is used: *I'll meet you outside the school.* **Prison, jail, court, and church** work in the same way: *Her husband spent three years in prison.*
- ◊ note at **COLLEGE, HOSPITAL**

school age noun [U] the age or period when a child normally attends school: *children of school age* ◊ *school-age children*

Figure 5: Two data boxes in a single article of the OALD

Where a data box is presented as postcomment in the article of a lemma representing a polysemous lexical item it is not always clear whether the treatment presented in the box is directed at only one or more senses or at all the senses of the word. This problem will be elaborated on in section 5.4.2.4.

5.4.2.2 Data boxes at the beginning of the article

The lemma sign of an article is part of the main outer access structure of the

dictionary and a guiding element of the specific article. Irrespective of the type of information a user wants to retrieve from a dictionary article as search area the access to that item, especially in a printed dictionary, proceeds via the lemma sign onto the search paths of the inner access structure. Items positioned in search zones in close proximity of the lemma are in salient positions. The framed or coloured appearance of data boxes adds focus to their occurrence as text constituents of an article and if such a box is awarded a position close to the lemma sign it further elevates the salience of the data in the box.



Figure 6: The article of *get together* in the MPVP

The *Macmillan Phrasal Verbs Plus* (Rundell 2005), (henceforth abbreviated as MPVP) uses data boxes in a slot close to the beginning of an article, to provide an overview of the senses where the phrasal verb represented by the lemma sign has five or more than five senses. This box differs from other boxes because it does not contain the kind of data typically allocated to data boxes. It displays items conveying the menu of the comment on semantics and they function as navigational entries in this article. Alternative strategies for the provision of such navigational guidance could for instance be shortcuts, signposts and guidewords and will be discussed in detail in Part 2 of this series. The box as seen in figure 6 functions in such an article as a precomment.

Data boxes are also included as precomments of articles in the *Macmillan English Dictionary* (Rundell 2007), (henceforth abbreviated as MED). In figure 7, the article of the lemma sign *above*, the data box contains data regarding the use of the word *above* in its occurrence in different parts of speech. It makes the user aware of the differences between these uses.

above /ə'ʌv/ adj, adv, preposition ★★★

Above can be used in the following ways:

- as a **preposition** (followed by a noun): *The birds were flying high above the trees.*
- as an **adverb** (without a following noun): *She stared up at the stars above.*
- as an **adjective**: *Please reply to the above address.*

1 used for saying where sb/sth is 1a. at a higher level than something or directly over it: *We lived in the room above the shop.* ♦ *Curran's leg was broken above the right knee.* ♦ *the snow-covered hills above the village* **1b.** in an earlier part of a piece of writing or higher up a page: *Many of the documents mentioned above are now available on the Internet.* ♦ *David and Brenda Mitchell, pictured above, celebrated their golden wedding anniversary last weekend.* ♦ *Interviews will be held at the above address on 2 December.*

2 higher in amount or standard 2a. more than a

Figure 7: The article of *above* in the MED

To ensure that the target user of this learner's dictionary will focus on the differences between the uses of the word, this data can at best be presented in a precomment instead of being subsumed under the cotextual items in different comments or subcomments on semantics of one of the three partial articles of this twofold complex dictionary article, cf. Wiegand and Gouws (2011: 242).

Because the beginning of an article is a significant position of salience in a dictionary article lexicographers should carefully consider the type of data boxes to include in that search zone.

5.4.2.3 Data boxes as article windows

Lexicographers often have innovative ideas regarding the way of presenting data in their dictionaries. In the OALD several types of data boxes are used, with some types reserved for specific types of data. An article window is a type of data box reserved for a single type of data, i.e. word families, as seen in figure 8:

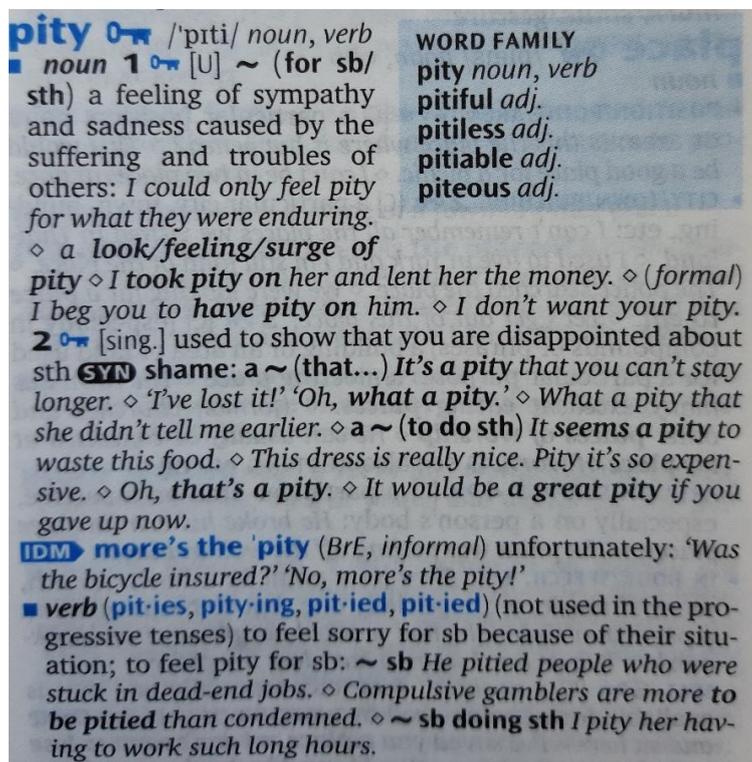


Figure 8: An article window in the OALD

In the front matter text "Key to dictionary entries" it is said that: "Word families show words related to the headword." Although articles could in principle have windows in different article positions, e.g. top right, top centre, top left, middle right, middle centre, middle left, bottom right, bottom centre and bottom left, the OALD primarily uses the top right corner for its window data boxes. However, the layout of a dictionary article on the page could also have an influence on the position of the window. The editors of the OALD are consistent in always placing data boxes presented in article windows in the top right corner of a text block, albeit not necessarily of the article. Where an article

commences at the bottom of a column there could be a lack of space for a window in the top right corner of the article, as seen in figure 9.

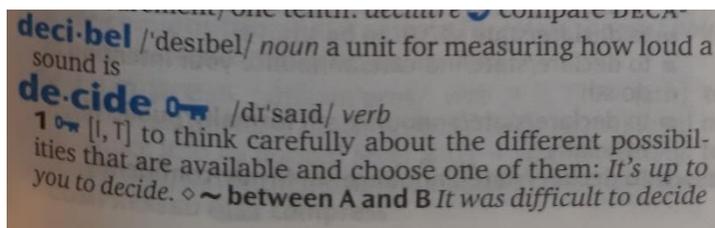


Figure 9: Article in the OALD

However, the window is then positioned in the top right corner of the text block, containing the remainder of the article, that starts in a new column, as seen in figure 10.

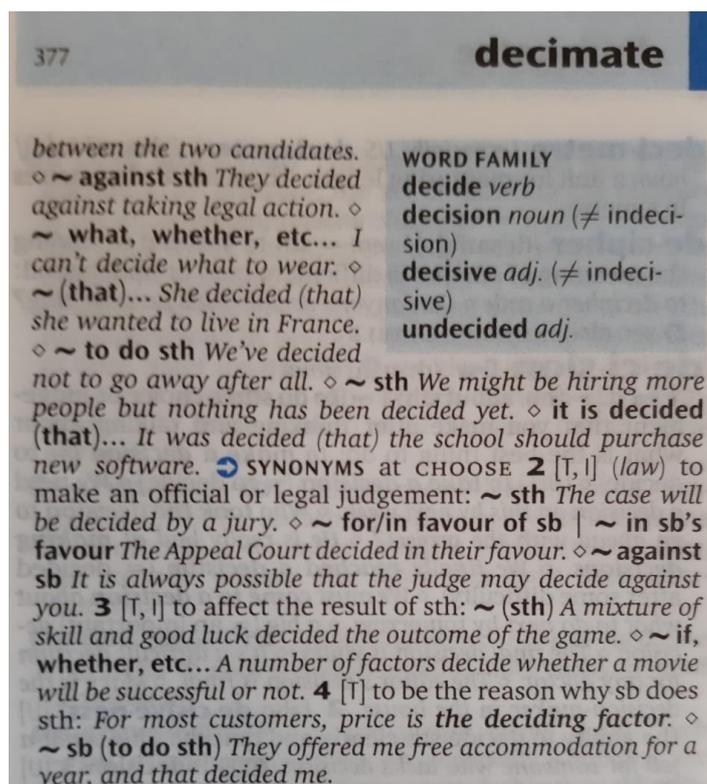


Figure 10: Article window in the OALD

Technically such a window occupies the right centre of the article. From a layout perspective it could be argued that the OALD always presents its windowed data boxes in the top right corner of a text block, preferably but not always the top right corner of the article.

A brief identification of word families as found in the article windows can be valuable assistance to the target users of this dictionary. Users should be aware of this assistance but unfortunately the OALD does not provide an outer text in the front matter section with a list of articles that contain these windows. In spite of the focus that a data box puts on this data, there is no defined search path besides the main alphabetical outer access structure to guide users to these articles that contain article windows. This lack of predictable access impedes the added value of these article windows.

