

Heidi Agerbo*

A Re-Examination of the Lemma through Digital and Functional Lenses

Abstract

The lemma is one of the core terms in lexicography. It is the lexicographical item that users of paper dictionaries normally utilise to look for the dictionary article that can help them satisfy their information need, and lexicographers consider the lemma an essential and indispensable part of a lexicographical tool. However, as new technologies are continuously being developed and many dictionaries (or lexicographical data) are incorporated into other kinds of digital tools, e.g. writing software and e-readers, it is necessary to re-examine the notion of lemma because these changes have consequences for the definition of this term. Furthermore, research into the operative function – one of the main functions in the lexicographical function theory – has shown that the lemma should not only be associated with words, but also other types of needs than those related to linguistic signs, something which e-dictionaries more easily are able to accommodate compared to paper-based dictionaries. This contribution, therefore, re-examines the lemma and investigates its role in e-dictionaries, and it suggests a provisional definition of this core lexicographical term that is applicable in e-dictionaries and in digital tools based on lexicographical data.

Keywords

lemma; digitalisation; e-dictionaries; searching; search options; navigating; the lexicographical function theory; operative function

1. Introduction

Discussions about the definition of *lemma* are not new (see e.g. Wiegand 1983), and all textbooks and handbooks on lexicography dedicate either a section or a whole chapter to the description of the lemma, explaining the different ways it can be incorporated into lexicographical tools (see e.g. Atkins/Rundell 2008: 160-199; Svensén 2009: 93-108). However, though a number of recent books on lexicography have focussed on lexicography in the digital era, e.g. Fuertes-Olivera/Bergenholtz (2011) and Granger/Paquot (2012), none of these have discussed what a lemma is when we look at it through the lens of digital technology. Furthermore, if we combine this digital lens with a function-theoretical lens (see e.g. Tarp 2008), even more critical questions arise to the traditional definitions of the lemma. Therefore, such points will be highlighted and discussed in this contribution. Firstly, I will present two traditional definitions of the lemma (Section 2), and secondly, based on the highlighted components of these definitions, I will provide some examples and new descriptions of the lemma from a technological and a functional perspective (Sections 3-5). Finally, I will suggest a tentative definition of *lemma* that incorporates the points discussed throughout this contribution and which is applicable in e-tools (Section 6).

2. Traditional Definitions

A definition of *lemma* is typically provided in handbooks on lexicography, and many of these definitions resemble each other. The definitions presented below are representative examples from two dictionaries of lexicography, and they will form the point of departure for the re-

* Heidi Agerbo
E-mail: missagerbo@hotmail.com

examination of this term provided in this contribution. The first one is a definition from *Dictionary of Lexicography* published in 1998:

Definition 1

The position at which an entry can be located and found in the structure of a REFERENCE WORK. The relationships of the lemma in the reference work are two-way: within the overall (e.g. alphabetical) MACROSTRUCTURE it constitutes the point of ACCESS where the compiler can place and the user can find the information listed; within the MICROSTRUCTURE it establishes the ‘topic’ on which the rest of the entry is a ‘comment’, e.g. the definition of the HEADWORD

Some authorities favour including all information preceding the definition within the notion of the lemma, i.e. all ‘formal’ items such as spelling, pronunciation and grammar, while others use the terms as a synonym for ‘headword’ or even the whole ENTRY. (Hartmann and James 1998: 83)

As expressed in the last part of this definition, lexicographers place themselves in either one of two camps according to what is included in the lemma: (1) the lemma functions as the guiding element (the headline) of a dictionary entry and is the primary treatment unit, and different information categories (e.g. grammar, orthography and definition) enable the user to retrieve lexicographical information about the lemma (e.g. Bergenholtz/Tarp 1994: 11; Gouws 2003: 36; Atkins/Rundell 2008: 162; Svensén 2009: 93-94); and (2) a lemma consists of everything preceding the first explanation or sense number in an entry – thus, it often contains information on orthography, spelling, inflection and etymology (e.g. Zgusta 1971: 249-252; Ilson 1988: 73). For sense number one, *lemma* and *headword* (or *entry word*, *look-up word* or *keyword*) are synonymous, and this is the most frequent use of the term, especially in contemporary meta-lexicography. This sense also corresponds to the second definition, which is from *Nordic Lexicographical Dictionary* published in 1997 (my translation):

Definition 2

lemma the expressional aspect of the linguistic sign, which introduces a dictionary article

– Lemma (Latin plural form: *lemmata*) is the meta-lexicographical term applied to what in workshop language and in the outside matter of dictionaries is called headword or entry word. Through the lemma, you find the entry you are looking for as the lemma also functions as a kind of heading to the entry. Except for cases such as proper nouns, etc., one specific form of the linguistic sign is usually chosen as the lemma, making this lemma form represent the full linguistic sign. In Western lexicographical tradition, nouns are normally lemmatised in the indefinite singular form, verbs in the infinitive and adjectives in the indefinite form, though a few dictionaries apply a different approach. In case of inflectional paradigms with word forms that deviate much from the lemma form, e.g. forms with umlaut or vowel shift, these forms may be included in the lemma list. In niche and nest articles, the lemma is often presented as an expression to which a tilde is added so the form only represents a part of the expressional aspect of the linguistic sign. = headword, entry word; lemma sign; run-on entry. (Bergenholtz et al. 1997: 172)

From these two definitions, we see that the most typical topics related to the notion of lemma are structuring (macrostructure, microstructure, niching, nesting, etc.), cf. Definition 1, and lemmatisation, cf. Definition 2. Two other topics typically raised in relation to the lemma are lemma selection and lemma lacuna (see Bergenholtz et al. 1997: 172-174). These topics have, to some extent, been discussed in relation to e-dictionaries, e.g. that lemma selection can benefit from logfile analyses of user look-ups in e-dictionaries (e.g. Wolfer et al. 2014), and that the lemma

stock of an e-dictionary and its entries are not limited by space restrictions (e.g. Tarp/Fuertes-Olivera 2016: 286-287). However, to this author's knowledge, no one has really questioned what consequences the e-environment has on our traditional understanding of the lemma itself, i.e. an understanding that is based on the paper environment. Therefore, in connection to the lemma, the following three points and their related sub-points derived from the two above-mentioned definitions will be investigated and discussed in this contribution: (1) dictionary structures in e-dictionaries, (2) e-dictionary consultation and (3) the "nature" of the lemma. In the following three subsections, which will each focus on one of the three points, examples of search options and entries from randomly investigated Danish and English e-dictionaries are included to show the strengths and weaknesses of these lexicographical solutions and, in some cases, to compare these solutions to hypothetical ones. The purpose of presenting these selected examples is to demonstrate that existing digitally-based solutions and possible future solutions force us to question the traditional definition of the lemma.

3. Dictionary Structures in E-Dictionaries

3.1 Macrostructure

The lemma is traditionally closely connected to the notions of *macrostructure* and *microstructure* as it constitutes a type of link between these two types of structures, cf. Definition 1. Figure 1 from Hausmann and Wiegand (1989: 329) illustrates a suggested connection between microstructure, macrostructure, lemma and entry in a paper dictionary, highlighting their interrelatedness.

