

# The Presentation of Word Formation in General Monolingual Dictionaries

Henning Bergenholtz, *Centre for Lexicography, Aarhus University,  
Denmark (hb@asb.dk)*

and

Rufus H. Gouws, *Department of Afrikaans and Dutch,  
Stellenbosch University (rhg@sun.ac.za)*

---

**Abstract:** Word formation in the dictionary belongs, at the latest since the contribution of Mugdan (1984), to the topics frequently discussed by lexicographers. Unfortunately the results have not always been satisfactory. This applies both to the formulation of lexicographic theory and to the lexicographic practice because the lexicographic terms have the linguistic terms as point of departure and questions are put as to how these phenomena should be presented in dictionaries. Instead, one should rather ask which information needs dictionary users experience in which types of user situations and then decide where and how items giving word formation can be presented in order to benefit the envisaged target user of a given dictionary.

The lexicographic practice is also in an unsatisfactory position with regard to items giving word formation because the typical polyfunctional dictionaries have too many items giving word formation for text reception problems and far too few for text production problems or for the real general knowledge needs regarding word formation in the specific language. This paper gives suggestions regarding the theoretical approaches that could lead to a better user-directed lexicographic practice.

**Keywords:** AFRIKAANS DICTIONARIES, COGNITIVE FUNCTION, COMPLEX FORM, COMPOUND, DERIVATIVE, DICTIONARY FUNCTION, ELECTRONIC DICTIONARIES, TEXT PRODUCTION, TEXT RECEPTION, USER NEEDS, WORD FORMATION

**Opsomming: Die aanbieding van woordvorming in algemene verklarende woordeboeke.** Woordvorming in woordeboeke is ten minste sedert die bydrae van Mugdan (1984) 'n onderwerp wat gereeld deur leksikograwe bespreek word. Ongelukkig was die resultate hiervan nie altyd bevredigend nie. Dit geld sowel die formulering van leksikografiese teorie as die leksikografiese praktyk want die leksikografiese terme het linguïstiese terme as vertrekpunt en vrae word gestel oor hoe hierdie verskynsels in woordeboeke aangebied moet word. In stede daarvan behoort daar eerder gevra te word watter inligtingsbehoefte gebruikers in watter tipes gebruikersituasies ervaar en dan moet daar besluit word waar en hoe aanduiders van woordvorming aangebied moet word om die beoogde teikenbruiker van die woordeboek ten beste te bevoordeel.

Die leksikografiese praktyk is ook in 'n onbevredigende situasie met betrekking tot aanduiders van woordvorming. Die tipiese polifunksionele woordeboeke het te veel aanduiders wat woordvorming vir teksbegrip bied en veels te min wat op die oplossing van teksproduksiepro-

bleme gerig is of op die werklike algemene kennisbehoefes oor woordvorming in 'n bepaalde taal. Hierdie artikel bied voorstelle oor teoretiese benaderings wat sou kon lei tot 'n beter gebruikersgerigte leksikografiepraktyk.

**Slutelwoorde:** AFLEIDING, AFRIKAANSE WOORDEBOEKE, ELEKTRONIESE WOORDEBOEKE, GEBRUIKERSBEHOEFES, KOGNITIEWE FUNKSIE, KOMPLEKS, KOMPOSITUM, TEKSBEGRIP, TEKSPRODUKSIE, WOORDEBOEFUNKSIE, WOORDVORMING

### 1. The function of data regarding word formation in monolingual dictionaries?

The question regarding the function of word formation in monolingual dictionaries constitutes the core of existing lexicographic contributions dealing with this theme. However, in the majority of cases the question is formulated the other way round. The point of departure is linguistic terms like derivative, compound, prefix, affixoid, etc. and the question relates to the way in which these phenomena are presented in existing dictionaries and often also how this could have been done in a better way. This is e.g. the approach followed by most contributors in the recent seminal publication *Wortbildung im elektronischen Wörterbuch*, edited by Annette Klosa (2013a), e.g. contributions by Eichinger (2013), Elsen (2013), Splett (2013) and Klosa (2013c). This does not imply that the user has been completely forgotten. Although mention has in some instances been made of the functions, it was done from the perspective of linguistics that focuses primarily on general knowledge regarding word formation. The following citation illustrates this point:

The presentation of word formation in a dictionary is primarily motivated by the fact that it enables a display of relations and interconnections of words. By means of the inclusion of compounds and derivatives, by means of the lemmatization of affixes and by means of the description of word formation rules in the dictionary grammar the interconnectivity of the vocabulary can be successfully indicated although the alphabetical ordering of head words in the dictionary can only display these relations in an inadequate way. In general the usability and effectiveness of dictionaries should be increased by the inclusion of word formation. The reception of items giving word formation can e.g. lead to an enrichment of the vocabulary especially for learners. (Klosa 2013b)

This is true but also confusing. Someone with e.g. a reception problem looks for the meaning of the word — also if it is a derivative or compound. When the user finds the meaning in a printed dictionary in its alphabetical position or in an electronic dictionary directly without an alphabetical access, his/her problem is solved and he/she can continue reading the text. It could obviously happen that the user forgets the initial problem or moves it to the background in order to explore some more general aspects regarding word formation. This could be due to a general interest in the language or to improve his/her con-

versation in the language or his/her reading and writing skills. In polyfunctional dictionaries all of this is available in one and the same dictionary. But one can also make, especially in the case of electronic dictionaries, monofunctional information tools, i.e. dictionaries, available to the users. This tendency can be seen in the contributions by Bergenholtz (2013) and Ten Hacken (2013).