5.4.2.4 Data boxes within a comment of an article

Whether an article-internal data box contains an item or an item text, the contents of that data box is presented as part of a treatment procedure. This procedure is directed at one or more specific treatment units that constitute the address within this procedure of addressing. The positioning of article-internal data boxes directly reflects the relevant addressing relations in which data boxes can become involved.

Although lemmatic addressing, i.e. with the lemma as address, is the most frequently used addressing procedure in dictionaries, other types of addressing also prevail, especially non-lemmatic addressing. Various aspects of addressing are dealt with in Hausmann and Wiegand (1989), Louw and Gouws (1996), Gouws and Prinsloo (2005), Wiegand (2006; 2011) and Gouws (2014).

Lexicographers utilise data boxes to emphasise certain entries not accounted for by items in search zones of the obligatory microstructure of a dictionary. These items or item texts do not always have the lemma as address, but often occur within one of the comments of an article, especially the comment on semantics or a subcomment on semantics, where they participate in relations of non-lemmatic addressing. A less optimal article-internal positioning of a box of which the entry or entries are not addressed at all the senses of the word represented by the lemma sign of the article may diminish the added value the data box is supposed to have. In figure 11 from the *Handwoordeboek van die Afrikaanse Taal* (HAT) (Odendal and Gouws 2005) the data box is included as postcomment at the end of the article. The contents of the box only applies to the use of the word *instansie* in its second and third senses. From this presentation it is not clear that the box is not addressed at the first sense.

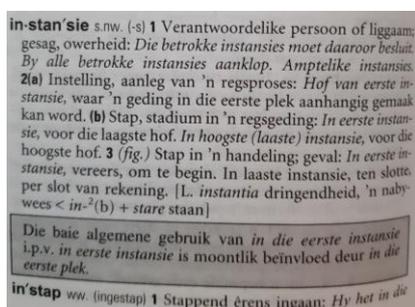


Figure 11: Article of the lemma *instansie* in the HAT

Figure 12 is a partial article of the lemma *open*² from the *Longman Exams Dictionary* (Summers 2006), (henceforth abbreviated as LED). This article contains two data boxes, so-called study notes. Both these data boxes are presented within single subcomments on semantics for the first and the second sense of the polysemous word *open*. By positioning them there the lexicographer ensures that the user can know what the exact address of the data box is:

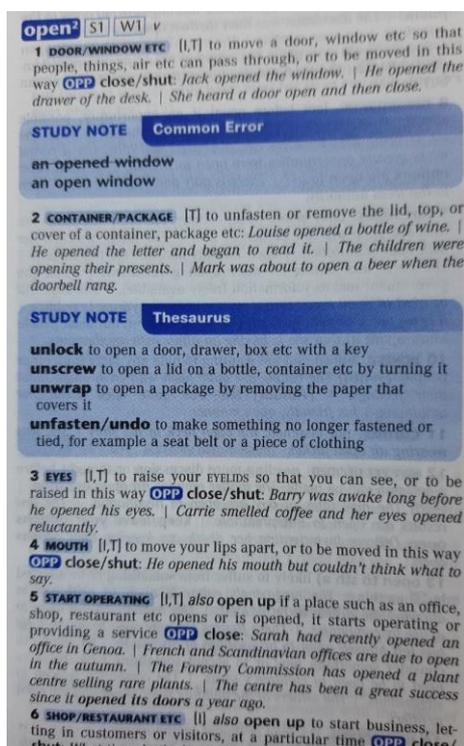


Figure 12: Partial article of the lemma *open* in the LED

In the *Longman South African School Dictionary* (Bullon et al. 2007, (henceforth abbreviated as LSASD)) the article of the lemma sign *name*¹ has three subcomments on semantics. The thesaurus data box is positioned at the end of the first subcomment on semantics. The thesaurus guidance in the data box is only applicable to the first sense of the word *name*.

name¹ /naym/ *noun* **1** the word that people call someone or something: *I can't remember the name of the hotel.*

Thesaurus

first name your personal name, not your family name

surname/family name/last name the name that everyone in your family has, not your personal name

middle name the name between your first name and your surname

full name your complete name: *He wrote his full name, James Campbell Giffen, in the space on the form.*

maiden name a woman's family name before she got married and changed it: *Mrs. Robertson's maiden name was Hansen.*

nickname a name your friends and family use for you, not your real name: *He runs so fast that his nickname is "Flash".*

2 the opinion that people have of a person or organization (**SYNONYM reputation**): *The company has a very good name.* **3** (informal) a famous person, company etc: *The film made him a big name in Hollywood.*

Figure 13: Article of the lemma *name*¹ from the LSASD

Figure 14 from BW shows a data box that only has the first sense of the word *opmerking* as its address. Such a procedure of immediate addressing is to the benefit of the target user of the dictionary and demands less dictionary using skills compared to correctly interpreting the address of a data box that is not positioned in close vicinity to its address, even if the user was able to identify the appropriate sense:

opmerking (opmerkings) *nw.* ① Wanneer 'n mens 'n opmerking oor iets maak, sê jy iets, of jy lewer kommentaar daaroor. *Mense hou oor die algemeen nie daarvan dat iemand persoonlike opmerkings oor hulle maak nie. Daar was heelwat positiewe opmerkings oor die betrokke dosent se klasse. Die opmerking wat hy oor haar gewig gemaak het, was heeltemal taktloos.*

Opmerking en aanmerking kan mekaar vervang wanneer iets negatiefs gesê word: *Hy maak altyd aanmerkings/opmerkings oor die kos. Opmerking* kan egter ook vir positiewe of neutrale kommentaar gebruik word.

② 'n Opmerking is dit wat 'n mens opmerk 1. *Die student se opmerking dat daar 'n makliker manier is om die som uit te werk, het sy dosent heeltemal verstom. Al haar opmerkings oor die betrokke gebied se plantegroei word vir haar navorsing aangeteken.*

opmerklik (opmerklike; opmerkliker, die opmerklikste) *adj.* Iets wat opmerklik is, is maklik om

Figure 14: Article of the lemma *opmerking* from BW

Where the treatment in a data box is directed at only one of the senses of a polysemous word or at only one of the items in the article and the data box is not positioned in such a way that an immediate addressing relation is possible, the lexicographer should clearly indicate what the address of the contents of the data box is. Such an approach is seen in figure 15, the article of the lemma *mad* in the *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary* (Pheiffer 2007), (henceforth abbreviated as OSD):

mad *adjective* (**madder, the maddest**)
1 waansinnig. ♦ The author went **mad** towards the end of his life. *Na die einde van sy lewe het die skrywer waansinnig geraak.*
2 gek. ♦ You are **mad** to go out in this rain! *Jy is gek om in hierdie reën uit te gaan!*
3 (*informal*) boos. ♦ My dad gets really **mad** when I don't do my homework. *My pa raak regtig boos as ek nie my huiswerk doen nie.*
4 (*informal*) dol, gek. ♦ Many young girls are **mad** about that actor. *Baie jong meisies is gek oor daardie akteur.* ♦ Spiro is **mad** about cooking. *Spiro is dol oor kook.*
drive someone mad iemand gek maak.
■ **madness** *noun* gekheid.

Mad in the first sense may be offensive. Use "mentally ill" instead.

Figure 15: Article of the lemma *mad* in the OSD

In this article with its four subcomments on semantics, the data box has been included as postcomment but it is clearly stated that it is addressed at the first sense.

Where a word can be used in more than one part of speech, the lexicographic treatment typically results in a complex article with a partial article for the occurrence of each of these parts of speech (Wiegand and Gouws 2011: 243). Data boxes could then be positioned in a search zone of one of the partial articles so that the other part of speech occurrence of the word falls beyond the scope of the data box. This can be seen in figure 16 from BW where the data box is positioned at the end of the first partial article of the article with the lemma *smaak* as guiding element.

smaak I (-) nw. ① 'n Mens se smaak is die sintuig wat dit vir jou moontlik maak om te proe. *Die warm koffie het my mond so gebrand dat ek glad nie meer smaak gehad het nie.* ② (**smake**) Die gewaarwording wat 'n mens kry wanneer jy iets proe, word die smaak daarvan genoem. *Die kos het 'n gebrande smaak. Ek verkies 'n soet smaak bo 'n suur smaak.* ③ Om iets met smaak te doen, beteken dat jy dit met lus en oorgawe doen. *Hulle het met smaak die lekker kos geëet. Dié skinderstorie sal jy nou seker met smaak gaan oortel.* ④ Iemand se smaak is die oordeel waarvolgens hy keuses maak om te wys waarvan hy hou of nie hou nie. *As dit by klere kom, verskil haar smaak heeltemal van dié van haar ma. Sy smaak vir musiek staan my glad nie aan nie. As almal dieselfde smaak gehad het, was die lewe maar baie eentonig. Dit help nie om baie geld maar swak smaak te hê nie.* **UITDR.:** Daar is 'n smakie aan: dit is nie meer goed/vars nie. *Daar is 'n smakie aan die melk, daarom moet jy dit liever nie gebruik nie.* In die smaak val: Daarvan hou. *Die predikant se preek van vanoggend het in almal se smaak geval.*

As jy sê dat iemand goeie/slegte smaak het, gee jy daardeur jou oordeel oor so 'n persoon se smaak.

