*Henning Bergenholtz & Heidi Agerbo**

Meaning Identification and Meaning Selection for General Language Monolingual Dictionaries

Abstract

The traditional way for lexicographers to deal with polysemy in dictionaries is by applying the terms lumping and splitting. We will not follow this tradition. Instead, we argue that the identification and selection of meaning items (= polysems) should be treated in the same way as the identification and selection of lemmas. Identifying meaning items is comparable to identifying different words, the only difference being that meaning items share the same orthographic form. When identifying meaning items, we do not at the outset assume that a somewhat abstract meaning can be split up. Instead, we always assume that there may be many meaning items connected to a lemma, and we try to identify them – though for some lemmas, it is only possible to identify one meaning item. The process of identification involves a method that combines analyzing corpora and establishing a meaning relationship to references in the world (in this contribution called things), followed by a meaning formulation of the identified meaning items which can be used for reception situations. Not always – as in the case of lemma selection – will all the identified meaning items be included in the dictionary. The selection of identified meaning items will depend on the genuine purpose of the dictionary.

1. Lemma selection and polysem selection

In dictionary reviews, there is a clear tendency to emphasize lemma selection. This was the conclusion reached in a survey of reviews of monolingual dictionaries printed in the lexicographic journal LexicoNordica from 1994 to 2002, in which lemma selection constituted 13.9% of the 3,260 statements concerning lexicographic data, cf. this excerpt from the survey in Bergenholtz (2003: 20):

	Total	%
• price	30	0.9
• lemma selection	453	13.9
• macrostructure	72	2.2
• access	86	2.6
• grammar	289	8.9
• spelling	74	2.3
• semantics	247	7.6
• examples	99	3.0
• collocations	67	2.1
• idioms	40	1.2
• synonymy/antonymy	41	1.3

Naturally, it is relevant to the dictionary user and therefore also important for a reviewer to highlight lemma lacunas, but it is still striking that the number of comments on meaning descriptions only constitute half the number of comments on lemma selection. It is even more noteworthy that not even one of these comments focuses on polysem selection. Whether or not a polysem is miss-

* *Henning Bergenholtz & Heidi Agerbo*
Centre for Lexicography
Department of Business Communication
School of Business and Social Sciences, Aarhus University
Jens Chr. Skous Vej 4
DK-8000 Aarhus C
hb@bcom.au.dk – heap@bcom.au.dk

ing from a dictionary should have the same status as whether or not a lemma is missing from a dictionary – but this opinion is not shared by the reviewers.

A similar result is reached in user investigations in which users are asked to prioritize the different types of dictionary data according to importance as none of these investigations (that we are familiar with) mention polysem selection, a term which in this contribution is referred to as meaning item identification and selection. The same may be said about log file surveys. Most of these surveys only focus on lemma searches, see Bergenholtz/Johnsen (2005). Though in such surveys an added focus on dictionary functions (see Bergenholtz/Johnsen 2007) provides a more accurate view of the situations that prompt users to seek help in a dictionary, these log file analyses have yet not considered the point that a high number of dictionary articles contain two or more polysem articles that each has the same function as an individual dictionary article. There are several theoretical as well as practical arguments for presenting each polysem as a unique article, that is, dispense with both polysemy and homonymy in monolingual dictionaries; this view is argued in more detail in Bergenholtz/Agerbo (2014b).

Typically, different dictionaries will provide the same lemma with a different number of meaning items (polysems). This is only natural and therefore expected as different lexicographers apply different methods and are affected by different empirical backgrounds in the selection of polysems. However, what remains to be investigated is the connection between the formulation of the definition(s) and the chosen selection of identified meaning items. In the following section, this will be illustrated with a number of specific examples.

2. Searching for meaning items for a monolingual dictionary

As mentioned above, when dictionaries are compared, it is easy to demonstrate that there is a difference in the number of meaning items assigned to the same lemma in different dictionaries. Sometimes, this is based on an error made by the lexicographer working on one of the dictionaries (Bergenholtz 2005). This is not the type of comparison that this contribution will discuss. Instead, this contribution will focus on those differences for which it cannot easily be determined if a certain number of meaning items or a certain type of formulation is suitable. The following four e-dictionary entries of the lemma *calcium* exemplify this type of difference:

1. **Cambridge Advanced Learner's Dictionary & Thesaurus, British English**
   a chemical element that is present in teeth, bones, and chalk