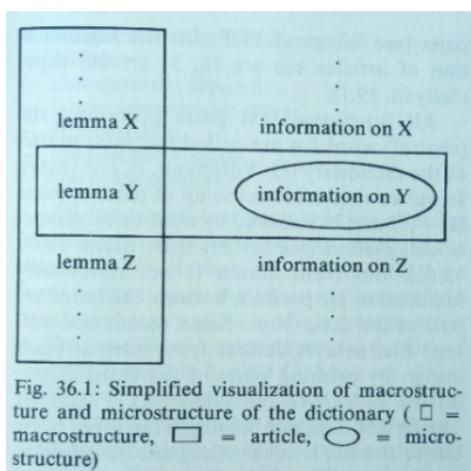


Figure 1. The lemma in relation to macro- and microstructure (Hausmann/Wiegand 1989: 329)

However, in e-lexicography, the concept of macrostructure is not nearly as important or even relevant for the discussion about lemmas as it is in relation to paper-based lexicography, at least not if we focus on digitally-born dictionaries. We may distinguish between two main ways of accessing an e-dictionary, which are connected to the medium of the dictionary and also how advanced its design is. The first approach is applicable both in e-dictionaries and paper dictionaries, whereas the latter is only applicable in e-dictionaries:

- through an alphabetical list (static or dynamic), i.e. an alphabetical macrostructure
- through a search field



Figure 2. Traditional lemma list and a search field in *The Danish Dictionary*

Figure 2 is a screenshot from the online and digitalised version of *The Danish Dictionary*, which offers both types of search facilities. On the front page, the user accesses a dictionary article by writing a word in the search field and clicking the button called “Søg” (Eng. Search). When this is done, the dictionary provides the user with the relevant entry to the left on the screen together with a list of neighbouring words to the right, i.e., an alphabetical lemma list known from paper dictionaries. Above this list, the search field (in a smaller format) and the search word are presented. The user can scroll up and down this list and click on a word if he wants to access the relevant entry. Though the user is provided with a lemma list, the default access function is the search field. In this case, we cannot talk about lemmas in relation to macrostructures in the traditional sense because the database cards from which dictionary articles are derived are not structured alphabetically or systematically (see Bergenholtz/Nielsen (2013), who explain how database cards are technically connected in a lexicographical relational database). Furthermore, related terms such as niching and nesting, cf. Definition 2, are somewhat outdated as e-dictionaries can easily lemmatise expressions that would traditionally have been considered sublemmas. For example, if we consider the idiom *kick the bucket*, *Ordbogen.com*’s English-Danish dictionary (Jørgensen/Levin 2018) is somewhat old-fashioned as it incorporates this into the same article as *bucket*, in this way following a typical paper-based incorporation and structuring of idioms – in this case, the user can either use the search term *bucket* or the idiom *kick the bucket* (as in Figure 3) to derive the same dictionary article with data about this idiom. Instead, in the online *Oxford Dictionary*, *kick the bucket* is a lemma that heads its own article in the dictionary, i.e., the user does not have to assess which word is the most essential one (*kick* or *bucket*) in order to find out where to look for the idiom in the dictionary (Figure 4). Thus, clearly, certain lexicographical terms related to the structuring and ordering of lemmas must be reassessed in an electronic environment. For example, the niche-alphabetic macrostructure with its niche entry lemma and the nesting macrostructure with its nest entry lemma could be made redundant in an e-dictionary.

kick the bucket

Engelsk (2) Tysk (2) Fransk (2) Spansk (1) Svensk (3)

Engelsk-Dansk

bucket *substantiv* <-S>

Oversættelser

- › **spand** - Spanden er fyldt med vand
- › **gulvspand** - Han fik en gulvspand i indflyttergave
- › [Søfart] **pøs** - Sejleren brugte pøsen til at tømme jollen for vand
- › **spandfuld** - Han tog en spandfuld sand fra stranden
- › **grab** - Gravemaskinens grab knækkede under vægten
- › (på entrenærmaskine) **graveskovl** - Graveskovlen havde sat sig fast
- › (på frontlæsser) **frontskovl** - Frontskovlen var læsset med sten
- › (på mudderapparat) **bøtte** - Han fyldte sten og mørtel i bøtten
- › (for transport af materiale) **hejsespand** - Tagdækkeren trak den tomme hejsespand på tage
- › (celle på vandhjul) **skovl** - Vandhjulets skovl blev fyldt med vand
- › **kick the bucket** » **dø** - Hun er bange for at dø

Figure 3. Searching with *kick the bucket* in Ordbogen.com’s English-Danish dictionary

[Home](#) > [British & World English](#) > [kick the bucket](#)

Definition of *kick the bucket* in English:

kick the bucket

PHRASE

informal

Die.

Figure 4. Searching with *kick the bucket* in *Oxford Dictionary*

3.2 Microstructure

From both definitions in Section 2, it can be inferred that lemmas are mandatory in dictionaries: The lemma heads the entry, providing the topic to which the rest of the data make a comment; i.e., all the data in an article are addressed to and therefore in some way dependent on the lemma. This is supported by Gouws (2003: 35), who argues that “[t]he lemma sign entry is the only compulsory entry in a dictionary article [...]”, and also by Wiegand (1983: 430) and Wolski (1989: 363) – the latter writes that “the lemma is a mandatory building block of the text in a dictionary article” (my translation). However, more than 10 years ago, Tarp (2008: 110-113) discussed the status of the lemma in relation to flexible article structures in lexicographical e-tools. He introduced the term *invisible lemma* (or *zero lemma*) in e-lexicography, that is, a lemma that is not directly expressed in the presentation of search results – what in this contribution will be referred to as an *unmentioned lemma*. For example, in many e-dictionaries viewed on a computer screen, a search field with the applied search string appears above the entry that is presented as a result of this search – in such cases, it is questionable whether the incorporation of the lemma is actually necessary or what the relation between the search string and the lemma actually is (see also Section 4.2). The e-dictionaries *The Danish Meaning Dictionary*, *Oxford Dictionary* and *Macmillan Dictionary* apply three different approaches in their presentation of search strings together with the lemma:

push-up

Engelsk (1) DDNO (1) Italiensk (1) DDGR (1) DDBO (1) Flere resultater ▾

Den Danske Betydningsordbog

push-up *substantiv*

1. teknisk indretning i en bh, som skubber brysterne op fra siden og på den måde giver en mere markant kavalergang
2. teknisk indretning, som løfter en kropsdel, fx ballerne, op, så de formes og fremstår mere struttende
3. styrkeøvelse, hvor man liggende med håndfladerne samt tæerne i gulvet og med lige ryg løfter kroppen op og ned ved hjælp af armræfterne

Figure 5. *push-up* in *The Danish Meaning Dictionary*

DICTIONARY

press-up

NOUN

British

An exercise in which a person lies facing the floor and, keeping their back straight, raises their body by pressing down on their hands.

Figure 6. *press-up* in *Oxford Dictionary*

CAREERLAND Take your career to the next level

press-up - definition and synonyms

NOUN [COUNTABLE] BRITISH

a physical exercise in which you lie down with your face towards the floor and use your arms to raise and lower your body

Figure 7. *press-up* in *Macmillan Dictionary*

In the first example (Figure 5), the user could have managed fine only with the search string, i.e., the dictionary could have applied the unmentioned lemma because the search string appears close to the data provided and, therefore, the lemma is redundant. The second example (Figure 6) resembles the first one, but in this case, the search field with the search string moves up and down when the user scrolls up and down in the entry (the search field has been locked), which is quite useful in a long entry in which case the user cannot see the lemma when moving further down the page. An interesting question arising from this example is whether the search string turns into a (dynamic) lemma as it overlaps with and replaces the static lemma; the search string becomes the heading to which the data in the article are addressed. Another interesting point in this example is that if a search is made with the plural form *press-ups*, then the search string in the search field is changed into the singular form *press-up* when the article with the lemma *press-up* is presented

(Figure 8). Thus, it could be argued that, in this case, the search string is actually transformed into a lemma in the search field. And then, again, the lemma (or the search string) is redundant.