As earlier indicated, the majority of contributions to this theme do not pose the questions that should, according to our point of view, actually be put by the lexicographer: What kind of problem does the user have? How can a dictionary satisfy his/her information needs? Most contributions, not only in Klosa (2013a) but e.g. almost all in Barz, Schröder and Fix (2000) regard word formation in dictionaries from the perspective of a linguist. As has been argued by Bergenholtz and Tarp (2005) this is motivated by the fact that many lexicographers regard lexicography as a subdiscipline of linguistics and therefore want to bring as much from the field of linguistics into the dictionary. It can be useful but this is not necessarily so. The question should be which theory, linguistic or otherwise, is appropriate to ensure that the selection, method and lexicographic presentation for a given dictionary with a specific genuine purpose can be achieved in an optimal way.

The typical user questionnaires are conceptualised in a similar way. The question is often put, typically to linguists or students of linguistics, which items are rated as the most important or which ones are used the most frequently. Even with 2 200 test persons, as in Töpel (2013), no representative result follows when the test group has not been selected according to a principle of representativeness but have rather voluntarily responded to become test persons following an internet invitation. Töpel (2013) asks which of, among others, word formation, pronunciation, the lemma, meaning, grammar, typical uses or synonyms, etc. are regarded as the most important. Surprisingly enough the result is not that the lemma is important for everyone. In the results obtained from that study the lemma is much less important than meaning or grammar, with pronunciation being the least important. But in these results word formation has also been ranked on the lower end of the popularity scale. What does that say? According to us, nothing because the question should actually have been: Which items are extremely important when you use a dictionary as an aid to solve a reception problem? Or: Which items are extremely important when you learn a language?

## **2. The presentation of word formation products in current dictionaries**

The following paragraphs will focus on the way in which some current dictionaries present word formation products. This discussion does not intend to portray a representative view of the lexicographic treatment of word formation products. However, albeit that examples are only taken from a few randomly selected dictionaries, the presentation found in these dictionaries should not be seen as isolated procedures restricted to these dictionaries or dictionaries of the

respective languages. A similar or comparable presentation can be found in many other dictionaries — of the same and of other languages. Lexicographers dealing with languages with derivatives and compounds as word formation products could do well to embark on innovative presentation procedures to enhance the access of their intended target users to these items, to respond to the needs of these users and to improve the way in which the given dictionary satisfies the identified lexicographic functions.

## 2.1 Printed dictionaries

Monolingual Afrikaans dictionaries like the *Verklarende Handwoordeboek van die Afrikaanse Taal* (HAT) and *Verklarende Afrikaanse Woordeboek* (VAW) follow comparable ways with regard to the presentation of derivatives and compounds. In both these dictionaries a list of derivatives and compounds is often attached to the article of a lemma representing the first stem of the relevant derivative or compound. The users' guidelines text of VAW gives no explanation of this kind of presentation. HAT does motivate this presentation by indicating that these words are *unexplained headwords*. They are not lemmatised and explained separately because the meaning of the word formation product is self-evident if the different stems are included and explained in their respective alphabetical positions, cf. Gouws (1989).

Attached to the article of the lemma *herstel*<sup>2</sup> the VAW has the following sequence of words:

**~afdeling, ~baar, ~depot, ~krag, ~kuil, ~ler, ~ling.**

In the users' guidelines text it is mentioned that the tilde is a place-keeping symbol for the form represented by the lemma. The use of the tilde increases the degree of textual condensation and makes it even more difficult for the user, especially the occasional dictionary user, to have a rapid access to the required form. In this list the words *herstelbaar*, *hersteller* and *herstelling* are derivatives whereas the rest are compounds. The list is presented in an alphabetical ordering without distinguishing between derivatives and compounds. From a linguistic point of view a distinction between these two types of word formation products might have been helpful. For the user looking for the specific word an alphabetical ordering ensures easier access. However, a significant problem lies in the fact that the word *herstel* has no less than eight polysemous senses allocated to it in the specific dictionary article. There is no way the user will know which sense applies in the different complex forms, i.e. derivatives and compounds. Albeit that the primary sense is usually the one used in word formation products, cf. Gouws (1988), this is not always the case. The complex forms included in this list are not also included as lemmata although a number of compounds with *herstel-* as first stem have been included. The user is at a loss when having to decide where to find a given compound or deriva-

tive: attached to the article of the lemma representing the first stem or included in its own alphabetical position as guiding element of a separate article. From a text reception perspective this way of presentation leaves much to be desired. Where the word formation demands the inclusion of a linking morpheme VAW presents the linking morpheme as part of the second component of the complex form, as can be seen in the article of the lemma *kind* that has an *-er-* as linking form:

**~eraand, ~erbal, ~erbottel, ~erdrag, ~ergebattel, ...**

to represent *kinderaand*, *kinderbal*, *kinderbottel*, etc. From a linguistic perspective this is unacceptable and from a user perspective the subsequent degree of textual condensation is extremely confusing. This form of textual condensation with the first component of the compound in a remote position makes it difficult for the average user to quickly interpret the compound correctly. If the linking morpheme does not occur in all complex forms the presentation confuses the user even more, cf. the list attached to the article of the lemma *meisie*:

**~(s)koshuis, ~agtig, ~span, ~stehuis**

as condensed forms of *meisiekoshuis* / *meisieskoshuis*, *meisieagtig*, *meisiespan*, *meisiestehuis*.