2. **Merriam-Webster Dictionary**
   a silver-white divalent metallic element of the alkaline-earth group occurring only in combination

3. **Oxford British & World English Dictionary**
   The chemical element of atomic number 20, a soft grey metal. (Symbol: Ca)
   Calcium is one of the alkaline earth metals. Its compounds occur naturally in limestone, fluorite, gypsum, and other minerals. Many physiological processes involve calcium ions, and calcium salts are an essential constituent of bone, teeth, and shells

4. **Stedman's Medical Dictionary**
   A metallic bivalent element; atomic no. 20, atomic wt. 40.078, density 1.55, melting point 842°C. The oxide of calcium is an alkaline earth, CaO, quicklime, which on the addition of water becomes calcium hydrate, Ca(OH)2, slaked lime.

Most people will agree that the first dictionary article is brief and easy to understand, but only to a small extent is it informative. If a person reads the following real text example, s/he will most likely be confused rather than informed if s/he seeks help for reception in the first dictionary:

- Calcium is used as a constituent in lead alloys used for bearings and the sheaths for electric cables. Calcium is also used in aluminum alloys.

The same problem occurs with the second dictionary article. It does not show that you may need to have calcium in your food; that you have calcium in your bones; or that calcium can build up in a kettle. Neither does the third dictionary article nor the fourth one help to clarify the meaning of calcium in the quoted example. None of these four dictionary articles is wrong, but neither is any one of them optimal in the user situation described. As is always the case when a lexicographer does not simply copy the results from an existing dictionary made by another lexicographer, these definitions are different; They are human products. The question is, though, whether they can be improved in order to make them much more useful in information tools produced to aid dictionary users in specific user situations.

This example with *calcium* is characterized by the fact that none of the dictionaries provides more than one meaning. Later in this contribution, we will demonstrate that this would have been a better solution. If we turn to the lemma *stargazer*, we find, as for *calcium*, a difference in the number of meanings and the type of meaning formulation applied, but here we also have a case of meaning lacuna completely parallel to the often debated problem with lemma lacuna:

1. **Collins English Dictionary, British English**
   someone who observes the stars, such as an astrologer or astronomer
2. **The American Heritage Dictionary of the English Language**
   1. an astronomer or astrologer
   2. a daydreamer
   3. any tropical marine fish of the family Uranoscopidae, having the eyes at the top of the head
3. **Oxford British & World English Dictionary**
   1. *informal* an astronomer or astrologer.
   2. *Australian informal* a horse that turns its head when galloping.
   3. a fish of warm seas that normally lies buried in the sand with only its eyes, which are on top of the head, protruding.

In The American Heritage Dictionary of the English Language and Oxford British & World English Dictionary, some of the definitions are very brief, sometimes only made up of one or two synonyms (an *astronomer* or *astrologer*, a *daydreamer*). The meaning 'astronomer' occurs in all three dictionaries, and the meaning 'fish' occurs in two of the dictionaries. The statement that the meaning 'a horse that turns its head when galloping' is only an Australian use of the word is questionable as it is a meaning which exists in many different languages, e.g. in Danish for the equivalent *stjernekigger*. In total, these three dictionaries offer four different meaning items of *stargazer*. As mentioned in the introduction, in parallel to the term *lemma lacuna* we find examples of *meaning lacuna* (or *polysem lacuna*) in all of the dictionaries as none of them have incorporated all four meaning items.

From these descriptions of *calcium* and *stargazer*, we can extract two fundamental questions:

1. How many meaning items can and should a dictionary include for a certain orthographic word?
2. How should the definition be formulated?

The first point is the focus of this contribution whereas the second point is treated in Bergenholtz/Agerbo (2014a).