II (het gesmaak) ww. ① Om te sê hoe iets smaak, hang af van die manier waarop dit jou smaak-sintuig prikkel en jy die smaak daarvan ervaar. *My ma se kos sal altyd vir my lekker smaak. As jy te veel sout op die vleis gooi, sal dit sleg smaak.* ② Om iets te smaak, is om dit te ondervind. *Die weduwee het al baie verdriet in haar lewe gesmaak. Saam met die pyn het sy darem ook vreugde gesmaak.*

Figure 16: Article of the lemma *smaak* from BW

5.5 Phased-in inner texts as data boxes

5.5.1 Article-external text boxes

Data boxes can also be accommodated in the word list of dictionaries as phased-in inner texts. These data boxes are immediate constituents of the article stretches of the word list and function as article-external data carriers. Data boxes presented as phased-in inner texts typically have a connection with a lemma that is the guiding element of an article occurring in the specific article stretch into which the data box is phased in. These data boxes can be items or item texts and participate in either primary or secondary addressing relations. In addition to having the lemma that is in close proximity as address, these data boxes can also participate in relations of remote addressing, cf. Louw and Gouws (1996), where one or more lemmata in the same or often in other article stretches can be the address. As is the case with article-internal text boxes, different types of data can be presented in these data boxes, cf. Wiegand et al. (2017: 140-144 (WLWF)).

Phased-in inner texts could be constituents of the word list of a dictionary with a single alphabetical macrostructure or constituents of different word lists of dictionaries with poly-alphabetical macrostructures with vertical or horizontal parallel alphabetical access structures, cf. Wiegand (1989: 402) and Wiegand and Gouws (2013: 88). In the remainder of this section the focus will primarily be on phased-in inner texts in single macrostructures but reference will briefly be made to their occurrence in dictionaries with poly-alphabetical macrostructures with vertical parallel alphabetical access structures.

5.5.2 Phased-in inner texts in the primary macrostructure

In an expanded word list different types of phased-in inner texts can be included to split the article stretches.

5.5.2.1 Data boxes phased out of the article

In paragraph 5.4.2.3 it was mentioned that the layout of a dictionary page could influence the positioning of text boxes presented as article windows. In a comparable way the structure and layout of a dictionary article and the physical size of a specific item could lead to such an item not being included as microstructural constituent but rather in an article-external data box. This is where a phased-in inner text is the result of an item phased out of an article and phased into the article stretch as a data box.

Due to the position of an article in a column and on a page and due to their size pictorial illustrations do not always fit into the boundaries of an article as search area. The application of a well-defined data distribution structure

may then lead to a situation where such a pictorial illustration is phased out of the article and presented within a data box included as phased-in inner text.

The MWALD often contains pictorial illustrations as items in its articles, as can be seen in figure 17, the article of the lemma *glove*:

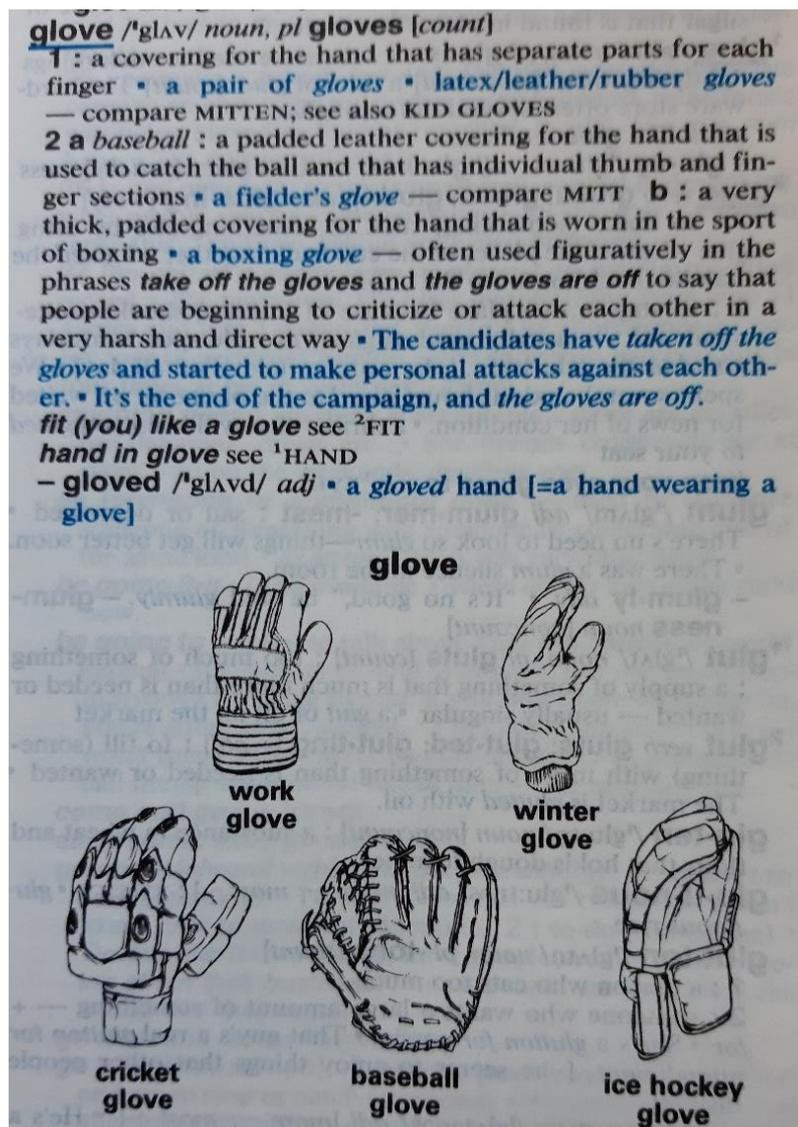


Figure 17: Article of the lemma *glove* from the MWALD

No pictorial illustration is presented within the borders of the article of the lemma sign *deer* in the MWALD. However, an illustration does occur on the same page, in an article-external position, as can be seen in figure 18 where the illustration is entered across two columns and splits the partial article stretch between the articles of the lemmas *defamatory* and *defame*:

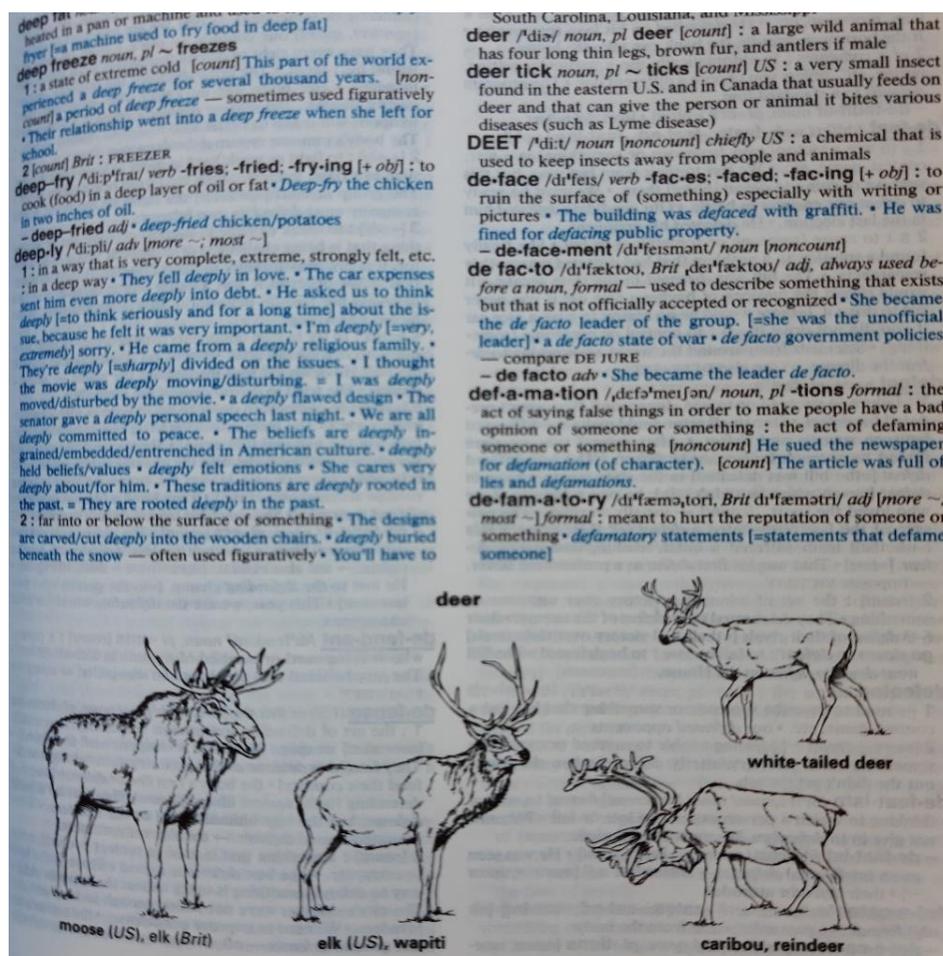


Figure 18: Pictorial illustration in the MWALD

It could have been problematic for the lexicographers to fit this illustration into the article of the lemma *deer*. Consequently it has been phased out of the article and positioned as a phased-in inner text in the article stretch, where its title has the same form as the article of the lemma *deer*. Because the pictorial illustration

is in close proximity of the article of the lemma *deer*, this article has no reference to the pictorial illustration.¹

In dictionaries a distinction can be made between single and synopsis articles, cf. Bergenholtz, Tarp and Wiegand (1999: 1780). The treatment in a single article is directed only at the lemma of that article, whereas the treatment in a synopsis article also contains data relevant for the treatment of the lemmata of other articles. Data boxes also have a single or a synopsis character. This applies to article-internal as well as article-external data boxes. Figure 19, the article for the lemma *lend* in the *Tweetalige aanleerderswoordeboek/Bilingual Learner's Dictionary* (Du Plessis 1993), (henceforth abbreviated as TAW), has an article-internal data box presented as postcomment. The contents of this data box apply to the lemma *lend* but it is also remotely addressed at the lemma *borrow*. Figure 20 shows the article of the lemma *borrow* with an item giving a cross-reference to *lend*.