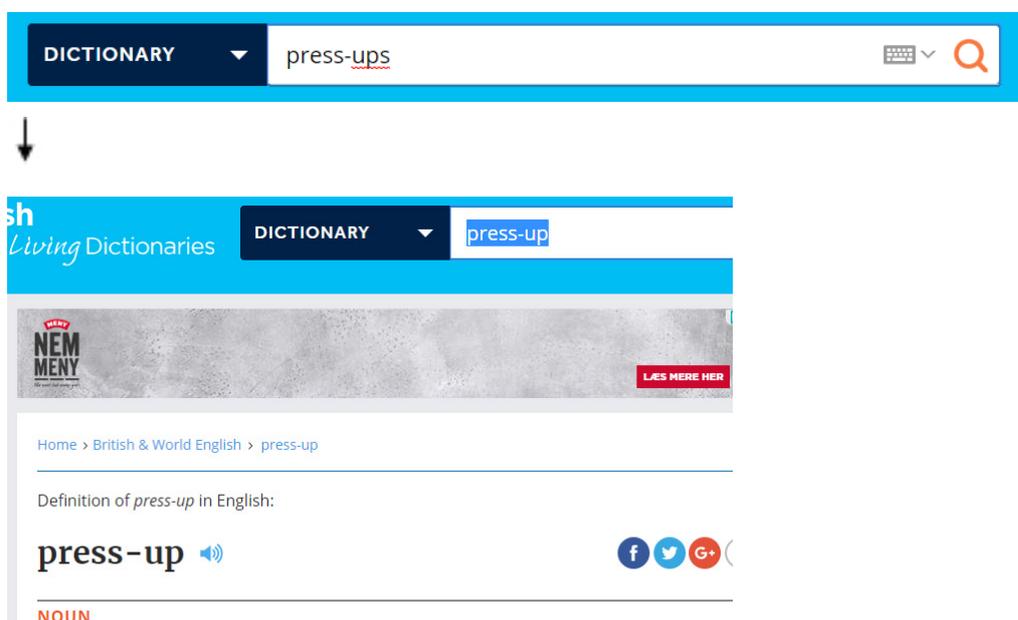


Figure 8. Searching with *press-ups* (plural) in *Oxford Dictionary*

In the third example (Figure 7), the search string is excluded from the search field in the presentation of the dictionary article. In this case, the word *press-up* was typed into the search field, but it was excluded from the presentation of the article. Thus, what element actually heads a dictionary article – or if something even needs to head an article – is questionable in an online environment, cf. Definition 2, which explains that the lemma functions as a type of headline in the dictionary article. Another current example concerning the role of the lemma as the headline is the presentation of dictionary articles in e-readers. Typically, if a person wants to understand a word in an e-text and accesses the integrated dictionary, he is presented with a traditional entry, cf. Figure 9, i.e., an entry that contains a vast amount of data. However, in this situation, all he needs is a meaning explanation; none of the grammatical information is needed – and not even the lemma because the reader can already see the relevant word in the actual text he is reading, i.e., the lemma becomes redundant. Thus, in such a digital tool, the lemma should not be considered a compulsory element in the presentation of lexicographical data intended to satisfy the user's information need. In this case, an unmentioned lemma would have been more suitable, especially as all the sense-related data can be presented in the same screen image.

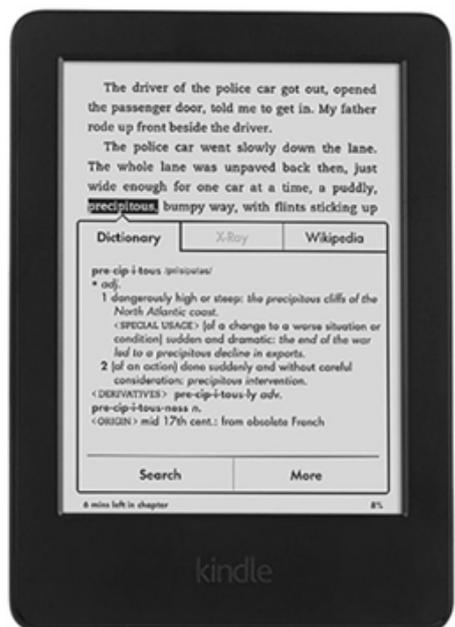


Figure 9. Using the integrated dictionary on a Kindle

4. E-Dictionary Consultation

4.1 Looking up, Searching and Navigating

When consulting a paper dictionary, we look a word up (e.g. by applying our knowledge about the alphabet) and try to find data that somehow explain this word or explain words or expressions that we think are grouped under this word, cf. the definition of *look-up* (*lookup*) in the online *Merriam-Webster's Dictionary*: 'an act, process, or instance of looking something up (as in a reference work or listing)'. Applying this definition, we can still talk about look-ups in an electronic setting, but it would be more accurate to use two related terms applied in information science: searching and navigating (Bothma 2011: 81).

Searching

The directed exploration of a defined information space with a defined objective in view and a defined search strategy. I.e. in either a simple or complex search interface, the user applies one or more search terms or phrases by means of which the search engine makes a search for articles that match this query.

Navigating

The exploration of a defined or undefined information space without using a defined strategy. There are two sub-categories:

- a) **Browsing**: The user has a defined objective, but has no defined search strategy
- b) **Surfing**: The user has no defined objective in mind, but is serendipitously following links that may seem interesting

According to Bothma (2011: 83),

Users usually combine searching and browsing to satisfy an information need. A search, be that through a search engine or in an e-dictionary, usually results in multiple results. Users then navigate

from the list of results to the actual full text article by following links. Based on the user's information needs, multiple links in an article to further articles can be followed (browsed) to enable the user to obtain sufficient detail to satisfy his/her information need. A user can also be distracted through interesting snippets and start surfing from article to article, not based on the original information need but on a serendipitous interest created by an interesting piece of information.

An important distinction could be made between *looking up* and *searching*: When a person makes a look-up, he looks through a list of words that function as the headings of dictionary articles, cf. Figure 10. The list contains selected words presented in their look-up form, cf. Definition 2. In the case of searching, a person does not look through a list of predetermined words, but types a search string into a search field. This search string may correspond to the exact form of the heading of the article, i.e., the lemma, but it could instead be one of its inflections or conjugations (Figure 11); or it could be a combination of words that lead to a relevant lemma (Figure 12). Thus, searching involves many more options than looking up; see also de Schryver (2012: 492) and Lew (2013).

Looking up

Going through a pre-selected list of words (a lemma list) that function as the headings to the entries and are, therefore, the entry points to the relevant data

Searching

Typing a search string, e.g., a specific word form or a combination of words, in a search field, which may lead to the data needed

- ▶ kvindefrigørelse sb.
- ▶ kvindefængsel sb.
- ▶ kvindefødt adj.
- ▶ kvindegruppe sb.
- ▶ kvindehad sb.
- ▶ **kvindehader** sb.
- ▶ kvindehandel sb.
- ▶ kvindehjem sb.
- ▶ kvindehjerte sb.
- ▶ kvindehold sb.
- ▶ kvindehoved sb.
- ▶ kvindehånd sb.



Figure 10. Alphabetical lemma list

Figure 11. Searching for *kvindehadere* (plural form)

fjendtlig +kvinder Søg

Søg med enkeltord eller med flere ved at skrive + eller - direkte foran søgeord.