Although HAT also makes no distinction between derivatives and compounds and also fails to link the unexplained derivatives and compounds to specific senses of the first stem the listing of unexplained forms is immediately preceded by a repetition of the first stem, e.g. attached to the article of the lemma *doop*:

**doop: ~bak, ~bediening, ~boek, ~diens, ...**

Where a linking morpheme is needed it is attached to this first stem that stands in the entrance position of the list of unexplained forms as in the list attached to the article of the lemma *meisie*. From both a linguistic and a user perspective this is a more acceptable and satisfying approach:

**meisie: ~agtig, ~jare, ~kind, ... ~stem. meisies: ~boek, ~drag, ~gesig, ...**

From a text reception perspective the same problems prevail as in VAW but the lesser degree of textual condensation impedes access to the unexplained forms in a lesser way.

In both these dictionaries the decision regarding explained or unexplained is done in an arbitrary way. Gove (1966) remarked that the *self-* in *self-explanatory* should refer to the interpreter of the word and not to the word itself. When planning and compiling a dictionary the *self-* in *self-explanatory* should refer to the user and not to the lexicographer. The application of such an approach will lead to the lemmatisation of many more complex forms.

*Nasionale Woordeboek* (NW) also includes complex forms in a list attached to the article of a lemma representing the first stem of the complex word. However, the ordering within the list is partially determined by morpho-semantic criteria, cf. the list attached to the article of the lemma *skeer*:

~**skeerder** (by 2); **skeerapparaat**, **-goed**, **-kwas**, **-mes**, ... (by 1), **-geld**, **-hok**, **-kraal**, **-skêr**, ... (by 2).

The markers (*by 1*) and (*by 2*) are semantic markers indicating which sense of the polysemous lemma applies in the specific complex form. The semicolon preceding *skeerapparaat* marks a division between derivatives and compounds. This is valuable linguistic information but it increases the degree of textual condensation. The inclusion of these complex forms is motivated in the users' guidelines text where it is stated that they are included to indicate their existence or to show their spelling. Yet again the distinction between explained and unexplained complex forms is done in a haphazard way because this dictionary also includes compounds and derivatives as guiding elements of fully-fledged articles.

The *Woordeboek van die Afrikaanse Taal* (WAT) also makes a distinction between explained and unexplained complex forms with the former being included as lemmata and the latter as entries in a list attached to the article of the lemma representing the first stem. In this list, a rudimentary treatment is allocated to the complex forms by means of an item giving the part of speech and a marker of the main stress of the word, cf. the list following the article of the lemma *ritme*:

**rit'megevoel** s.nw., **rit'meloos** b.nw., **rit'mepatroon** s.nw., **rit'mevariasie** s.nw., ...

In this list the word *ritmeloos* is a derivative and the rest are compounds. This is not indicated. The use of a much lesser degree of textual condensation by giving the full forms of the complex words enhances the chances for easy access to these forms.

In VAW, HAT, NW and the WAT the presentation of word formation products assist in text production with the user being able to comprehend something of the system followed in the formation of derivatives and compounds. But the lexicographers eschew the text reception function. Not only will users have problems in determining the relevant sense of the first stem of the unexplained compounds but even worse, no attention is given to the meaning of the second stem in these compounds where different senses could also come to the fore.

*Van Dale Groot Woordenboek der Nederlandse Taal* includes some complex forms as main lemmata as part of a straight alphabetical macrostructure whereas other complex forms are included as sublemmata being the guiding elements of horizontally ordered niched articles. The word *moeras* is included as

main lemma. Attached to its article a horizontally ordered niched article cluster follows with sublemmata like *moerasachtig*, *moerasaloë*, *moerasandijvie*, *moeras-andoorn*, ... *moerashoenders*. Each niched article contains at least a paraphrase of meaning but in addition also entries like items giving grammatical data or example sentences. The article of the niched lemma *moerashoenders* is followed by the main lemmata *moerashoorn* and *moerasijzer* and then again a niche of horizontally ordered articles with compounds as lemmata. This represents a typical example of multiple niching, cf. Gouws (2005). Although a strict alphabetical ordering is maintained the distinction between vertically and horizontally ordered lemmata remains unclear to the user.

*The New Oxford Dictionary of English* (NODE) includes compounds as main lemmata whilst derivatives are included within the article of a lemma representing a form from which the derivative was derived. Derivatives are given in a specific article zone and they are preceded by the marker "- DERIVATIVES". The article of the lemma *lemma* has no derivatives but the subsequent article, i.e. that of the lemma *lemmatize* has the noun *lemmatization* as derivative. Derivatives receive a rudimentary treatment, e.g. items giving pronunciation and part of speech. For both text reception and text production purposes the presentation in this dictionary is helpful.

Even when a user is familiar with the system used by a given dictionary with regard to the presentation of derivatives, it is not always that easy to achieve a rapid access to the required form. In its explanatory notes, *Webster's Ninth New Collegiate Dictionary* (W9) explains the inclusion of derivatives as run-on entries, following a "main entry". They are not defined "because their meanings are readily derivable from the meaning of the root word". One of the problems the user experiences with this system is that some derivatives are entered as run-on entries and others as main lemmata. The user does not know where to find a required form. The article of the lemma *incline* contains the run-on entry *incliner*. The user looking for the derivatives *inclinable*, *inclination*, and *inclinational* will not find them in this article. The derivatives *inclinable* and *inclination* are guiding elements of their own articles whereas *inclinational* is presented as a run-on entry in the article of *inclination*. Whilst *inclinable* is presented as lemma and not as run-on entry in the article of *incline*, *includable* is a run-on entry in the article of the lemma *include*, with *inclusion* included as separate lemma.