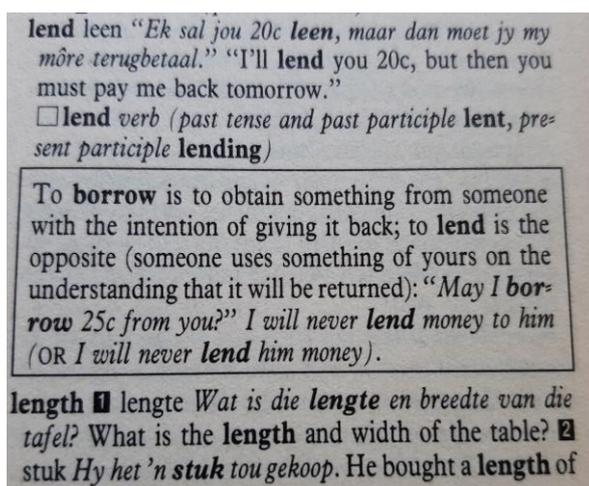


Figure 19: Article of the lemma *lend* in the TAW

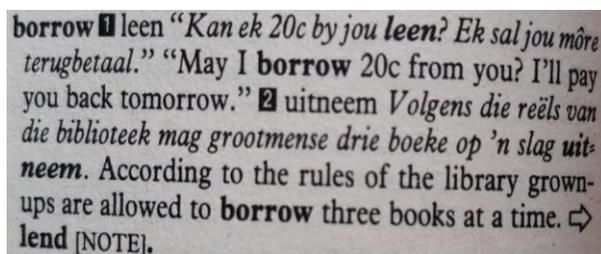


Figure 20: Article of the lemma *borrow* in the TAW

The data box in the article of the lemma *lend* is a synopsis data box. In a similar way the data box titled *deer* is a synopsis data box. The pictures are also relevant for the treatment of lemmata like *moose*, *elk* and *caribou*. Consequently the articles of these lemmata have an item giving a cross-reference "see picture at DEER".

Positioning the contents of the data box *deer* not within the article but as an article-external phased-in inner text, can probably best be motivated on article and page layout grounds. A similar use of data boxes where the data have been phased out of an article and into a data box functioning as phased-in inner text, is found in the *Cambridge International Dictionary of English* (Procter 1995), (henceforth abbreviated as CIDE). Figure 21 shows the framed data box *Molluscs* presented as a phased-in inner text. The contents of this box could have been included in the nearby article of the lemma *mollusc*. The layout of the column might have been disturbed by this illustration and consequently it was phased-out into a data box in the partial article stretch but still in close vicinity of the lemma sign. This is also a synopsis data box that functions as the cross-reference address for cross-reference items in the articles of lemmata like *cuttle fish*, *octopus* and *snail*:

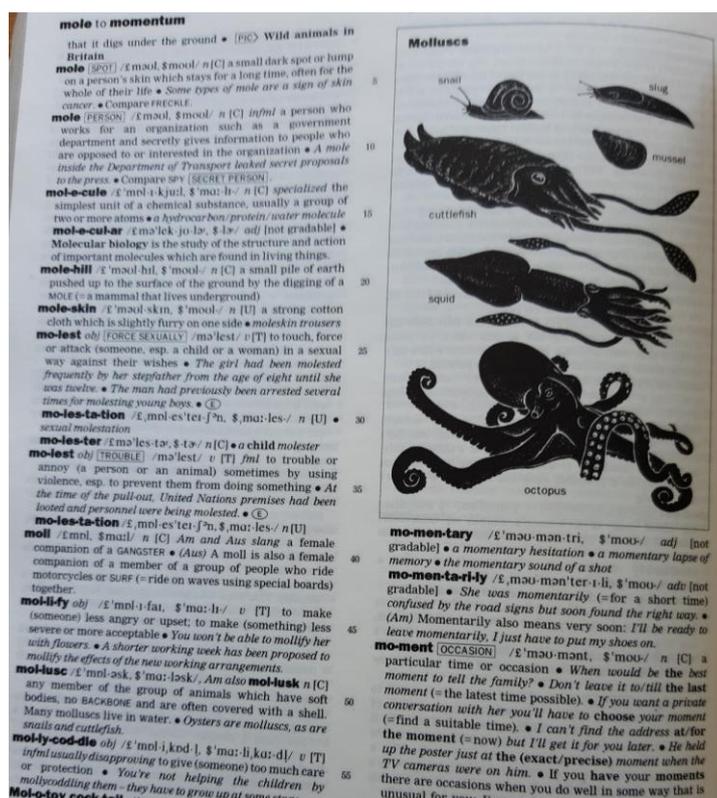


Figure 21: Data box for *Molluscs* in the CIDE

Phasing-out procedures do not only target non-boxed data that could have been included in a search zone of an article and allocate them to an article-external data box. These procedures can also target data boxes that could have had an article-internal position and phase them out to function as article-external data boxes and phased-in inner texts. The OALD frequently includes a text box containing synonyms in a postcomment position of an article, as can be seen in figure 22, the article of the lemma *painful*:

pain·ful 0-w /'peɪnfl/ *adj.*
1 0-w causing you pain: *Is your back still painful?* ◊ *a painful death* ◊ *My ankle is still too painful to walk on.*
2 0-w ~ (for sb) (to do sth) | ~ (doing sth) causing you to feel upset or embarrassed: *a painful experience/memory* ◊ *Their efforts were painful to watch.* **3** 0-w unpleasant or difficult to do **SYN** trying: *Applying for jobs can be a long and painful process.*

Synonyms

painful
sore • raw • inflamed • excruciating • burning • itchy
These words all describe sth that causes you physical pain.

painful causing you physical pain. **NOTE** Painful can describe a part of the body, illness, injury, treatment or death: *Is your knee still painful?* ◊ *a series of painful injections* ◊ *a slow and painful death*

sore (of a part of the body) painful and often red, especially because of infection or because a muscle has been used too much: *a sore throat* ◊ *Their feet were sore after hours of walking.*

raw (of a part of the body) red and painful, for example because of an infection or because the skin has been damaged: *The skin on her feet had been rubbed raw.*

inflamed (of a part of the body) painful, red and hot because of an infection or injury: *The wound had become inflamed.*

excruciating extremely painful. **NOTE** Excruciating can describe feelings, treatments or death but not parts of the body: *an excruciating throat/back/knee.*

burning painful and giving a feeling of being very hot: *She felt a burning sensation in her throat.*

itchy giving an uncomfortable feeling on your skin that makes you want to scratch; having this feeling: *an itchy rash* ◊ *I feel itchy all over.*

PATTERNS

- sore/inflamed/itchy eyes
- raw/inflamed/itchy skin
- a painful/an excruciating death
- a painful/burning sensation
- excruciating/burning pain

Figure 22: The article of the lemma *painful* from the OALD

This can be regarded as the default positioning of a data box that contains synonyms. The data box for synonyms is often phased out of the article and positioned as a data box and phased-in inner text in the textual position imme-

diately preceding the article. This is probably due to article and page layout considerations, a column break and the approach not to allow column or page breaks to divide a data box into two sections but always to present such a data box as an uninterrupted text block. This can be seen in figure 23, the article of the lemma *surprise*:

Synonyms

surprise
startle • amaze • stun • astonish • take sb aback • astound
These words all mean to make sb feel surprised.

surprise to give sb the feeling that you get when sth happens that you do not expect or do not understand, or sth that you do expect does not happen; to make sb feel surprised: *The outcome didn't surprise me at all.*

startle to surprise sb suddenly in a way that slightly shocks or frightens them: *Sorry, I didn't mean to startle you.* ◊ *The explosion startled the horse.*

amaze to surprise sb very much: *Just the huge size of the place amazed her.*

stun (rather informal) (often in newspapers) to surprise or shock sb so much that they cannot think clearly or speak

astonish to surprise sb very much: *The news astonished everyone.*

AMAZE OR ASTONISH?
These two words have the same meaning and in most cases you can use either. If you are talking about sth that both surprises you and makes you feel ashamed, use **astonish**: *He was astonished by his own stupidity.*

take sb aback [usually passive] (especially of sth negative) to surprise or shock sb: *We were rather taken aback by her hostile reaction.*

astound to surprise or shock sb very much: *His arrogance astounded her.*

PATTERNS

- It surprises sb/startles sb/amazes sb/stuns sb/astonishes sb/takes sb aback/astounds sb
- to surprise/startle/amaze/stun/astonish/astound sb that...
- to surprise/amaze sb what/how...
- to surprise/startle/amaze/stun/astonish/astound sb to know/find/learn/see/hear...
- to be surprised/startled/stunned into (doing) sth

sur·prise 0-ɪ /sə'praɪz; NAmE sər'p-/ noun, verb

■ **noun** 1 0-ɪ [C] an event, a piece of news, etc. that is unexpected or that happens suddenly: *What a nice surprise!* ◊ *a surprise attack* ◊ *There are few surprises in this year's budget.* ◊ *I have a surprise for you!* ◊ *It comes as no surprise to learn that they broke their promises.* ◊ *Her letter came as a complete surprise.* ◊ *There are lots of surprises in store for visitors to the gallery.* ◊ *Visitors to the gallery are in for a few surprises.* 2 0-ɪ [U, C] a feeling caused by sth happening suddenly or unexpectedly: *a look of surprise* ◊ *She looked up in surprise.* ◊ ~ (at sth) *He gasped with surprise at her strength.* ◊ ~ (at seeing, hearing, etc.) *They couldn't conceal their surprise at seeing us together.* ◊ *I got a surprise when I saw the bill.* ◊ *Much to my surprise, I passed.* ◊ *To everyone's surprise, the plan succeeded.* ◊ *Imagine our surprise when he walked into the*