In the lexicographic literature, the focus has most frequently been directed towards the way that the definition should be formulated. What is termed polysemy has been offered much less attention, and the relatively few times this has been done, it has been reduced to a question of either lumping or splitting. Meaning formulation as the main theme is found in for example the following two special editions of two well-known lexicographic journals: Lexicographica 8, 1992, and

Dictionaries. Journal of the Dictionary Society of North America 14, 1993. Particularly interesting was the contribution by Wierzbicka (1993), in which she elaborates on the theme about definitions presented in Wierzbicka (1985, 39), in which she writes that ”an adequate definition must show fully what the word in question means”. In her book from 1985, she criticizes a specific dictionary article with a four line definition of the noun *cup* and suggests an 80 line definition, which she claims fulfills the requirement to definitions. This requirement as well as this suggested definition is criticized by Atkins (1993:9), who thinks that this type of definition is much too long. In Wierzbicka's article from 1993, she has possibly moved away from her former suggestion because in this article she now criticizes a definition of *dentist* for being too long:

> **dentist** a person who is skilled in and licensed to practise the prevention, diagnosis, and treatment of diseases, injuries, and malformations of the teeth, jaws, and mouth and who makes and inserts false teeth

This definition is, according to Wierzbicka, overly detailed, and it belongs in an encyclopedia rather than in a dictionary because a dictionary should inform about meaning and not, as in an encyclopedia, about general knowledge (knowledge and meaning). Therefore, Wierzbicka recommends and suggests ”a short definition”:

> **dentist** someone whose job is to look after teeth

It is noticeable that Wierzbicka does not even provide an argument for her claim that this (highly inadequate) definition is the best solution. She thinks that the criticized definition contains unnecessary elements, but this is not an argument. She approaches definitions as though the discussion about the boundary between semantic and encyclopedic knowledge has already been settled, but this is not the case; and is there even a difference between the two? Likewise, we may ask whether there even is a clear boundary between a dictionary and an encyclopedia. Concerning the last question, we refer to Bergenholtz (2012), who considers *dictionary* a term applicable to all lexicographic information tools. As for the first question, we refer to Haiman (1980) and Bergenholtz and Kaufmann (1996), who argue that it is not possible to distinguish between semantic and encyclopedic knowledge. Though we have not diverted from this opinion, we will express it differently in our search for meaning items: The question to be dealt with is not whether or not there is a boundary between semantic and encyclopedic knowledge; The crucial question to be taken as the point of departure for the identification and selection of meaning items (which is not even thematized in the theme issues of Lexicographica and Dictionaries) is: For what purpose is the dictionary going to be used? If a dictionary is expected to be an information tool, the answer is that the definition should contain exactly that which the user needs to know in order to fulfill his or her information need. Detailed investigations into this are provided in Bergenholtz/Agerbo (2014a).

Most printed as well as electronic dictionaries are polyfunctional. Simultaneously they try to fulfill information needs related to receptions problems, text production problems as well as the acquisition of knowledge about a thing or a term. In several articles, we have argued that monofunctional dictionaries, especially electronic information tools, are much better at helping the user solve his or her problem than polyfunctional ones, see Bergenholtz and Bergenholtz (2013a) and Bergenholtz and Bergenholtz (2013b). This is not the theme of the current contribution, but it is with the lexicographic function theory as our theoretical framework that we will discuss the criteria for the optimal reception dictionary.

In the rest of this contribution, we will focus on information tools whose genuine purpose is to provide help when a person has reception problems. In the case of text reception, the reader and thereby the potential dictionary user will be one of two types: 1. The user is almost certain about the meaning of the word but wants to make sure that this assumption is correct. In this case, a synonym may be sufficient. 2. The user has no idea what the word means. S/he needs a thorough explanation in order to understand the meaning of the word in a concrete text. Such a definition has to be detailed enough for the user to be sure that s/he can assign a meaning to the word in the text that s/he is reading.

Similar considerations can be made for a situation based on text production: 1. The user wants to use a word and is almost sure about its meaning, but wants to be absolutely sure and therefore needs a reminder, e.g. in the form of a synonym. 2. The user needs an explanation detailed enough for her/him to be able to use the word related to the exact reference that s/he has in mind. In this case, s/he will need a very long and exact definition, at least as exact as for the second kind of user for a text reception need as mentioned above. 3. Sometimes the user knows what s/he wants to say, but does not know the actual word. In this case, there are certain solutions for electronic dictionaries, see Bergenholtz (2013): With the help of a Boolean search, the user can type in different words in the search field and in this way look for lemmas with a certain meaning item that matches this search and in this way be guided towards the wanted word. For this type of search, the dictionary should also offer detailed and exact meaning items.

Based on these variations in user type and situation, we can conclude that a very abstract meaning - as the one proposed by Wierzbicka (1993) - cannot help solve reception problems for user type (2), nor for user types (2) and (3) in the case of text production problems. In the rest of this contribution, we will focus on the reception function in the sense of user type (2), but our tentative claim is that solutions for this function are normally also sufficient for text production.