In W9, main lemmata are typically given an item indicating the date of the earliest recorded use. This is unfortunately not done for run-on entries. It can therefore not be deduced whether the run-on entry has really been derived from the respective main lemma. The dating of derivations given as main lemmata shows the difficulty of guessing which form constitutes a derivation from another form. The following main lemmata have been included (with their dates of earliest recorded use in brackets): *premeditate* (1548), *premeditated* (1590), *premeditation* (15c), *premeditative* (1858). According to these dates the noun *premeditation* is the form from which the verb *premeditate* and the adjective *premedi-*

*tated* were derived. *Premeditator* is included as run-on entry in the article of *premeditate* but without an item giving its date of earliest recording. Yet again the user has problems in knowing where to find a specific derivative.

The *De Gruyter Wörterbuch Deutsch als Fremdsprache* (WDF) has an article zone identified as accommodating the word family of the word represented by the lemma. A structural indicator precedes this slot and marks it clearly for the user. A system of cross-referencing guides users from articles with different members of the word family as lemmata to the article where the complete family is given. The articles of the lemmata *Diplom*, *Diplomat* and *diplomatisch* have cross-reference items guiding the user to the article of the lemma *Diplomatie*. Here the list of word family members, i.e. *Diplom*, *Diplomatie*, *diplomatisch* is given. It is interesting that the word family list is not given in the article of the base form but rather in the article of a derivative. The system of cross-references allows the user access to all the derivatives. These lists do not always only contain items with the same first stem. The article of the lemma *Kritik* has the following word family list: *Kritiker*, *kritisch*, *kritisieren*, *kritteln*, *gesellschaftskritisch*, *Selbstkritik*, *selbstkritisch* and the article of the lemma *Platz* has the following items in its word family list: *platzieren*, *Arbeitsplatz*, *Campingplatz*, *Flugplatz*, *Parkplatz*, *Schauplatz*, *Sitzplatz*, *Stehplatz*, *Studienplatz*, *Platzkarte*. The inclusion of the different members of a word family as lemmata and the cross-referencing to the article where the full family is found, helps a user to find a given complex form.

From the preceding discussion it is clear that dictionaries have diverse ways of presenting word formation products. One often has the feeling that the needs of the intended target users and the lexicographic functions of the dictionary have not been taken into account sufficiently in the planning of the way in which these word formation products are to be presented. One of the problems for a user consulting an existing printed dictionary in order to find a given derivative or compound is to decide in which article the specific form will be found. Access to word formation products is often impeded by uncertainty as to where the form is entered in the dictionary and what the system of ordering of word formation products is. The typical user of a general dictionary should not have to negotiate the distinction between a base and the derivative forms. He/she merely wants to find the form he/she has encountered in a text and wants to find it as quickly as possible. The average dictionary user is familiar with the alphabetical ordering within dictionaries. Consulting the dictionary to find a specific word formation product should not demand knowledge of an alternative ordering system, e.g. one based on morpho-semantic criteria. Where space restrictions demand that compounds and derivatives should be included in a cumulative list attached to the article of the lemma representing the word that corresponds to the first stem of the compound or derivative the lexicographer should maintain the alphabetical ordering within that list. As many complex forms can be given as regarded necessary by the lexicographer to ensure the best possible presentation in a given dictionary.

Another problem regards the treatment allocated to word formation products. If the inclusion is restricted to their occurrence in a cumulative list the user with a text reception problem will have little assistance. Where text reception is the function a paraphrase of meaning is needed for a complex form. If the dictionary has a cognitive function, data like the date of the first recording of a derivative or compound could be of interest. Additional items could give guidance regarding the history of the complex form, the structure, i.e. the word formation process resulting in the specific complex form, the meaning of the complex form and its frequency of use.

## 2.2 Electronic dictionaries

In electronic dictionaries, a search for a specific complex form can usually have quicker success because of direct access to the required form, irrespective of it being included as lemma or as derivative/compound in the article of another lemma.

In *ellexico*, an electronic dictionary of German compiled by the Institut für deutsche Sprache, a variety of word formation forms are given. For the word *ernst*, a list is given in which combinations determined by the part of speech of the constituents are ordered alphabetically and for each one its usage frequency in the specific corpus is indicated, e.g. *bierernst* (836), *tiefernst* (1 010), *todernst* (760). Clicking on anyone of these forms takes the user to the specific lemma but also to an alphabetical list of complex forms indicating the alphabetical environment of the specific word.

This dictionary gives access to a wide-ranging selection of complex forms. Clicking on the word *Angst* leads to an article with links to both compounds and derivatives. In the section on compounds, there are alphabetical lists with *Angst-* as first and *-angst* as second stem, e.g. *Angstanfall*, *Angskauf*, *Angsluft* and *Altersangst*, *Bürgerangst*, *Herzangst*. A question can be asked regarding the choice of the form to be indicated as the base. In the article of the word *Liebe*, it is indicated that the verb *lieben* is the base form with the noun *Liebe* (with 211 723 occurrences in the corpus) as a derived form. This might be valuable linguistic guidance. The typical non-linguistic user is not interested in the distinction between base and derivative. In this dictionary, the word *Liebe* can be accessed directly or via the article of the base *lieben*. This is good lexicography. However, the route via *lieben* is not necessarily a quick road. When searching for the verb *lieben* the automatic access is to the noun *Lieben*. On the left of the screen there is a word list with the article stretch to which *Lieben* belongs alphabetically. This list also includes the verb *lieben* and clicking on this word takes the user to the lemma. The choice of *Lieben* as only destination when entering *lieben* as a search word seems to be an arbitrary choice that is not supported by the frequency of use of the respective forms.