Figure 23: Data box for the lemma *surprise* from the OALD

The biggest section of the article of the lemma *surprise* appears in the left column with a brief section of this article continuing in the right column. Had this section also been included in the left column there would probably not have been room for an uninterrupted data box as postcomment in that column. Consequently the data box was phased out to an article-external position, preceding the article. This data box is not positioned as precomment because it lies outside the article borders.

Because the planned data distribution structure of a dictionary needs to be executed in a meticulous and consistent way, it is not plausible to deviate from such a structure for column or page break or article or page layout reasons. Such a deviation impedes the access of users to the required data. The structural inconsistency resulting from the phasing out of items to article-external data boxes can be seen in the treatment of the lemma *save* in the OALD. In the OALD synonyms are frequently provided in data boxes, as seen in the treatment of *painful*. Irrespective of the degree of complexity of an article (whether it is a single or an n-fold complex article) or the number of senses treated in one or more subcomments on semantics, article-internal data boxes containing synonyms are presented as postcomments. Although such a positioning of the data box may cause addressing uncertainty in complex articles or articles with polysemous lexical items as lemmata, knowledgeable users will become familiar with this positioning of data boxes if this execution of the data distribution structure is done in a consistent way.

Column and page breaks result in deviations of this execution of the default data distribution structure. The treatment of the lemma *save* is presented in a threefold complex article with partial articles for the use of the word *save* as different parts of speech, namely as verb, noun, preposition and conjunction. The partial article treating the occurrence of *save* as a verb accounts for eight polysemous senses of this word, presented in different subcomments on semantics. The treatment is interrupted by a column and page break within the eighth subcomment on semantics. For the senses presented in the first two subcomments on semantics synonyms are provided in data boxes. The data box containing the synonyms for the first sense is included article-internally, according to the default data distribution of this type of text constituent, as a postcomment. However, the data box containing synonyms for the second sense has been phased out to a position immediately preceding the article of the lemma *save* as an article external phased-in inner text. This can be seen in figure 24 that shows the article as it is spread over the right column of the first and the left column of the second page:

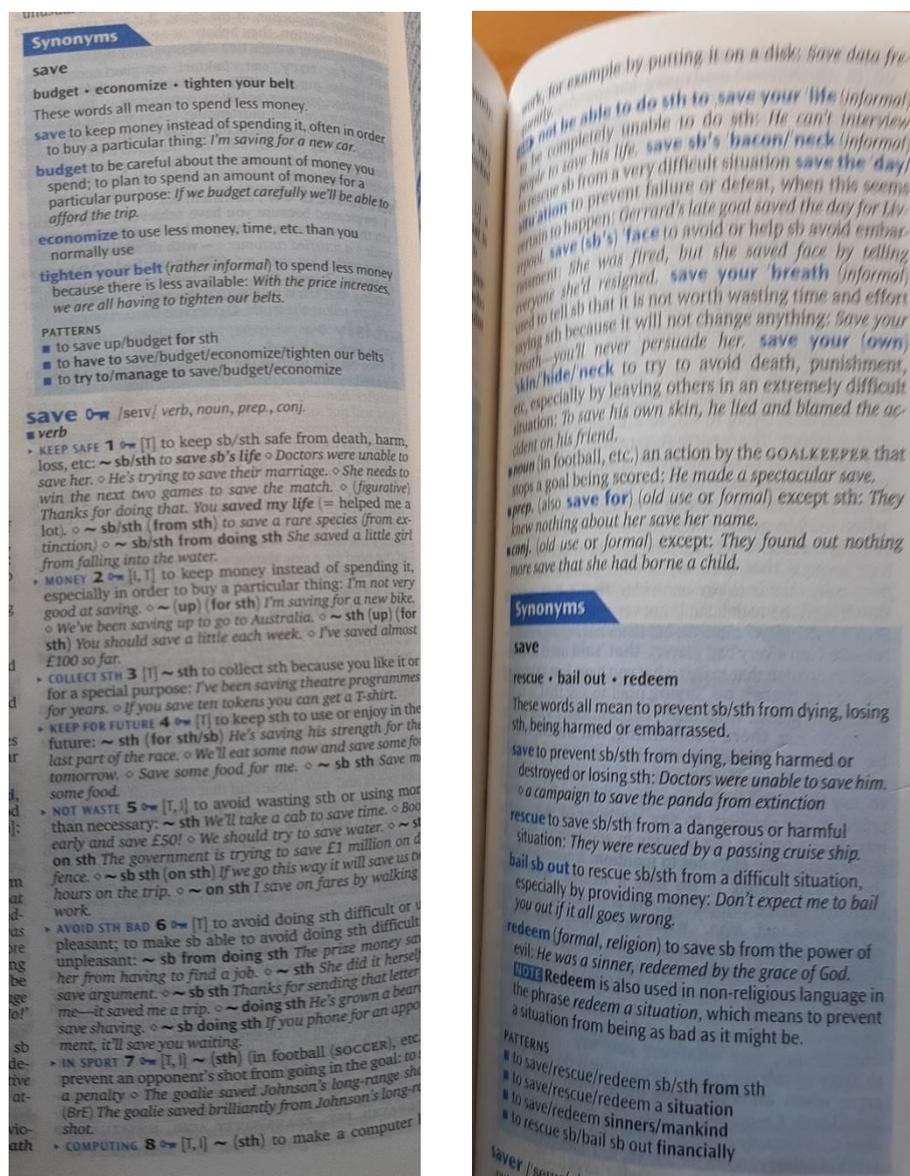


Figure 24: Article of the lemma *save* from the OALD

Successfully accessing this data will be challenging for even a knowledgeable user of this dictionary. This is a typical situation where procedures in the lexicographic practice, i.e. allowing column and page breaks to impede the execution of the default data distribution structure, has resulted in theory-determined

dictionary structure problems. A more user-friendly way of dealing with these two data boxes would have been to either add them both as postcomments (reflecting the order of the senses at which they are addressed) or, even better, include each data box in a slot in the relevant subcomment on semantics. This would have enabled procedures of immediate addressing that would have assisted users in a quicker and better retrieval of information.

In some articles the inconsistency in the positioning of synonym data boxes in the OALD on column and page break grounds, results in such an article-internal data box not being included as the default postcomment but rather within the comment on semantics. The lemma *reason* has a single complex article with partial articles for its occurrence as noun and as verb. The article is interrupted by a column and page break, as seen in figure 25:

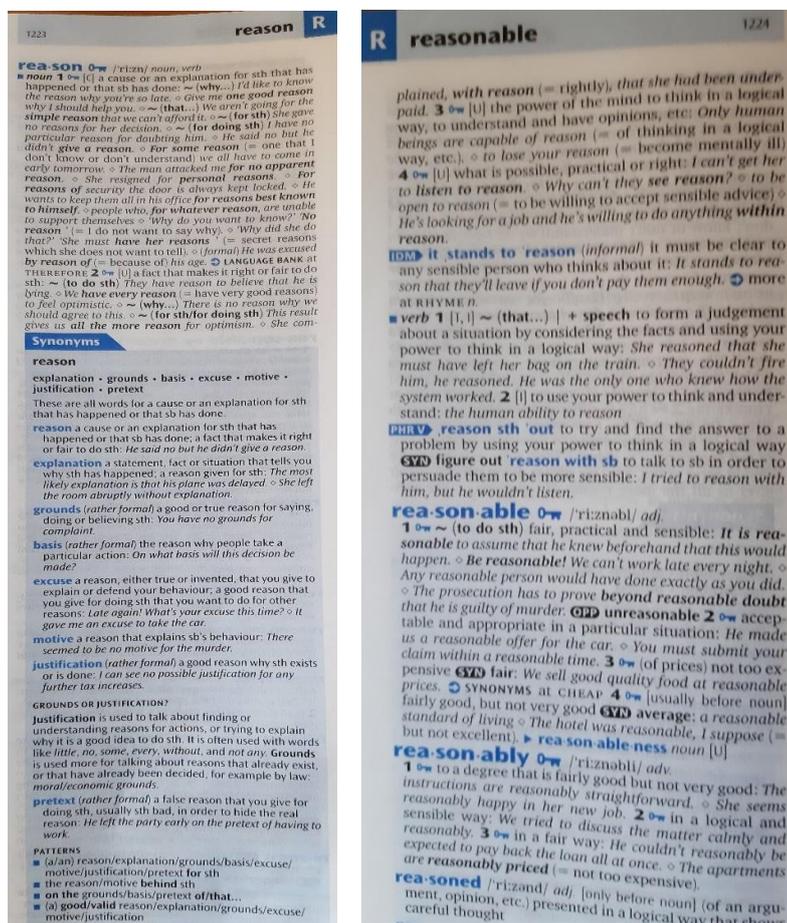


Figure 25: Article of the lemma *reason* from the OALD

The comment on semantics of the partial article in which the occurrence as noun is treated, makes provision for four polysemous senses of this word. A data box is provided that contains synonyms for the first sense of this word. In order to present an uninterrupted text block the data box containing synonyms is positioned within the comment on semantics of the partial article in which the occurrence of *reason* as a noun is treated. It is positioned at the end of the first column on the first page on which the article is given. The data box occurs within the subcomment on semantics in which the second sense is treated, even though the synonyms are addressed at the first sense of the word. This positioning defies logical and consistent addressing relations.