DDBS (4) Andre søgninger

Hvad er det nu, det hedder?

kvindehader *substantiv* <en; kvindehaderen, kvindehadere, kvindehaderne>

1. kvindehader

Betydning

person, der nærer stærk modvilje og fjendtlighed over for kvinder og tydeligt giver udtryk for disse følelser

Figure 12. Searching for a relevant word based on its meaning

It should be noticed that, in the electronic environment, it is not always the lemma that leads the user to the relevant dictionary article, which is contrary to the description in Definition 2. In Figure 13, the word *dislikes* is emphasised when the mouse hovers over the word, and it functions as a link in the definition. If the user clicks on this word, he is taken to the entry which is headed by this particular word, or rather, its look-up form *dislike*. In Figure 14, the synonym *kvindehader* (Eng. woman hater) and the antonym *mandehader* (Eng. man hater) function as links to the dictionary articles in which they are the headlines. Thus, the user can be taken to a given article not simply by looking up or searching for a specific lemma, but also by navigating articles and their links.

misogynist

noun [C] • US UK /miˈsɒdʒənɪst/

dislikes

a man who **dislikes** women very much

Figure 13. The lemma *misogynist* in *Cambridge Dictionary*

misogyn¹ *substantiv* <en; misogynen, misogyner, misogynerne>

Betydning

person, der nærer stærk modvilje og fjendtlighed over for kvinder og undertiden nærer et decideret had til kvindeskønnet, og som tydeligt giver udtryk for disse følelser

Synonymer

kvindehader

Antonymer

mandehader

Figure 14. The lemma *misogyn* (Eng. misogynist) in *The Danish Internet Dictionary*

We need to examine these new ways of accessing dictionary articles as the lemma is no longer the only item used for gaining access. One of the things to investigate is the role of links. For example, since a link in an entry (e.g. an antonym such as *mandehader* in Figure 14) can take the user directly to the entry which the word functioning as the link heads, could we argue that this is a lemma as it somehow constitutes the access point to the place where the user can find the data he is

looking for? Or perhaps it would be better to introduce the term *lemma link* to distinguish between a link that leads to a lemma and other types of link such as dictionary external website links and dictionary internal association links.

If we introduce such a term, it should be made clear that a lemma link can occur in two forms: the headline itself as in Figure 15 or a reference text such as “Se artikel” (Eng. See entry) as in Figure 16.

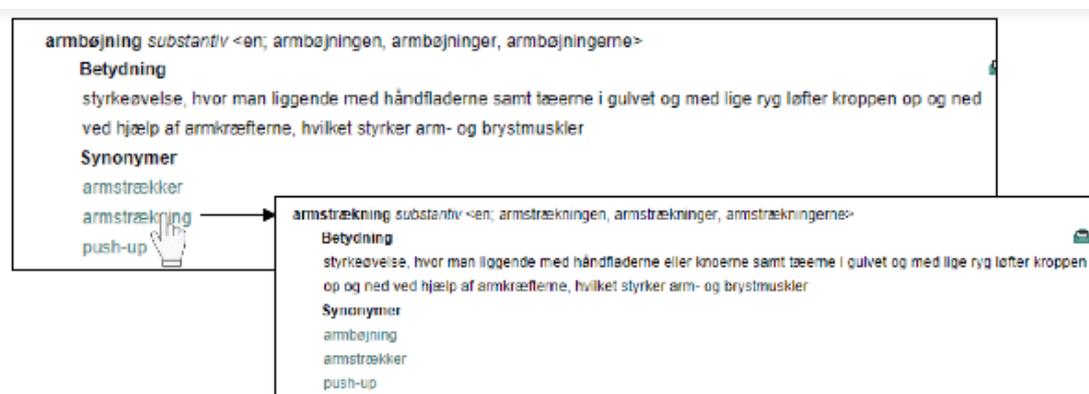


Figure 15. Using a lemma link, the word *armstrækning*, to access an entry



Figure 16. Using a lemma link, “Se artikel”, to access an entry

4.2 Lemmas, Search Options and Search Strings

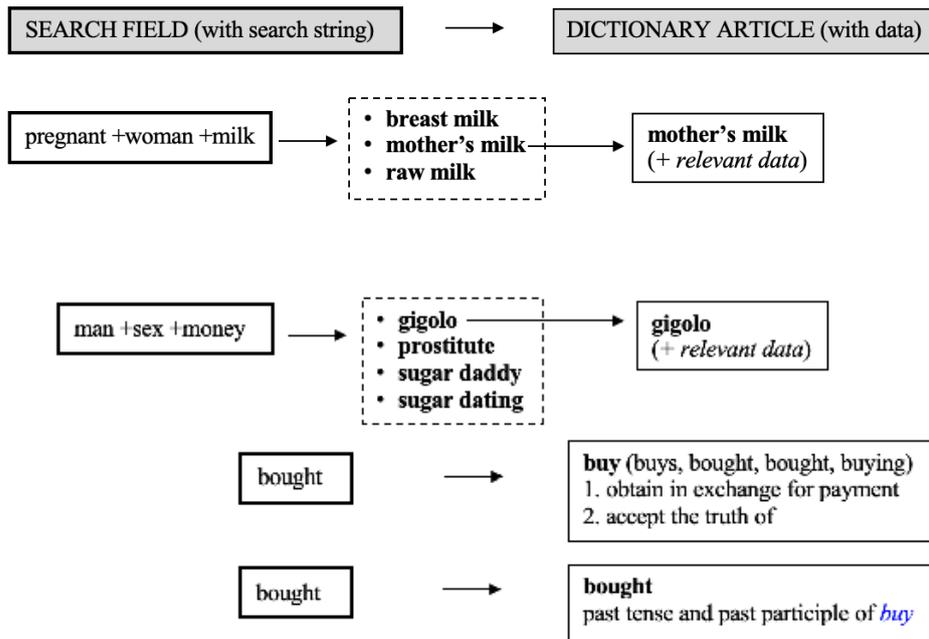
It is important to make a distinction between *lemma*, *search string* and *search option* (though see the discussion about the overlapping presentations of search strings and lemmas in Section 3.2) as the search item and the heading of a dictionary article are not always the same.

In the case of lemma versus search string, some modern e-dictionaries allow their users to search with keywords combined with Boolean operators, and many also allow their users to search with different word forms that lead the user to the relevant lemma (see some constructed examples below). Sometimes, the user is even presented with more than one (potentially) relevant lemma as a result of a complex search string. In Ordbogen.com’s Danish-English dictionary (Jørgensen/Levin 2018), the user can apply a search string consisting of several words (which, in this dictionary, cannot be combined with Boolean operators), and if these do not correspond to a lemma or sublemma, the user is presented with an entry for each search term on the same page. In Figure 17, a search was made with the string “hængende haver” (Eng. hanging gardens), and this resulted in two entries presented on the same page: one with *hængende* (Eng. hanging) as the lemma, and one with *have* (Eng. garden) as the lemma.

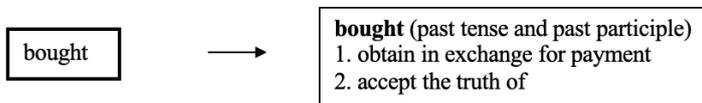


Figure 17. Result of using the search string *hængende haver*

In the first three of the constructed examples below, the search string and the lemma are different, whereas in the fourth example, the irregular form has been lemmatised and, therefore, the search string and the lemma are identical. In examples 1 and 2, the boxes with interrupted lines contain lists of suggested words from which the user can select a word, i.e., the user can click on one of the words to access the entry headed by this word.



It may also be possible to extract the exact search form (string) from the grammatical paradigm incorporated into the database and present this as the lemma in the dictionary article (a dynamic structure dependent on the specific search), i.e., leave out the traditional infinitive form as above and also leave out all the other verb forms (whether to include any grammatical information at all could be discussed); this would be useful in a reception situation in which the verb form applied in the text that the user is reading is the most important one.