The problem indicated in the Afrikaans dictionaries HAT, VAW, NW and WAT that the user does not know which sense of a polysemous word is acti-

vated in the occurrence of the corresponding stem when included as first or second constituent of a compound is cleverly avoided in *The Danish Writing Dictionary*. In this dictionary, the word *portræt* is treated as a polysemous word. The relevant paraphrase of meaning is followed by, among others, a list of complex forms in which the stem corresponding to the word represented by the lemma has the polysemous sense indicated in the specific paraphrase of meaning, e.g.

**portræt** (= portrait)

1. billede, som gengiver en eller flere personer i form af et maleri, tegning eller fotografi; også om et billede, der kun gengiver ansigtet (= picture reflecting one or more persons by means of a painting, drawing or photo; also a picture giving only the face)

**Orddannelser** (= word formation)

Nøgenportræt nude (= nude portrait)

portrætalbum (= portrait album)

portrætbillede (= portrait picture)

portrætfoto (= portrait photo)

portrætfotografering (= portrait photographing)

portrætfotografi (= portrait photography)

portrætgalleri (= portrait gallery)

portrætlighed (= portraitness)

portrætmaler (= portrait painter)

portrættere (= to make a portrait)

portrættering (= making a portrait)

portrættør (= maker of a portrait)

selvportræt (= self-portrait)

2. tekst, film, udsendelse eller lignende, som giver et indtryk af en persons eller en gruppe menneskers karakterer (= text, film, television broadcast or something similar that describes the character of a person or group of persons)

**Orddannelser** (= word formation)

portrætsamtale (= portrait discussion)

portrættere (= to make a portrait)

portrættering (= making a portrait)

portrætudsendelse sendung (= portrait broadcast)

Here the verb *portrættere* (to make a portrait) is given as a derivative. However, this word is also included as lemma and guiding element of its own article. This article also includes word formation products, including the form *portræt*. The user is not confronted with the distinction between a base and a derived form but the poly-accessibility of this dictionary makes it possible to have rapid access to the desired form but also to link a derived form to its base. Yet again each one of the paraphrases of meaning of the different polysemous senses of this word is allocated an own article zone. This zone also contains the relevant

word formation products and possible base forms:

**portrættere**

1. gengive en eller flere personer i form af et maleri, tegning eller fotograf (= one or more persons pictured as painting, drawing of photograph)

**Orddannelser** (= word formation)

portræt (= portrait)

portrættering (= to make a portrait)

portrættør (= maker of a portrait)

2. gengive og give et indtryk af en persons eller en gruppe menneskers karakterer i form af en tekst, film, udsendelse eller lignende; benyttes især i forbindelse med skuespil (= to characterise a person or group of persons, especially in a film or play)

**Orddannelser** (= word formation)

portræt (= portrait)

portrættering (= to make a portrait)

portrættør (= maker of a portrait)

The article zones for both polysemous senses offer the same word formation products. This will help the user with both text reception and text production needs because the user has no uncertainty with regard to the meaning of the first stem of the word formation products.

### 2.3 Where can information regarding word formation be found in dictionaries?

Dictionaries can assist their users with regard to word formation products in various ways. In the present discussion, the focus has primarily been on word formation products presented in dictionary articles. Although this is a frequent and important way of conveying word formation information, there are also other ways to do so.

The most typical way remains the lemmatisation of a derivative or compound. Lemmatisation typically leads to a treatment, albeit perhaps of a limited extent, but the user retrieves information that helps him/her to understand and use the complex word correctly. When planning the lemmatisation and treatment of compounds and derivatives, the lexicographer should also negotiate the possibility of variant forms of a given complex. A similar approach that is followed with regard to other variants should also be applied to compounds and derivatives. Both or all the variants could be lemmatised with the necessary cross-references given where needed. In VAW the word *meisie(s)koshuis* indicates that this compound can be used with or without the linking morpheme. No preference is indicated. In a proscriptive dictionary, one specific form could be indicated as the recommended variant. Guidance with regard to word formation can also be given by means of lexicographic comments. In

HAT a comment is added to the article of the word *lief* stating that *lief* is sometimes used as a suffix which is added to nouns, e.g. in *vaderlief* and *kindjief*. Such comments are helpful but they do not allow a systematic retrieval of information. As seen in some of the dictionaries discussed in the preceding sections, word formation items can also be presented in a word formation field or article zone. Separate fields for derivatives and compounds can be included in or attached to an article. Although a word formation product is not lemmatised, its inclusion in such a field or cumulative list can help the user to recognise the word and in some cases even to understand it although text reception is at best accomplished by means of a paraphrase of meaning and not by a mere listing of a so-called self-explanatory complex word. The treatment of word formation is also strengthened by the inclusion as lemmata of relevant terms from the field of word formation where these terms are explained in such a way that the target user of the dictionary can use them to increase his/her understanding of word formation types presented as part of the subject matter of the dictionary.

Lexicographers can also utilise outer texts to present data regarding word formation. One such possibility which is extremely helpful but unfortunately too often neglected is the inclusion of a separate section in the dictionary grammar presented as front or back matter text. This is for example done in the *Malagasy–German Dictionary*. Such an approach integrates grammar and dictionary and brings an added value to the dictionary as language learning instrument. Guidance regarding word formation can also be given within the users' guidelines text, typically presented within the front matter section. Such a discussion makes the user aware of the way in which word formation products are presented and treated in the given dictionary. Dictionaries can also give cross-references or links to relevant internet websites or other dictionary-external sources, e.g. text books or grammars where supplementary discussions of various word formation processes can be found. Such an approach emphasises the relations holding between different reference sources and confirms the fact that dictionaries do not have to be used in isolation but form part of a bigger reference network.