5.5.2.2 Article-external data boxes resembling articles or article-internal items

Data boxes included in the word list as phased-in inner texts are often inserted between two other articles in a partial article stretch where it adheres to the alphabetical ordering. These data boxes typically have a guiding element that looks like a lemma and the box contains a rudimentary treatment of that word. In the TAW the word *fridge* had not been selected as a lemma candidate. However, the lexicographer must have felt the need to make the users of this learner's dictionary aware of that informal English word. Consequently, a data box was included as a phased-in inner text and inserted in its alphabetical position in the article stretch between the articles of the lemmata *Friday* and *friend*, as seen in figure 26, with the frame clearly identifying it as a data box and not a default article:

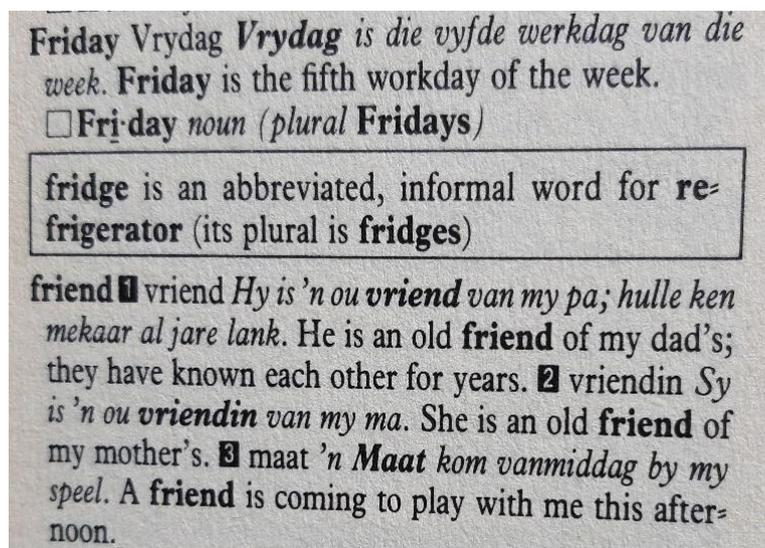


Figure 26: A data box for *fridge* in the TAW

Figure 27, a partial article stretch from *Groot Woordeboek/Major Dictionary* (Eksteen et al. 1997); (henceforth abbreviated as GW), contains the data box with *percentage point* as its guiding element. This box follows the article niche attached to the article of the lemma *percentage*. This niche also contains the sublemma *percentage point* in a condensed format. In the niche the treatment of the lemma is restricted to an item giving a translation equivalent. The lexicographer wants to increase the assistance to users regarding this lemma and consequently employs the phased-in data box to supply the expression *percentage point* with a paraphrase of meaning.

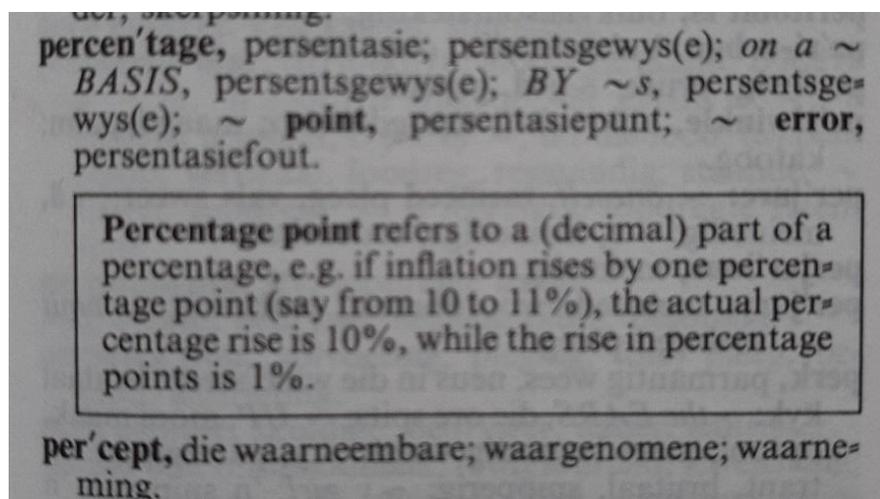


Figure 27: Data box for percentage point from the GW

Data boxes do not always look like articles but can occur in the alphabetical position of a word that has not been included as lemma, in order to make the user aware of some relevant feature of that word. In Afrikaans the word *huidiglik* (=presently) is often used. However, from a linguistic perspective the use of this word is not approved. The word does not qualify for inclusion as a lemma, but the lexicographers of the monolingual Afrikaans school dictionary the *Oxford Afrikaanse Skoolwoordeboek* (Louw 2012), (henceforth abbreviated as OSW) would like to make their users aware of the fact that the word should not be used. The word *tans* should rather be used. In the alphabetical slot where *huidiglik* would have been entered had it been a lemma in this dictionary, the lexicographers include an article-external data box, cf. figure 28, to sensitise users that they should not use the word *huidiglik*.

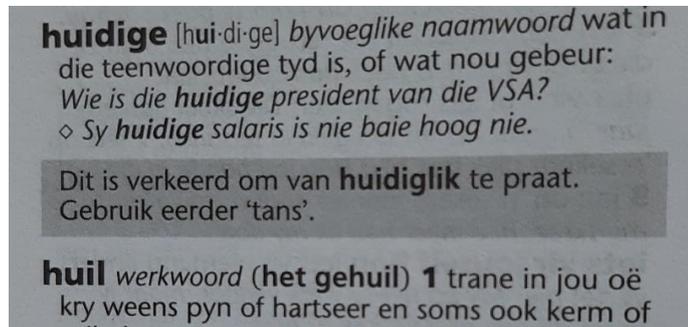


Figure 28: Data box for *huidiglik* in the OSW

The data box in figure 28 is not a postcomment in the article of the lemma *huidige* but rather a phased-in inner text.

Article-external data boxes often include data that do not represent a data type that belongs to the default microstructural items but it does add to the treatment of a given lemma or lemma cluster. Figure 29 presents a partial article stretch from HAT:

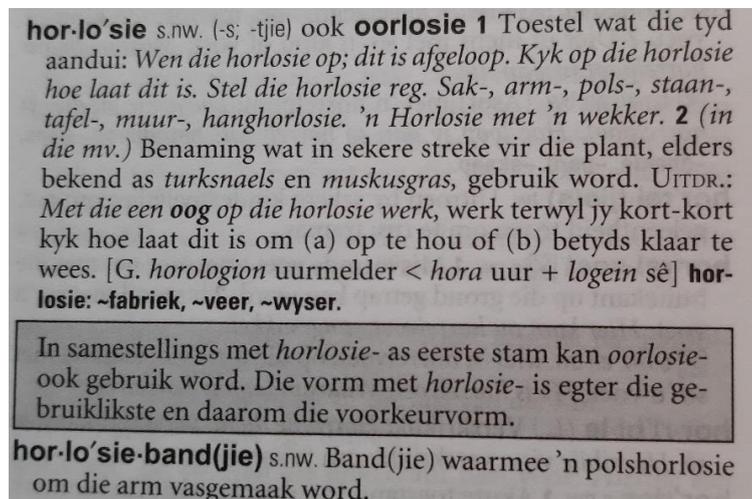


Figure 29: Partial article stretch from the HAT

In the comment on form of the article of the lemma *horlosie* (watch) it is indicated that *oorlosie* is a variant form of *horlosie*. Attached to the article of this lemma is a lemma cluster, presented as a first level nest that includes a condensed presentation of compounds with *horlosie-* as first stem. Between this nest and the following main lemma a data box has been inserted with guidance

regarding the formation of compounds with *horlosie-* as first stem. Although this data is relevant to the main lemma *horlosie-* and to the sublemmata presented in the nest, the data box is positioned outside the article of the main lemma as well as the nest of sublemmata and therefore it is a phased-in inner text.

5.5.2.3 Other article boxes included as phased-in inner texts

A variety of other types of data boxes also function as phased-in inner texts in dictionaries. A typical feature of many of these data boxes is their synopsis character whilst the data presentation in others complements and expands the default data coverage of a specific lemma. These data boxes convey data regarded by the lexicographer as important enough to include them in salient search venues like data boxes. The added value of these data boxes is unfortunately too often impeded by insufficient access guidance.