To illustrate the difference between search string and search option, we can take a look at *The Danish Sign Language Dictionary* (Kristoffersen/Troelsgård 2012). This dictionary applies a unique definition of *lemma* based on sign language. The user is provided with four different predefined search options, some of them with a limited number of elements from which to choose:

- Handshape
- Location
- Text, including phrase search (“...”) and wildcard search (* and ?)
- Topic
- Combinations of the above



Figure 18. Search options in *The Danish Sign Language Dictionary*

In this case, there is a difference between search *options* and search *strings*: For search *option*, the user is provided with a number of ways to look for the needed data (this is, for example, also the case in several online picture dictionaries in which the user is provided with two search options: topic or text), whereas for search *string*, the user is given the specific option of inserting a search string (text) into a search field, i.e., search option 3 in the above list. Thus, this is another good example of the way that the electronic medium has opened up for new ways of searching for and presenting data that deviate from the traditional understanding of the lemma. Based on this example, an additional method of looking for data should be added to searching and navigating (and looking up): selecting.

Selecting

Looking for relevant information by clicking on one or more available search options (i.e., predefined categories and subcategories), e.g. “topic”

Furthermore, the concept of lemma applied in this sign dictionary is unique compared to the traditional definition of the lemma as the sign lemmas are represented through:

- A video clip of the base form (and potentially of one or more variants)
- A photo
- A unique gloss, which represents the sign throughout the dictionary, regardless of its actual meaning in a given context
- Drawings of the first location and handshape occurring in the description of the sign’s manual features

In this case, the lemma is not just the written word, but a combination of different visual and written elements (see an example in Figure 19). Therefore, to describe this, the term *complex lemma* may be suggested as it is made up of a number of components that cannot stand alone. This point leads to the next topic of this contribution, which concerns the nature of the lemma in e-dictionaries.

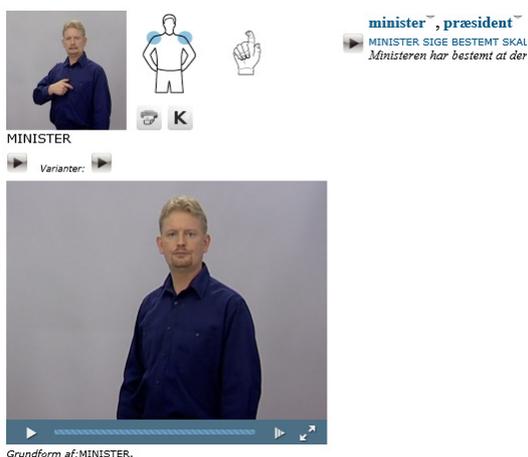


Figure 19. Presentation of a lemma in *The Danish Sign Language Dictionary*

5. The “Nature” of the Lemma

5.1 Linguistic and Non-Linguistic Lemmas

In his well-known discussion of the lemma, which was written more than three decades ago, Wiegand follows a linguistics-inspired approach as he makes a distinction between *lemma* and *lemma sign* (Wiegand 1983). According to his definitions, the lemma sign is the linguistic sign treated in a dictionary article, and the lemma is the graphic representation of the linguistic sign.

Strictly speaking, the information contained in the microstructure does not refer to the lemma but to the linguistic sign which the lemma arbitrarily represents in the macrostructure. The sign may be called the *lemma sign* [...]. Lemmatising will thus refer to the selection of one single morphological form whose function in the macrostructure is to represent the total set of grammatical and morphological forms of the linguistic sign treated in the microstructure [...] (Hausmann/Wiegand 1989: 329).

In addition to these two concepts, Wiegand (1989: 468) and Hausmann and Wiegand (1989: 329) also introduce the notion of Lemmazeichengestaltangabe, which Nielsen (1994: 123-124) translates as lemma form realization item, and which can be defined as the typographical realisation of the lemma sign found in a word list. The difference between these three lemma concepts can be illustrated with the following fictitious dictionary article for hippopotamus:

hip·po·pot·a·mus noun (countable), (plural: hippopotamuses, hippopotami)
a large African animal with a wide head and mouth and thick gray skin

Lemma = hip po pot a mus (a discontinuous lemma; see Wiegand (1983))

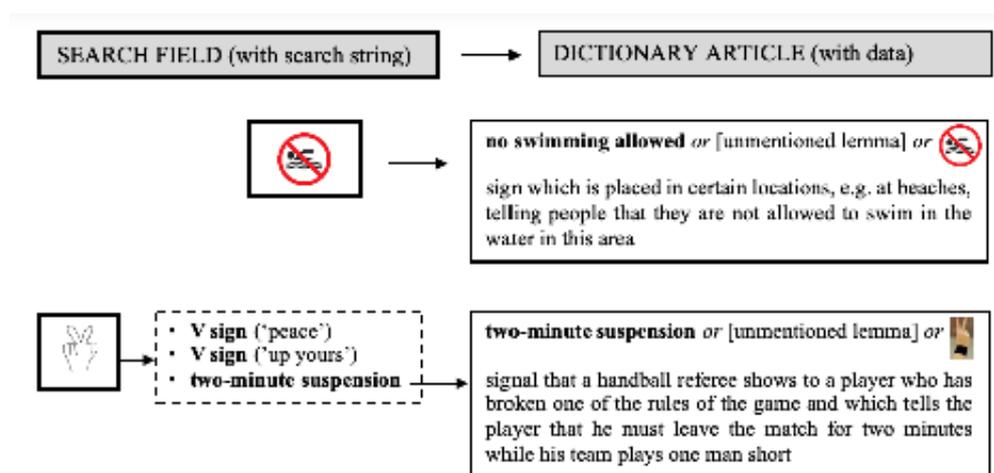
Lemma form realization item = hip·po·pot·a·mus

Lemma sign = the linguistic sign representing the concept introduced by the lemma hippopotamus

What is remarkable about this trichotomy as well as many more of Wiegand's points regarding the lemma is that these terms are clearly based on linguistic concepts and rest on a linguistic foundation. For example, the lemma sign resembles the traditional linguistic notion of lexeme, which is also defined as an abstract item containing a number of words (word forms) with the same basic meaning. Unfortunately, such linguistics-based discussions disregard other considerations concerning what can constitute a lemma, e.g. considerations related to technology (Section 5.1.1) and lexicographical theory (Section 5.1.2). Therefore, the rest of this section will focus on and discuss the lemma in relation to these two points.

5.1.1 Visual Lemmas

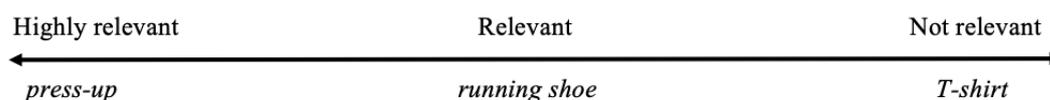
To this author's knowledge, no dictionary yet makes it possible for the user to search for information by "typing in", that is, uploading, drawing or uttering photos, illustrations or sounds even though the technology exists or is being developed, cf. the image and oral search options available through Google Chrome. Thus, perhaps in the future, a person could insert the following search strings and be presented with the following results in a comprehension dictionary (in the first example, a photo taken with the mobile phone has been uploaded; in the second example, the user has drawn the illustration directly into the search field):



These examples demonstrate that other types of sign than only linguistic ones can (and perhaps should) be incorporated as lemmas in future tools.