### **3. Word formation items as a tool to assist dictionary users in different types of situations**

#### **3.1 Word formation items as a tool to assist dictionary text reception problems**

If you do not understand a derivative or a compound you will not need to know whether it is a derivative or a compound. You do not need to know anything but the meaning. When the required word can be found as lemma in the dictionary the user has performed a successful dictionary consultation. When the word has not been lemmatised, Bergenholtz and Tarp (2005: 581), as

many other lexicographers in various contributions, believe that the dictionary does not offer any help. This is not completely correct; at least not in electronic dictionaries. When one looks for a word that has only been lemmatised in its occurrence with a linking morpheme, e.g. the German *Glaubenssache*, but not *Glaubensache* (without the linking -s-), the electronic dictionary can suggest, as we also know from Google searches: Do you mean *Glaubenssache*? In this case even a printed dictionary can help if the searched word is not given too far from the lemmatised word. But in some other cases only the electronic dictionary can help, e.g. in the case of a derivative where the derivative is only given as item presenting the word formation in the dictionary and not also as lemma. In *The Danish Meaning Dictionary* (2013), the form *ionosfærisk* (= ionospheric) has not been lemmatised. Because it is included as item giving the word formation in the article of the lemma *ionosfære* (= ionosphere) the user retrieves the following dictionary article with the following introduction:

No article *ionosfærisk* can be found. Instead, the following article is shown in which *ionosfærisk* occurs as item giving word formation:

**ionosphere** noun

the outer part of the atmosphere of about 80 km and more, where the air has been ionised, i.e. the air molecules have been split in ions and electrons.

The knowledgeable user will now realise that the adjective has more or less the same meaning. The user cannot know for sure that he/she can use the explanation found for the noun, but he/she will see that it fits the context of the not understandable word *ionisfærisk*.

However, it does not always work. And the problem is not only whether the lexicographer has presented an item giving word formation or not. It is not relevant to the user with a reception problem whether the following advice had been followed:

Thus, Dutch *handappel* 'eating apple' should be listed under *appel*, not under *hand*, because language users know that the word *handappel* stands for a subset of apple, not of hands, and will look for this word under the heading for *appel*. (Booij 2003: 254)

From a linguistic point of view the argument might be correct, but you cannot be sure that the intended user is thinking in exactly the same way. And much more important is the fact that you should not make any lexicographic choices without considering the intended dictionary functions. If you have a text reception problem, you do not understand the word *handappel*; it is difficult to look up under both the guiding elements *hand* and *appel*. You need the individual lemma *handappel* in order to find help in solving the reception problem. If you do not understand a compound at all and the meaning of the parts of the compound are not really related to the meaning of the compound as a whole, you can only be sure that your assumption or guess is correct, when you find

the word in the lemma list. That means that a really helpful reception dictionary must be a very large one with at least 200 000 lemmata. This is too comprehensive for general printed dictionaries but for electronic dictionaries it only remains a question whether the project has or had enough lexicographers.

### 3.2 Word formation items as a tool to assist text production problems

All the arguments in favour of including many word formations in order to help solve reception problems are also valid in terms of production problems. But here the use of a cumulative list of unexplained word formations with the lemma sign can be useful for the native speaker who only wants to know that the word formation exists and how it is spelled or whether it is used with a linking morpheme or not. In the latter case you often have variants for which it is not sufficient to only inform about the existence of variants. Instead, the user needs a recommendation for making his choice. At best, this can be done through a simple reference to the recommended word, e.g.:

**job** noun ...

Word formations:

bijob (= sideline job)

deltidsjob (part time work)

fuldtidsjob (= full time job)

fuldtidjob (= full time job) [this spelling without -s- is unusual and not recommended, use instead fuldtidsjob]

(In this article, the underlined forms constitute links that the user can follow to obtain a more comprehensive treatment of the specific word.)

Word formation items are important if the formation is irregular or somehow not predictable. This is often the case with nouns for people living in a country or a town or adjectives for something belonging to that part of the world. The user knows the name of the town or the country, but not the related noun or adjective, e.g.: People coming from or living in *Copenhagen (København)* are called *københavnner* (with an *-er* as derivation morpheme), people from *Silkeborg* are called *silkeborgenser* (with an *-enser* as derivation morpheme), people from *Århus* are called *århusianer* (with an *-ianer*) as derivation morpheme). In Afrikaans an inhabitant of *Durban* and *Paarl* is called a *Durbaniet* and a *Paarliet* respectively and an inhabitant of *Bloemfontein* is a *Bloemfonteiner* and wine from *Stellenbosch* is *Stellenbosse wyn*. This is not really a problem in a text reception situation, because you can guess what the words *københavnner*, *silkeborgenser*, *århusianer*, *Durbaniet*, *Bloemfonteiner*, etc. mean, but in a text production situation you cannot predict how the adjective derived from Stellenbosch or the noun indicating an inhabitant from *København* must be formed if you do not already know it. Therefore it is crucially important that city and country names should also be included as lemmata in order to be able to enter the relevant

word formation products. This is yet again clear when the word formation is not done in a systematic or predictable way, e.g. the inhabitants of the Danish city of *Ribe* are not *ribener* or *ribenser* (with a *b*), but *ripenser* (with a *p*). Also if *ripenser* is a lemma, it does not help the user, who only knows the name of the town. He/she will not find this lemma, but will need a presentation like the following:

**Ribe** PROPER NOUN

**Word formations** ripenser, ripensisk

(*Ribe* is a Danish town in Southern Jutland; a *ripenser* is someone from Ribe, *ripensisk* refers to something from Ribe)