In close proximity of the article of the lemma *plural* in CIDE one finds the data box titled "Plural of nouns", presented in figure 30:

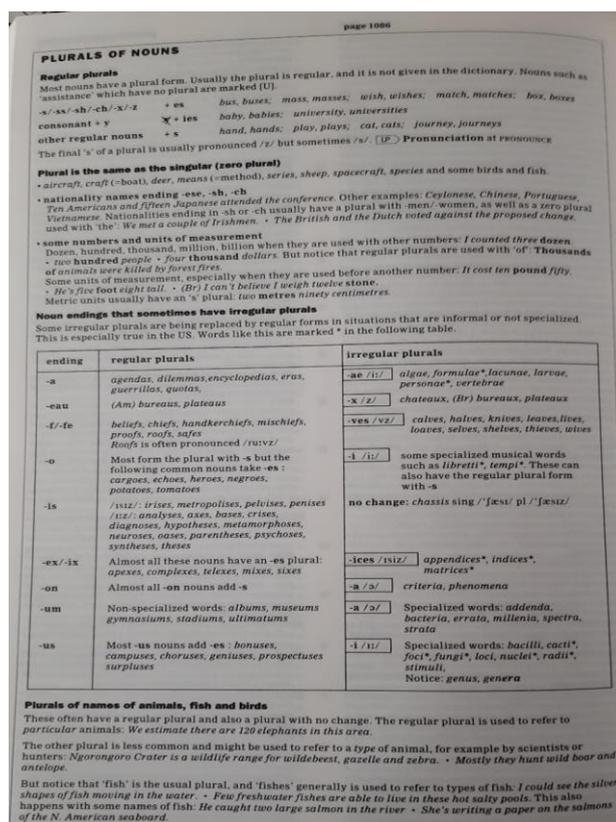


Figure 30: Data box in the CIDE

The treatment in this data box is not addressed at any single lemma but it adds value to the treatment of all nouns that take plural. The problem is that it is very difficult for users in need of this guidance to know where to find the help. CIDE has a back matter text that provides an alphabetical index with relevant page numbers of all pictures, language portraits and lists of false friends included in the word list. The list includes an entry "Plurals" but no entry "Plural of nouns", the title of the data box in figure 30. It will remain a challenge for the target users of this dictionary to optimally benefit from this data box. They will probably only have access to this data box if they accidentally consult that page, seeing that the lemma *plural* is on the previous page and its article has no cross-reference to the data box.

Having consulted the lemmata *meet* and *meeting* in the CIDE a user has to turn the page to find the data box "Meeting someone", given in figure 31:

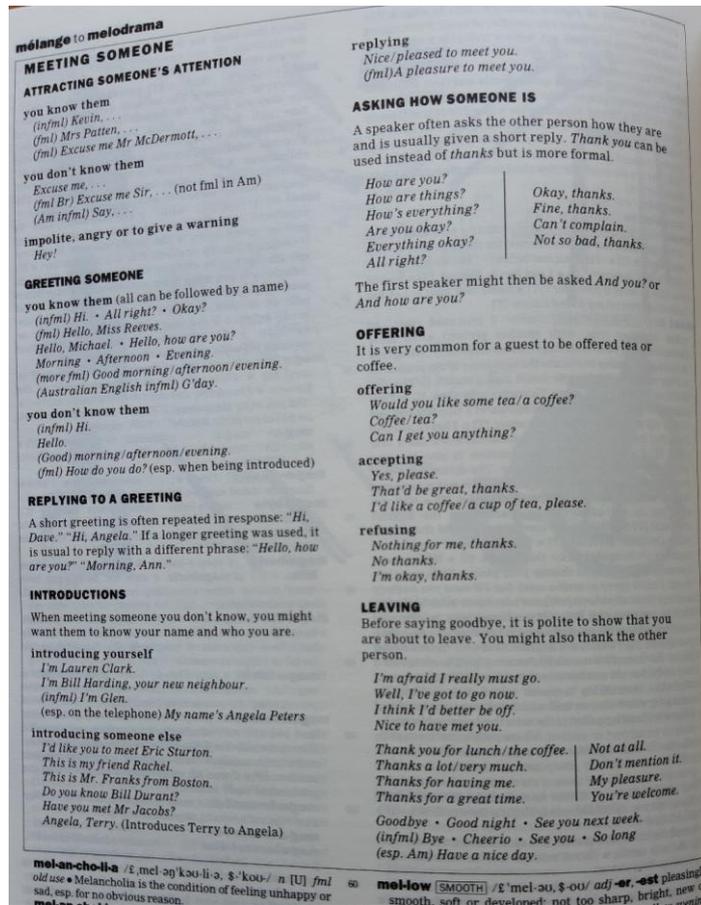


Figure 31: Data box for *Meeting someone* from the CIDE

The contents of a data box like this one can hardly be included in the article of the corresponding lemma. Data boxes like these elevate the quality of the lexicographic treatment.

Even if an article in a learner's dictionary displays an extended obligatory microstructure lexicographers have to restrict the extent of the data presented in the different search zones and should refrain from procedures of data overload. If lexicographers want to respond to the needs of the users of a learner's dictionary, like CIDE, for some linguistic guidance typically contained in text books, they can use data boxes included as phased-in inner texts to convey this kind of data. Figure 33, a partial article stretch from CIDE, contains articles of the lemmata *homograph* and *homophone*. In each one of these articles there is an item giving a cross-reference to the language portrait (LP) *Homophones and homographs*.

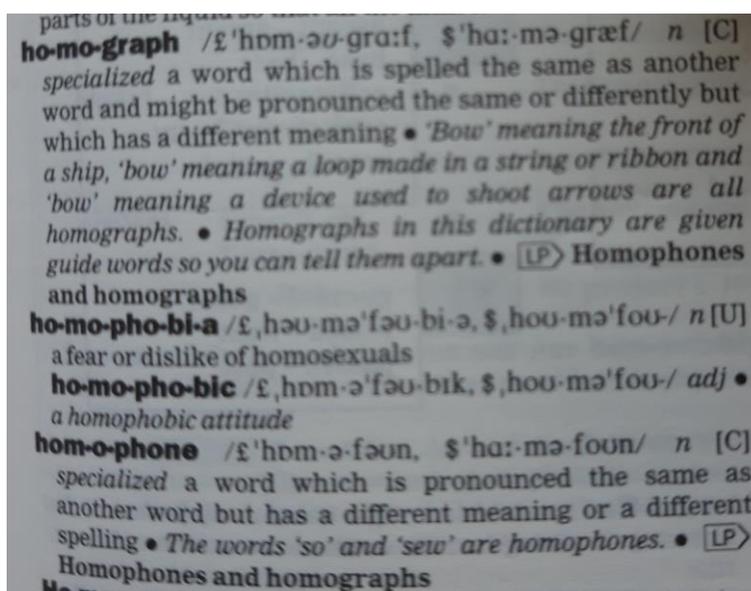


Figure 33: Partial article stretch from the CIDE

This language portrait is presented as the data box seen in figure 34. This data box can also be accessed via the index in the back matter text. It has a synopsis character that could assist users in clearly distinguishing between homophones and homographs.

HOMOPHONES AND HOMOGRAPHS

A homophone is a word with the same pronunciation as another word, but with a different spelling and meaning. For example, *son* and *sun* are both pronounced /sʌn/. Usually the meaning of words like this is clear from the sentences they are in: *My son is a doctor.* • *The sun is really hot today.* Homophones are sometimes used humorously, for example in newspaper headlines. For example *The cent of success* might be the headline of a story of a successful perfume and cosmetics business.

The following table lists some common words that are homophones in both British and American pronunciation. They are grouped according to the sound (usually a vowel) that is differently spelled:

i:	be - bee bean - been heal - heel - he'll meat - meet peace - piece scene - seen steal - steel weak - week	e	berry - bury bread - bred lead - led weather - whether	əʊ soʊ	know - no loan - lone pole - poll road - rode role - roll sew - so
ɔ: ɔɪ	board - bored coarse - course morning - mourning warn - worn	ei	brake - break stake - steak male - mail sale - sail tale - tail rain - rein - reign wait - weight way - weigh	ɛər ɪər	air - heir bare - bear fair - fare pair - pare - pear stare - stair ware - wear
u:	blew - blue flew - flu - flue threw - through to - too - two root - route	aɪ	aisle - I'll - isle buy - by - bye eye - I hi - high hire - higher write - right	n	knew - new knight - night knot - not
ɜ: ɜɪ	berth - birth curb - kerb fir - fur heard - herd	aʊ	aloud - allowed foul - fowl	s	ascent - assent base - bass cell - sell cent - scent - sent cereal - serial
ʌ	some - sum son - sun			st	guessed - guest leased - least passed - past

• She picked a large bunch of honeysuckle. [U] • It's an unusual honeysuckle with long orange flowers. [C]
honk [SOUND] /ɛˈhɒŋk, \$hɑ:ŋk/ v, n (to make) a short loud unpleasant noise, like that of a GOOSE or a car horn • The noise of traffic and cars honking (their horns) drifted up to
hon-ori-fic /ɛˈɒn-ərɪˈfɪk, \$,ɑ: nərɪˈfɪ-/ adj [before n] fml showing or giving honour or respect • an honorific post-title
hon-our [RESPECT] Br and Aus Am and Aus **honor** /ˈɒn-

Figure 34: Data box from the CIDE

5.5.3 Phased-in inner texts in the secondary constituent of a parallel macro-structure

Dictionaries can have more than one word list and therefore more than one macro-structure. More than one macrostructure could prevail within the same alphabetical constituent of a dictionary. One such type identified by Wiegand (1989: 402) and Wiegand and Gouws (2013: 88) is the poly-alphabetical macrostructures

The text boxes presented as phased-in inner texts are titled *Words in action*. They are addressed at some of the source and target language items in the articles of the primary macrostructure and they contain different types of data, e.g. pragmatic, usage, linguistic and lexical guidance.