3. A critical discussion of some concrete examples of meaning items

Below, we have found some more real text examples in which the word *calcium* occurs:

1. Calcium is the most abundant mineral in the body, but that doesn't necessarily mean everyone is getting enough of it in their diet.
2. Every element of a period has the same number of electron shells. Calcium (Ca) is in Group 2.
3. The negative effect of calcium is that it creates scale on pipes, hardware, and surfaces. This leads to high energy costs for heaters and expensive repairs for ice machines, coffee machines, and other appliances.

If a dictionary user looks up *calcium* in one of the dictionaries previously mentioned, it will be difficult for him or her to understand what the word means in these specific text examples. The first dictionary article (Cambridge Advanced Learner's Dictionary and Thesaurus, British English) will be able to help the user understand text example 1. For text examples 2 and 3 it is useless as it does not explain what *calcium* means in these contexts. It is not enough for the dictionary user only to be told that *calcium* is a "chemical element that is present in teeth, bones, and chalk" because how does this definition explain that it *creates scale on pipes, hardware, and surfaces*? The second dictionary article (Merriam-Webster Dictionary) describes it as a metal. This will – to a certain extent – help the user understand text example 2, but neither 1 nor 3. The third dictionary article (Oxford British & World English Dictionary) combines a number of characteristics of *calcium*, but this makes it difficult to understand what *calcium* really is in each one of the text examples: It is both a chemical element, more precisely a metal, and also an essential component for the development and maintenance of strong bones and teeth. Neither is the fourth dictionary article (Stedman's Medical Dictionary) able to help the user understand what the word means in text examples 1 and 3. What we find for these four definitions of *calcium* is that there is both a big difference in the description of the meaning item(s) as well as in the number of meaning items assigned to the lemma. In some of the dictionary articles, the problem is a lack of information in the definition if we assume that the user is expected to use it in a reception situation; in some of the other articles, the lexicographer has described several meaning items as only one meaning item, a description that is more or less useless as it does not explain the exact meaning of the word in the text examples above.

The most important criterion for being able to understand the use of a word has to be to connect it to that part of the world to which the word relates. This can be done fairly easily for concrete nouns by considering whether or not different applications of the word also relate to differences

in form and function (what does the object to which the word refers look like, and for what purpose is the object to which the word refers used?) of the thing to which the word refers. In addition to this, collocations, synonyms and word formations also provide great help in the process of identifying meaning items. For the lemma *calcium*, the collocations *to supply calcium to the body* and *calcium and magnesium content of ground water* point in two different directions, i.e. there are two different meaning items. Similarly, *lime* can be described as the synonym for two meaning items whereas *Ca* is a synonym for a third meaning item.

Taking the Danish monolingual dictionary Den Danske Netordbog as our point of departure, we suggest that the lemma *calcium* should be described as shown below, making it possible to understand all of the text examples mentioned above. The meaning formulations are based on the selection of collocations, synonyms and word formations taken from a Danish corpus, in the following translated into English:

1. element with the atomic character Ca and the atomic number 20; it occurs at standard temperature and air pressure as a soft, silvery-white metal that is easily affected by oxygen, thereby turning gray
synonym(s): Ca [= Ca]
collocation(s):
prikformel for calcium [= dot formula for calcium]
word formation(s):
calciumhydroxid [=calcium hydroxide]
calciumnitrat [= calcium nitrate]

2. vital mineral in the body, which is bound to the bones and strengthens the bone tissue; it is obtained through food, especially dairy products, but can also be purchased as a dietary supplement
synonym(s): kalk [= lime]
collocation(s):
behov for calcium [= calcium requirement]
depot for calcium [= calcium depot]
indeholde calcium [= contain calcium]
mangel på calcium [= calcium deficiency]
tilføre calcium til kroppen [= to supply the body with calcium]
tilskud af calcium [= calcium supplement]
variere indtag af calcium [= vary the calcium intake]
word formation(s):
calciumbeholdning [= calcium content]
calciumindtag [= calcium intake]
calciumtilskud [= calcium supplement]