But in principle it also applies to a text production dictionary that it ideally contains all word formation products of a language as lemmata. It is unforgivable to exclude opaque compounds as the earlier mentioned Dutch word *handappel* or the Afrikaans form *broekskeur* (literally to tear pants, meaning "difficult"). It is of little assistance to the user whether the word *handappel* is presented as an item giving word formation at *hand* or at *appel*. If he/she does not know the word and even if he/she finds it as word formation item but does not understand it, he/she cannot use the word. The more lemmata the dictionary contains the better are the chances to help the user in those cases, not so infrequent, where the user is able to form the word but is not sure whether the word exists in the language, i.e. whether it can actually be used in the language. Alternatively, as word formation item, the dubious word can assist the user realising that the word does exist. Admittedly Google nowadays has a comparable function when the user no longer trusts the dictionary.

### 3.3 Word formation items as a tool to assist knowledge needs

To satisfy a cognitive need, the user would often not consult a dictionary but rather a grammar or a linguistic text book. If a dictionary does have an integrated dictionary grammar, one can find in one information tool both a systematic presentation of the grammar and possible cross-references to the description of the single phenomenon in the dictionary, or explicit cross-references from concrete dictionary articles to the more comprehensive presentation in the dictionary grammar. In a dictionary directed at cognitive needs, the complementary items to individual items giving word formation, especially in the form of comments, play a much more significant role than in communication dictionaries. Furthermore the description of the lemmatised word formation terms also plays an important role. Finally, comprehensive comments regarding linguistically interesting problems are also important, e.g. with regard to the relation between morphological variants. These problems can be illustrated by the following two examples. The first example shows a lemmatised derivation morpheme in the *Malagasy–German Dictionary* (1991). Note in particular here the reference to a dictionary grammar, i.e. a dictionary grammar written in

both languages as part of the dictionary:

- a verbal morpheme which is added to a word stem with two different functions
  1. used for forming action (agisive) verbs, e.g. adaboka (give up) → §1.5.4
  2. used for forming instrumental verbs, e.g. aharatra (shave) → §1.5.6

The second example shows a grammatical comment directed at a variant problem from *The Danish Internet Dictionary*. It is a polyfunctional dictionary with "Knowledge regarding the Danish language" as part of one of its functions. It also is a proscriptive dictionary, i.e. neither a purely descriptive nor a prescriptive dictionary:

**massemedium** = Massemedium (= mass media)

Medium that reaches the majority of the population of a country or a region, e.g. a news paper, television or radio.

This word formation product is not recommended, rather use → *massemedie*.

**Comment:** In most cases of two words with the same meaning but with the suffixes *-ium* or *-e*, the variant with the *-ium* is recommended, e.g. the word *gymnasium*. But in words with the stem *medi* and the two variants *medium* and *medie* there is a different proposal. Generally the use of *-medie* is recommended, as also in this case, where a Google search in March 2011 shows 1.980 citations with *massemedium* and 25.900 citations with *massemedie*. In this regard our advice is in line with the language use. But the proposal does not mean that the other variant, i.e. *massemedium*, should be regarded as incorrect. It is merely recommended that the other variant should be used.

It is important that the presentation of word formation products should be done according to a specific and well-defined concept. This is co-determined by whether the envisaged dictionary is compiled for linguists, students of linguistics, linguistically lay people interested in language or e.g. learners of a language. In the above-mentioned Malagasy and Danish examples two articles have been cited from dictionaries compiled for linguistic lay persons. Corresponding articles e.g. for students of linguistics could and should be much more comprehensive. In a similar way articles dealing with linguistic terms regarding word formation should be sufficiently exact and comprehensive.

#### 4. Conclusion

Diversity in lexicography implies different dictionaries for different needs in different situations of use. In specific situations of use specific users consult a dictionary to find a compound or derivative and to retrieve information regarding that word. Current dictionaries display a variety of ways in which

complex forms are presented and treated. In many instances this presentation is determined by linguistic criteria and not necessarily by the needs of the intended target users of the dictionary. The discussion in this paper illustrates the current situation and argues in favour of a presentation and treatment determined by the needs of the users. Compilers of electronic dictionaries should utilise the means at their disposal to give a more comprehensive selection of compounds and derivatives as lemmata and a treatment directed at the functions of the dictionary. In the treatment of word formation products dictionaries should not be seen as isolated products but the use of links and references to other sources should emphasise the role of a dictionary as part of an interactive network of reference sources.