The data boxes help to introduce new comments to enhance the quality of the original dictionary.

Conclusion

Data boxes occur frequently and in diverse ways in dictionaries. This is especially the case in printed dictionaries, although some electronic dictionaries, especially those that are based on printed dictionaries, also employ this type of lexicographic entry. Although various aspects of the use of data boxes have been discussed in metalexicographic literature, much still needs to be done in this regard. This paper focused on what data boxes are, identified them as a type of lexicographic entry and indicated where they are positioned in dictionaries and that they should be used in such a way that they add value to dictionaries. This can serve as an aid for future lexicographers who wish to employ data boxes in their dictionaries.

Endnote

1. The value of a picture presenting different types of deer (but not all types) is not discussed here. The decision to give a pictorial illustration rests with the lexicographer. Therefore it cannot be expected that a specific word will be illustrated or illustrated in the same way in all dictionaries. For example, the online version of the MED does not contain such an illustration.

Acknowledgement

This work is based on the research supported in part by the National Research Foundation of South Africa (Grant specific unique reference numbers (UID) 85434 and (UID) 8576). The Grantholders acknowledge that opinions, findings and conclusions or recommendations expressed in any publication generated by the NRF supported research are that of the authors, and that the NRF accepts no liability whatsoever in this regard.

References

Dictionaries

- BW = Gouws, R., I. Feinauer and F. Poneelis. 1994. *Basiswoordeboek van Afrikaans*. Pretoria: J.L. van Schaik.
- CIDE = Procter, P. (Ed.). 1995. *Cambridge International Dictionary of English*. Cambridge: Cambridge University Press.

- GW = Eksteen, L.C. et al. (Eds.). 1997¹⁴. *Groot Woordeboek/Major Dictionary*. Cape Town: Pharos.
- HAT = Odendal, F.F. and R.H. Gouws. 2005⁵. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Cape Town: Pearson Education South Africa.
- JBS = Stoman, A. et al. (Eds.). 2018. *Junior Tweetalige Skoolwoordeboek/Junior Bilingual School Dictionary*. Cape Town: Pharos.
- LED = Summers, D. (Ed.). 2006. *Longman Exams Dictionary*. Harlow: Pearson Education.
- LSASD = Bullon, S. et al. (Eds.). 2007. *Longman South African School Dictionary*. Harlow: Pearson Education.
- MED = Rundell, M. (Ed.). 2007². *Macmillan English Dictionary for Advanced Learners*. Oxford: Macmillan.
- MPVP = Rundell, M. (Ed.). 2005. *Macmillan Phrasal Verbs Plus*. Oxford: Macmillan Education.
- MWALD = Perrault, S.J. (Ed.). 2008. *Merriam-Webster's Advanced Learner's English Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- OALD = Turnbull, J. (Ed.). 2010⁸. *Oxford Advanced Learner's Dictionary of Current English*. Oxford: Oxford University Press.
- OSD = Pfeiffer, F. et al. (Eds.). 2007. *Oxford Afrikaans–Engels/English–Afrikaans Skoolwoordeboek/School Dictionary*. Cape Town: Oxford University Press Southern Africa.
- OSW = Louw, P. (Senior Editor). 2012. *Oxford Afrikaanse Skoolwoordeboek*. Cape Town: Oxford University Press Southern Africa.
- RWD = Grobbelaar, P. 1987. *Reader's Digest Afrikaans–Engelse Woordeboek/English–Afrikaans Dictionary*. Cape Town. The Reader's Digest Association, South Africa (Pty) Ltd.
- TAW = Du Plessis, M. 1993. *Tweetalige aanleerderswoordeboek/Bilingual Learner's Dictionary*. Cape Town: Tafelberg.
- WLWF = Wiegand, H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer and W. Wolski (Eds.). 2017. *Wörterbuch zur Lexikographie und Wörterbuchforschung/Dictionary of Lexicography and Dictionary Research*. Vol. 2. Berlin/New York: Walter de Gruyter.

Other references

- Bergenholtz, H., S. Tarp and H.E. Wiegand. 1999. Datendistributionsstrukturen, Makro- und Mikrostrukturen in neueren Fachwörterbüchern. Hoffmann, L. et al. (Eds.). 1999. *Fachsprachen. Ein internationales Handbuch zur Fachsprachenforschung und Terminologiewissenschaft/Languages for Special Purposes. An International Handbook of Special-Language and Terminology Research, Bd./Vol. 2: 1762-1832*. Berlin: De Gruyter.
- Gouws, R.H. 2014. Expanding the Notion of Addressing Relations. *Lexicography: Journal of ASIALEX* 1(2): 159-184.
- Gouws, R.H. 2018. Expanding the Data Distribution Structure. *Lexicographica* 34: 225-237.
- Gouws, R.H. and D.J. Prinsloo. 2005. *Principles and Practice of South African Lexicography*. Stellenbosch: SUN PReSS, AFRICAN SUN MeDIA.
- Gouws, R.H. and D.J. Prinsloo. 2010. Thinking out of the Box — Perspectives on the Use of Lexicographic Text Boxes. Dykstra, A. and T. Schoonheim (Eds.). 2010. *Proceedings of the XIV Euralex International Congress, Leeuwarden, 6–10 July 2010: 501-511*. Leeuwarden: Fryske Akademy.

- Gouws, R.H. and S. Tarp.** 2017. Information Overload and Data Overload in Lexicography. *International Journal of Lexicography* 30(4): 389-415.
- Gouws, R.H. et al. (Eds.).** 2013. Dictionaries. *An International Encyclopedia of Lexicography. Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography.* Berlin/New York: De Gruyter.
- Hausmann, F.J. and H.E. Wiegand.** 1989. Component Parts and Structures of Monolingual Dictionaries: A Survey. Hausmann, F.J. et al. 1989-1991: 328-360.
- Hausmann, F.J. et al. (Eds.).** 1989-1991. *Wörterbücher. Dictionaries. Dictionnaires. An International Encyclopedia of Lexicography.* Berlin: Walter de Gruyter.
- Louw, P.A. and R.H. Gouws.** 1996. Lemmatiese en nielemmatiese adressering in Afrikaanse vertalende woordeboeke. *Suid-Afrikaanse Tydskrif vir Taalkunde* 14(3): 92-100.
- Taljad, E., D.J. Prinsloo and R.H. Gouws.** 2014. Text Boxes as Lexicographic Device in LSP Dictionaries. Abel, A., C. Vettori and N. Ralli (Eds.). 2014. *Proceedings of the 16th EURALEX International Congress: The User in Focus, Bolzano/Bozen, Italy, 15-19 July 2014:* 697-705. Bolzano/Bozen: EURAC.
- Tarp, S.** 2004. Basic Problems of Learner's Lexicography. *Lexikos* 14: 222-252.
- Wiegand, H.E.** 1989. Aspekte der Makrostruktur im allgemeinen einsprachigen Wörterbuch: alphabetische Anordnungsformen und ihre Probleme. Hausmann, F.J. et al. 1989-1991: 371-409.
- Wiegand, H.E.** 1996. Das Konzept der semiintegrierten Mikrostrukturen. Ein Beitrag zur Theorie zweisprachiger Printwörterbücher. Wiegand, H.E. (Ed.). 1996. *Wörterbücher in der Diskussion II:* 1-82. Tübingen: Max Niemeyer.
- Wiegand, H.E.** 2006. Adressierung in Printwörterbüchern. Präzisierungen und weiterführende Überlegungen. *Lexicographica* 22: 187-261.
- Wiegand, H.E.** 2011. Adressierung in der ein- und zweisprachigen Lexikographie. Eine zusammenfassende Darstellung. Kürschner, W. and M. Ringmacher (Eds.). 2011. *Aus Ost und West:* 109-234. Frankfurt: Peter Lang.
- Wiegand, H.E., S. Beer and R.H. Gouws.** 2013. Textual Structures in Printed Dictionaries. An Overview. Gouws, R.H. et al. (Eds.). 2013: 31-73.
- Wiegand, H.E. and R.H. Gouws.** 2011. Theoriebedingte Wörterbuchformprobleme und wörterbuchformbedingte Benutzerprobleme I: Ein Beitrag zur Wörterbuchkritik und zur Erweiterung der Theorie der Wörterbuchform. *Lexikos* 21: 232-297.
- Wiegand, H.E. and R.H. Gouws.** 2013. Addressing and Addressing Structures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 273-314.
- Wiegand, H.E. and M. Smit.** 2013. Microstructures in Printed Dictionaries. Gouws, R.H. et al. (Eds.). 2013: 149-214.