5.1.2 Function-Based Lemmas

The function theory has mainly discussed communicative functions, i.e., focussing on the understanding, writing or translation of linguistic signs. If, instead, we turn the focus towards lexicographical needs related to acting non-communicatively, we find that in relation to this function, a new understanding of the lemma will have to be introduced – or a completely different term is needed. In the case of an operative need, a person wants to know what to do or how to do something in a specific situation. At first sight, it appears as if some words are more relevant for dictionaries with an operative function than others. If we consider the words below – *T-shirt*, *running shoe* and *press-up* – these can be placed at each their point on a scale according to their relevance in an operative lexicographical tool for physical training:



For a lemma such as *press-up*, to a small extent, the definition of the word (useful for reception) will involve an explanation of how to do such an exercise. Though no lemmas are in themselves operative, it is quite clear that a word such as this might be connected to an operative need as the word is related to some kind of acting. This is the same for words such as *rock-paper-scissors*, which involves the act of playing, and *yodelling*, which involves the act of singing. A word such as *running shoe* does not involve an explanation on how to do something in the same way as *press-up* does. However, there are clearly operative needs that involve running shoes, e.g., how to choose which running shoes to buy or use, or how to clean your running shoes. A word such as *T-shirt* does not seem to be connected to any kind of operative need. People usually do not have problems finding out how to wash or choose *T-shirts* in relation to training as there are no specific requirements.

Though it may appear as if the words selected for lemmatisation in the examples above for a tool with an operative function are *press-up*, *running shoe* and *T-shirt*, it would be more precise to conclude that *how to do a press-up*, *how to wash running shoes* and *how to select running shoes* are the lemmas. It is actually not the words themselves that are related to an operative need. This means that it is not a specific word that forms the lemma, but rather a specific action. We could still call this a lemma, but in order to distinguish it from traditional understandings of the lemma, we could instead replace it with a different and practically applicable term such as “action”.

If we return to the topic of lemmatisation mentioned earlier, or perhaps, “actionisation”, we could “actionise” operative needs as full questions or they could be actionised as short phrases. These actions would then be connected to a specific nucleus. This approach is similar to the one applied in *The Dictionary of Fixed Expressions*, which involves the connection between a nucleus and the fixed expressions (see also Bergenholtz/Bothma/Gouws 2011: 36). In the example below, we could call *aben* (Eng. the monkey) the nucleus, and all the mentioned fixed expressions containing the word *aben* are the “lemmas” (if we choose to apply this word in this context) to which the meaning explanation is connected.

Figure 20. A search with *aben* in *The Dictionary of Fixed Expressions*

Thus, if someone has an operative need, the person should be able to make a search in an operative dictionary by typing in a search string (one or more keywords and/or images) and then the list of results should contain the different actions related to this search string. Instead of focusing on lemmas, we could distinguish between the following three core terms:

Search string

What the user types in the search field in the tool, i.e. word(s) and/or image(s), perhaps combined with Boolean operators

Nucleus

An abstraction in the form of a word to which one or more actions are connected

Action

An abstraction that represents the operative need that the user wants to satisfy and which functions as the name of the article that contains the relevant data

Some lexicographers might argue that the nucleus is the same as a lemma, but people are not searching for words in relation to operative needs; they do not want to act in relation to a word. E.g., if you want to know how to do a burpee, this has nothing to do with the word *burpee*, but instead the exercise burpee. However, in order to make it possible for lexicographers to collect and administer data on operative needs in a database, it is necessary to connect the data to these nuclei – if this is not done, there will be too many cards in the database to administer.

We also need to distinguish between (1) the database cards that contain different database fields, (2) search result lists and (3) dictionary articles. If we take *running shoe* as an example, a person may want to know how to clean his running shoes, i.e., this is the operative need he experiences. Let us imagine there are three database cards with the nucleus *running shoe* as these represent three different types of operative needs that are relevant in a lexicographical training tool: cleaning, choosing and tying running shoes, respectively. Each of the cards contains one or more actions, i.e. different written forms that all represent the same operative need (“synonyms”). No matter which action the user types into the search field, he will be presented with the relevant entry. Below, five different search strings (shoe; running shoe; clean; running shoe AND clean; How do I clean my running shoes?) have been applied to show the different result lists that appear from these searches.

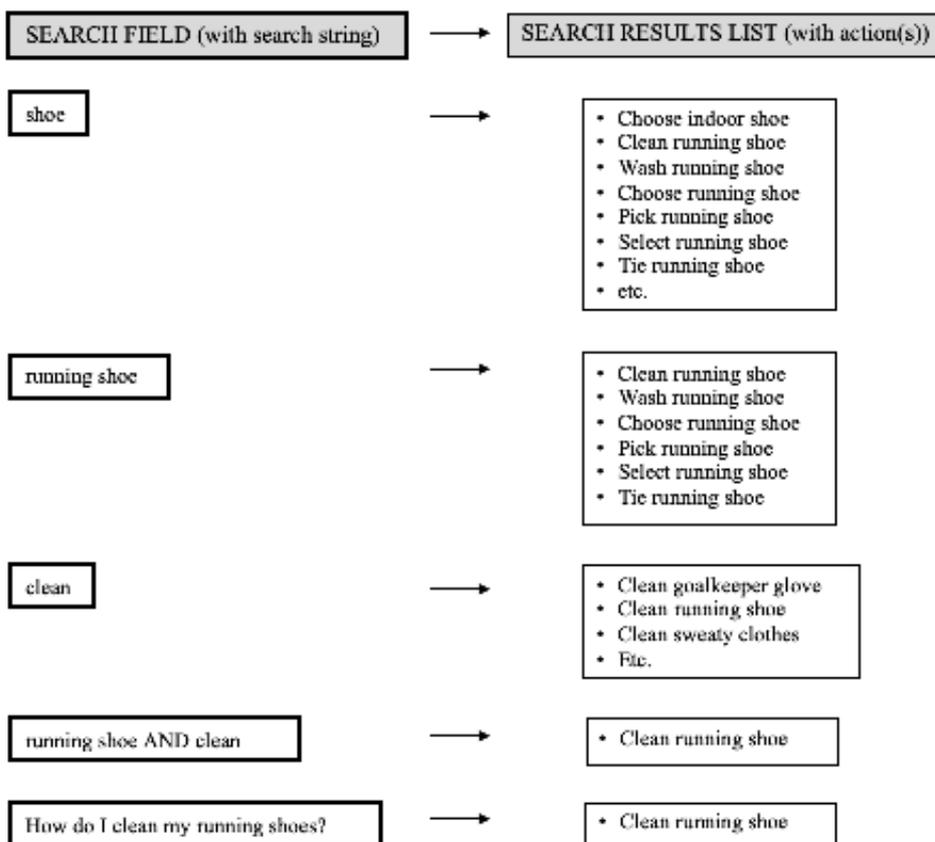
Database (3 database cards)

Nucleus	running shoe
Action(s)	choose running shoe
	pick running shoe
	select running shoe
Instruction	... [Instruction 1]
Level of difficulty	...
Equipment	...
Etc.	

Nucleus	running shoe
Action(s)	clean running shoe
	wash running shoe
Instruction	... [Instruction 2]
Level of difficulty	...
Equipment	...
Etc.	

Nucleus	running shoe
Action(s)	tie running shoe
Instruction	... [Instruction 3]
Level of difficulty	...
Equipment	...
Etc.	

Dictionary (5 different search strings)



Dictionary article (the action *clean running shoe* has been selected)

clean running shoe	
Instruction	...
Equipment	...
Variations	...
etc.	