## 5. References

- Barz, I., M. Schröder and U. Fix (Eds.).** 2000. *Praxis- und Integrationsfelder der Wortbildungsforschung*. Heidelberg: Winter.
- Bergenholtz, H.** 2013. Wortbildungsangaben als Hilfe für den Zugriff auf andere Datentypen und als Hilfe bei kommunikativen und kognitiven Informationsbedürfnissen. Klosa, A. (Ed.). 2013a: 133-156.
- Bergenholtz, H. and S. Tarp.** 2005. Dictionaries and Word Formation. Brown, K. (Ed.). 2005. *Encyclopedia of Language and Linguistics. Volume 3*: 580-583. Second edition. Oxford: Elsevier.
- Booij, G.** 2003. The Codification of Phonological, Morphological, and Syntactic Information. Van Sterkenburg, P. (Ed.). 2003. *A Practical Guide to Lexicography*: 251-259. Terminology and Lexicography Research and Practice 6. Amsterdam/Philadelphia: John Benjamins.
- De Gruyter Wörterbuch Deutsch als Fremdsprache** = Kempcke, G. et al. (Eds.). 2000. *De Gruyter Wörterbuch Deutsch als Fremdsprache*. Berlin/New York: Walter de Gruyter.
- Eichinger, L.M.** 2013. Wortbildung im Wörterbuch. Aus der Sicht eines Grammatikers. Klosa, A. (Ed.). 2013a: 63-86.
- ellexico** = *ellexico* Online Wörterbuch zur deutschen Gegenwartssprache. Compiled by the Institut für deutsche Sprache, Mannheim. Available: <http://www.owid.de/wb/ellexico/start.html> (Accessed 6 April 2013).
- Elsen, H.** 2013. Problemzonen der Wortbildung und der Eintrag im Wörterbuch. Klosa, A. (Ed.). 2013a: 87-103.
- Gouws, R.H.** 1988. Compounds in Dictionaries: A Semantic Perspective. Snell-Hornby, M. (Ed.). 1988. *ZüriLEX '86 Proceedings. Papers Read at the EURALEX International Congress, University of Zürich, 9-14 September 1986*: 91-97. Tübingen: Francke Verlag.
- Gouws, R.H.** 1989. *Leksikografie*. Pretoria/Cape Town: Academica.
- Gouws, R.H.** 2005. Multiple Niching. Gottlieb, H., J.E. Mogensen and A. Zettersten (Eds.). 2005. *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen*: 261-275. Lexicographica. Series Maior 115. Tübingen: Max Niemeyer.
- Gove, P.B.** 1966. Self-explanatory Words. *American Speech* 41: 182-198.
- Klosa, A. (Ed.).** 2013a. *Wortbildung im elektronischen Wörterbuch*. Tübingen: Gunter Narr.
- Klosa, A.** 2013b. Einleitung. Klosa, A. (Ed.). 2013a: 7-11.

- Klosa, A.** 2013c. Wortbildung in elexiko: Gegenwart und Zukunft. Klosa, A. (Ed.). 2013a: 175-196.
- Malagasy–German Dictionary** = Bergenholtz, H. in cooperation with Rajaonarivo, S., R. Ramasomanana and B. Radanielina as well as Richter-Johanningmeier, J., E. Olszowski and V. Zeiss in collaboration with Ranaivoson, H., N. Rasoarimanana, R. Ravololomboahangy and M. Razafiarivony. 1991. *Rakibolana Malagasy–Alema/Madagassisch–Deutsches Wörterbuch*. Antananarivo: Leximal/Moers: aragon.
- Mugdan, J.** 1984. Grammatik im Wörterbuch: Wortbildung. Wiegand, H.E. (Ed.). 1984. *Studien zur neuhochdeutschen Lexikographie IV*: 237-308. Hildesheim/Zürich/New York: Olms.
- Nasionale Woordeboek** = De Villiers, M. et al. (Eds.). 1987. *Nasionale Woordeboek*. Cape Town: Nasou.
- Splett, J.** 2013. Grundlegende Bemerkungen zu einem auf einer pragmatischen Sprachtheorie fußenden Wortfamilienwörterbuch als legitimen Ort einer integrierten Wortbildung. Klosa, A. (Ed.). 2013a: 117-129.
- Ten Hacken, P.** 2013. Wortbildung in elektronischen Lernerwörterbüchern. Klosa, A. (Ed.). 2013a: 157-174.
- The Danish Internet Dictionary.** 2013. = Bergenholtz, H. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, K.B. Eriksen, H.R. Gudmann, H.H. Jensen, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com.
- The Danish Meaning Dictionary.** 2013. = Bergenholtz, H. in cooperation with Jensen, H.H. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, K.B. Eriksen, H.R. Gudmann, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Betydningsordbog*. Odense: Ordbogen.com.
- The Danish Writing Dictionary.** 2013. = Bergenholtz, H. in cooperation with Eriksen, K.B. in collaboration with Agersnap, O., A. Bodilsen, S.W. Dohrn, H.R. Gudmann, H.H. Jensen, J.H. Jensen, M. Holst, C. Kjølhedede, J. Nguyen and H.A. Pedersen. Database: Almind, R., R.T. Styrk and P. Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com.
- The New Oxford Dictionary of English** = Pearsall, J. (Ed.). 1998. *The New Oxford Dictionary of English*. Oxford: Oxford University Press.
- Töpel, A.** 2013. Die Wortbildungsangaben im Online-Wörterbuch und wie Nutzer sie beurteilen — eine Umfrage zu elexiko. Klosa, A. 2013a: 197-214.
- Van Dale Groot Woordenboek der Nederlandse Taal** = Geerts, G. et al. (Eds.). 1992. *Van Dale Groot Woordenboek der Nederlandse Taal*. Utrecht: Van Dale Lexicographie.
- Verklarende Afrikaanse Woordeboek** = Labuschagne, F.J. et al. (Eds.). 2010. *Verklarende Afrikaanse Woordeboek*. Cape Town: Pharos.
- Verklarende Handwoordeboek van die Afrikaanse Taal** = Odendal, F.F. and R.H. Gouws (Eds.). 2005. *Verklarende Handwoordeboek van die Afrikaanse Taal*. Fifth edition. Cape Town: Pearson Education South Africa.
- Webster's Ninth New Collegiate Dictionary** = Mish, F.C. (Ed.). 1987. *Webster's Ninth New Collegiate Dictionary*. Springfield, Massachusetts: Merriam-Webster.
- Woordeboek van die Afrikaanse Taal** = Botha, W.F. (Ed.-in-Chief). 2009. *Woordeboek van die Afrikaanse Taal, Volume 13*. Stellenbosch: Bureau of the WAT.