The database cards are structured around one nucleus and one or more actions that are different representations of the same operative need. A dictionary article only contains one action (even if the connected database card contains more than one), but similar actions will be connected to the same data. Thus, searches made with synonymous actions, e.g. “clean running shoe” and “wash running shoe”, will result in similar entries. The only difference between the entries will be the action that functions as the heading.

If we follow this lemma approach in operative tools, i.e., abandoning the (traditional) lemma and incorporating actions (or at least, redefining the lemma), what are the consequences for dictionaries with other functions? Is the action just as useful a concept for non-operative tools as for operative ones? If we follow a typology of dictionary functions in which we distinguish between comprehension, acting (communicative and non-communicative) and cognition¹, we may draw the following picture of the relation between functions, needs and lemmas:

¹ This typology deviates a little from the usual typology applied by most function theorists, who distinguish between communication (including reception, production, etc.), cognition, operation and interpretation.

• Comprehension

Understanding linguistic or non-linguistic signs: the need is “I want to understand this sign”. The needs are similar here (it is all a matter of understanding something) and, therefore, they can be expressed just with the sign representing this need:

- I want to understand the meaning of *elephant* → lemma = *elephant*
- I want to understand the meaning of *red card* → lemma = *red card*
- I want to understand the meaning of *kick the bucket* → lemma = *kick the bucket*
- I want to understand the meaning of  → lemma =  (and *no swimming sign*)

• Communicative acting

Producing, revising or translating texts (where “text” is used in a broad sense) containing linguistic and/or non-linguistic signs: the need is “I want to **use** this sign communicatively”. This type of acting is related to the specific sign, e.g. a word or a traffic sign, and there is a limited number of ways to act with these (e.g. related to spelling, grammar and equivalents). Therefore, we could decide to let the sign represent the need(s):

- I want to spell the word *X* → lemma = *X*
- I want to know how to write *X* in the plural → lemma = *X*
- I want to know the grammatical gender of *X* → lemma = *X*
- I want to know the equivalent of the English word *X* in Danish → lemma = *X*
- I want to know the equivalent of  in Denmark² → lemma  = (and *dead end sign*)

• Non-communicative acting (= Operation)

The need is “I want to **act** in a certain way that does not involve communication”. The needs are quite diverse when we move beyond operative needs related to the “Highly relevant” category mentioned above, and the needs are not related to the words or signs themselves, only to nuclei. Therefore, these needs have to be expressed with actions instead of lemmas:

- I want to know how to do a **burpee**
→ action instead of lemma, i.e. *how to do a burpee* (nucleus = burpee)
- I want to know how to cut a **tree**
→ action instead of lemma, i.e. *how to cut down a tree* (nucleus = tree)
- I want to know how to plant a **tree**
→ action instead of lemma, i.e. *how to plant a tree* (nucleus = tree)
- I want to know how to make an **Easter letter**
→ action instead of lemma, i.e. *how to make an Easter letter* (nucleus = Easter letter)

² Notice that it is the sign to which we want to find an “equivalent”, not the word *dead end*. In Denmark, this traffic sign looks different.

• Cognition

The need is “I want to **gain knowledge** about something”. This “something” could be a word (e.g. if you are looking for etymological information about a word or want to know the difference in meaning between two words), but it could also be about something completely different, e.g. knowledge about the abortion limit in Denmark or the four stages of sleep. In *The Danish Internet Dictionary*, the information to satisfy such needs are found in the so-called lexical remarks in the entries for *svangerskabsafbrydelse* (Eng. termination of pregnancy) and *REM-søvn* (Eng. REM sleep). In these examples, the cognitive needs are represented by a lemma, or rather, by a word, which may not be the most useful solution as the user cannot know beforehand whether the articles represented by such lemmas will contain the relevant information. Cognitive needs are very diverse, and, unfortunately, not much research has been made into this type of need. Therefore, it is also unclear what the best solution would be in terms of applying actions and lemmas in a cognitive lexicographical tool. Perhaps a different term representing such cognitive needs should be introduced, e.g. *enquiry*. Examples of such needs could be:

- I want to know in what country the city **Mogadishu** is placed
- I want to know the etymology of the word *wife*
- I want to know about the rules of **football** for a school paper I am working on
- I want to know why it is impolite to use the word *Negro*
- I want to know if I can improve my running speed through **backwards running**
- I want to know the Danish rules for termination of **pregnancy**

For example

Nucleus: *pregnancy*

- Cognitive need 1: I want to know when you are allowed to terminate pregnancy in Denmark → enquiry instead of lemma, i.e. *when is it legal to terminate pregnancy*
- Cognitive need 2: I want to know how long it usually takes to become pregnant → enquiry instead of lemma, i.e. *how long to become pregnant*
- Etc.

5.2 User-Defined Lemmas

As a last short comment to the nature of the lemma in e-dictionaries, it is also important to mention Tarp’s introduction of the *user-defined lemma* or *ad hoc-lemma* (2008: 112). This is a lemma that is not predefined by the lexicographer, but is unique to the specific search made by the user and can be extracted from the database based on the available data and the way they are organised. Tarp presents the example of searching for nouns without singular forms in which case the ad-hoc lemma to which the search results are connected is *nouns without singular forms*. The following is a hypothetical search result of such a query.

nouns without singular forms

- amends
- clothes
- glasses

- pyjamas
- scissors
- etc.

This unique search result with a lemma and its assigned data is not derived from a fixed (pre-defined) structure, but has been possible to derive due to search options that can be provided in a digital tool. This lemma may appear awkward compared to traditional lemmas, but it is plausible that, for example, students of English would conduct such a search. As a result of this lemma approach, “the whole idea of what constitutes a lemma needs to be reconsidered” (Tarp 2008: 112).

6. Conclusion: A Function-Based and Technology-Based Redefinition of *Lemma*

It should be clear from the descriptions above that the role of the lemma in modern digital tools has needed to be questioned and investigated. From a function-based and technology-based viewpoint of the lemma, this contribution suggests the following provisional definition of *lemma* in e-tools:

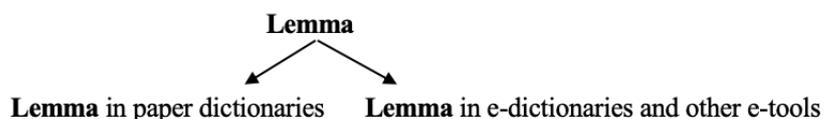
Lemma (in an e-dictionary)

The lemma expresses a lexicographical **information need** either implicitly (e.g. a need for understanding the meaning of a word or other type of sign – in this case the lemma is the representation of this word or sign) or explicitly (e.g. a need for information on how to act non-communicatively – in this case the need is expressed as a phrase or a question and could be called an **action**) in a standalone lexicographical tool or in a different type of e-tool that incorporates (integrated) dictionaries or lexicographical data. A lemma may function as the heading of an entry or be excluded if it is redundant (**unmentioned lemma**). A lemma can be (1) predefined by the lexicographer, (2) defined by the user (**ad hoc-lemma**) though his or her choice of lemma is limited by the data categories incorporated into the specific dictionary, or (3) semi-automatically extracted as it can be derived from some of the data incorporated by the lexicographer and used as a specific search string by a user.

This elaborate definition downplays the role of the lemma in e-dictionaries compared to paper dictionaries. In this digital context, it is not considered a compulsory component in an entry (it is not always needed as a heading), cf. Definition 2; it is not viewed as a linguistic concept, cf. Definition 2; and it is not seen as an element whose main role is to establish the relation between the macrostructure and microstructure, cf. Definition 1. Furthermore, it is not a static entity – which makes it difficult to determine the exact lemma stock (the amount of chosen lemmas) in a dictionary – and, most importantly, the lemma is not the only entity with which users can access the needed data from a lexicographical tool. It is clear that this definition of the lemma is not applicable for paper dictionaries. Instead, the following general description of the lemma can be applied – and this can then be divided into lemmas for paper dictionaries and lemmas for e-dictionaries and other e-tools:

Lemma

lexicographical unit that directly or indirectly expresses and represents an information need and which is used for finding and presenting data in a lexicographical tool or some other digital tool incorporating lexicographical data



The definitions of *lemma* provided at the beginning of this contribution were derived from two dictionaries published in the late 1990s. In general, these dictionaries base their definitions on the paper environment, but lexicography needs a new dictionary of lexicographical terms that focuses on the digital environment. I hope that this article will lay the foundation for the production of such a tool.

7. References

7.1 Dictionaries

- Bergenholtz, Henning/Agerbo, Heidi (eds.) 2018: *The Danish Internet Dictionaries [De Danske Internetordbøger]*. Odense: Ordbogen.com. www.ordbogen.com. Accessed 9 October 2018.
- Bergenholtz, Henning/Bjærge, Esben (eds.) 2017: *The Dictionary of Fixed Expressions [Ordbogen over Faste Vendinger]*. Odense: Ordbogen.com. www.ordbogen.com. Accessed 9 October 2018.
- Bergenholtz, Henning/Cantell, Ilse/Fjeld, Ruth V./Gundersen, Dag/Jónsson, Sigurd/Svensén, Bo 1997: *Nordic Lexicographical Dictionary [Nordisk Leksikografisk Ordbok]*. Oslo: Universitetsforlaget.
- Bergenholtz, Henning/Halkjær Jensen, Helene (eds.) 2018: *The Danish Meaning Dictionary [Den Danske Betydningsordbog]*. Odense: Ordbogen.com. www.ordbogen.com. Accessed 9 October 2018.
- Burkhanov, Igor 1998: *Lexicography: a dictionary of basic terminology*. Rzeszow: Wydawnictwo Wyższej szkoły Pedagogicznej.
- Cambridge Learner's Dictionary*. <https://dictionary.cambridge.org/dictionary/learner-english/>. Accessed 9 October 2018.
- Hartmann, Reinhard R.K./James, Gregory 1998: *Dictionary of Lexicography*. London: Routledge.
- Jørgensen, Britt West/Levin, Jacqueline R. (eds.) 2018: *English-Danish/Danish-English Dictionary*. Odense: Ordbogen.com. www.ordbogen.com. Accessed 9 October 2018.
- Kristoffersen, Jette H. (ed.) 2016: *The Danish Sign Language Dictionary [Ordbog over Dansk Tegnsprog]* (3rd edition). <http://tegnspog.dk/>. Accessed 9 October 2018.
- Macmillan Dictionary*. <https://www.macmillandictionary.com/>. Accessed 12 September 2018.
- Merriam Webster's Dictionary*. <https://www.merriam-webster.com/dictionary/dictionary>. Accessed 12 September 2018.
- Oxford Dictionary*. Oxford Dictionaries Online. <http://oxforddictionaries.com/>. Accessed 9 October 2018.
- The Danish Dictionary [Den Danske Ordbog]*. Ordnet.dk – Dansk Sprog i Ordbøger og Korpus. <http://ordnet.dk/ddo>. Accessed 9 October 2018.

7.2 Other literature

- Atkins, Sue/Rundell, Michael 2008: *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bergenholtz, Henning/Bothma, Theo/Gouws, Rufus 2011: A Model for Integrated Dictionaries of Fixed Expressions. In Kosem, Iztok/Kosem, Karmen (eds.), *Proceedings of eLex 2011*. Ljubljana: Trojina, Institute for Applied Slovene Studies, 34-42.
- Bergenholtz, Henning/Nielsen, Jesper 2013: What Is a Lexicographical Database? In *Lexikos* 23, 77-87.
- Bergenholtz, Henning/Tarp, Sven 1994: *Manual i fagleksikografi. Udarbejdelse af fagordbøger – problemer og løsningsforslag*. Herning: System
- Bothma, Theo: Filtering and Adapting Data and Information in an Online Environment in Response to User Needs. In Fuertes-Olivera, Pedro/Bergenholtz, Henning (eds.), *e-Lxicography. The Internet, Digital Initiatives and Lexicography*. London, Continuum International Publishing Group, 71-102.
- Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.) 2011: *e-Lxicography: the internet, digital initiatives and lexicography*. London/New York: Continuum.
- Gouws, Rufus 2003: Types of articles, their structure and different types of lemmata. In van Sterkenburg, Piet (ed.), *Practical Guide to Lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 34-43.
- Granger, Sylviane/Paquot, Magali (eds) 2012: *Electronic Lexicography*. Oxford: Oxford University Press.
- Hausmann, Franz Josef/Wiegand, Herbert Ernst 1989: Component Parts and Structures of General Monolingual Dictionaries: a Survey. In Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.), *Dictionaries. An International Encyclopedia of Lexicography, Vol. 1*. Berlin/New York: de Gruyter, 328-60.

- Illson, Robert 1988. Contributions to the Terminology of Lexicography. In Snell-Hornby, Mary (ed.), *Euralex '86 proceedings*. Zürich: A. Francke Verlag, 73-80.
- Kristoffersen, Jette/Troelsgård, Thomas 2012: The electronic lexicographical treatment of sign languages: The Danish Sign Language Dictionary. In Granger, Sylviane/Paquot, Magali (eds.), *Electronic Lexicography*. Oxford: Oxford University Press, 293-318.
- Nielsen, Sandro 1994: *The Bilingual LSP Dictionary: Principles and Practice for Legal Language*. Tübingen: Gunter Narr Verlag.
- Svensén, Bo 2009: *A Handbook of Lexicography: The Theory and Practice of Dictionary-Making*. Cambridge: Cambridge University Press.
- Tarp, Sven 2008: *Lexicography in the Borderland between Knowledge and Non-Knowledge*. Tübingen: Max Niemeyer Verlag.
- Tarp, Sven/Fuertes-Olivera, Pedro A. 2016: Advantages and Disadvantages in the Use of Internet as a Corpus: The Case of the Online Dictionaries of Spanish Valladolid-UVa. In *Lexikos* 26, 273-295.
- Wiegand, Herbert Ernst 1983. Was ist eigentlich ein Lemma? In Wiegand, Herbert Ernst (ed.), *Studien zur neuhochdeutschen Lexikographie*. Hildesheim: Georg Olms Verlag, 401-465.
- Wiegand, Herbert Ernst 1989: Arten von Mikrostrukturen im allgemeinen einsprachigen Wörterbuch. In Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.), *Dictionaries. An International Encyclopedia of Lexicography, Vol. 1*. Berlin/New York: de Gruyter, 462-501.
- Wolfer, Sascha/Koplenig, Alexander/Meyer, Peter/Müller-Spitzer, Carolin 2014: Dictionary users do look up frequent and socially relevant words. Two log file analyses. In Abel, Andrea/Vettori, Chiara/Ralli, Nataschia (eds.), *Proceedings of the XVI EURALEX International Congress: The User in Focus*. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism, 281-290.
- Wolski, Werner 1989: Das Lemma und die verschiedenen Lemmatypen. In Hausmann, Franz Josef/Reichmann, Oskar/Wiegand, Herbert Ernst/Zgusta, Ladislav (eds.), *Dictionaries. An International Encyclopedia of Lexicography, Vol. 1*. Berlin/New York: de Gruyter, 360-371.
- Zgusta, Ladislav 1971: *Manual of Lexicography*. Prague: Mouton.