3. calcium hydrogen carbonate dissolved in water, which when it is heated turns into a hard coating of calcium carbonate, such as scale in a kettle; it exists naturally in groundwater, for which the degree of hardness describes the amount of calcium hydrogen carbonate and magnesium
synonym(s): kalk [= lime]
collocation(s):
calcium i vand [= calcium in the water]
grundvandets indhold af calcium og magnesium [= calcium and magnesium content of the ground water]
varieret indhold af calcium i forskellige prøver [= varied calcium content of different samples]

We see that the first part of the lexicographic process for producing meaning descriptions consists of the following components: connecting the word to the world and identifying collocations, synonyms and word formations in a corpus. These make it possible to identify and select the three meaning items for *calcium* that we find in Den Danske Netordbog. Following this, the definitions of the meaning items are to be formulated, which involves the lexicographer identifying and selecting those components with which people normally talk about the word. For meaning item 2 of *calcium*, the lexicographer has chosen to write that it is a mineral in the body and that it can be consumed in different ways. Based on the analyzed corpus data, the lexicographer has assessed that these are the most important pieces of information that the user will need in order to under-

stand the meaning. It could also have been mentioned how much calcium a person needs and the consequences it has if a person does not get enough of this mineral. However, the lexicographer has assessed that this is not information that is necessary for the user to be able to understand the meaning, but it could possibly be added in a lexical note, which makes it possible for the user to gain more information if s/he is interested. This type of note, however, will not occur in a dictionary for reception, but in a dictionary directed towards users who want to gain knowledge about the word. What should be noticed is that when identifying and selecting lexicographic meaning items, the lexicographer applies a holistic working method; it is neither a linear nor an isolated process. In the first part of the process, the lexicographer makes an initial identification and selection based on collocations, text examples, word formations and synonyms while also relating the word to the world; but the formulation process may also lead to another round of identification and selection, which again is connected to the world, collocations, text examples, word formations and synonyms.

While the lexicographic process for describing *calcium* is relatively clear and straightforward, the process for the next example, *pigtail*, has more facets. In a number of selected English monolingual dictionaries, it is possible to find between one and four meaning items of *pigtail*; of these, Dictionary.com Unabridged includes subpolysems and The American Heritage Dictionary of the English Language only provides a synonym for one of its meaning descriptions, which is a reference to another dictionary article:

1. **Oxford British & World English Dictionary**
   - **1.** a plaited lock of hair worn singly at the back or on each side of the head
   - **2.** a short length of braided wire connecting a stationary part to a moving part in an electrical device
   - **3.** a thin twist of tobacco
2. **Cambridge Advanced Learner's Dictionary & Thesaurus, British English**
   length of hair that is tied at the back of the head or at each side of the head, sometimes in a plait (= twist)
3. **Merriam-Webster Dictionary**
   - **1.** tobacco in small twisted strands or rolls
   - **2.** a tight braid of hair
4. **Dictionary.com Unabridged**
   - **1.** a braid of hair hanging down the back of the head
   - **2.** tobacco in a thin, twisted roll
   - **3.** *Electricity*
   - a. a short, flexible wire used in connecting a stationary terminal with a terminal having a limited range of motion
   - b. a short wire connected to an electric device, as a lead or ground
5. **The American Heritage Dictionary of the English language**
   - **1.** A plait of braided hair
   - **2.** A twisted roll of tobacco
   - **3.** See flamingo flower

Below, there are 15 text examples from a Google search (November 2013) in which the word *pigtail* or the plural form *pigtails* occurs[1] . These examples from the corpus provide a preliminary overview, which may lead to the identification of meaning items for a dictionary:

1. With this dish you could add any kind of meat but I personally like to eat this with either saltfish or *pigtail*, the real traditional way.
2. Many times she would finish with a ribbon tied around the bottom of each *pigtail*.
3. Bit hesitantly, he stretched out his hand which held the flower, and gently put it in her *pigtail*.
4. The Manchu emperor ordered Chinese men to grow a *pigtail* and braid it otherwise you were disloyal.
5. Shop for your Chinese Hat With *Pigtail* 57511 at Easleys.com – the one stop shop for all your costuming needs!
6. This has made him even more garrulous than usual and he sits at one side of the table wearing a knitted red Tibetan hat with woven *pigtails* making snide remarks to each member of the team in turn.
7. But the *pigtails* said they'd take no more, because it made 'em silly.
8. …one of whom declared, that Captain Hemingway could take the ship to the White Mountains, gather a freight of cool air, and return in a given time with his eyes shut, as easy as he could twist an inch of *pigtail* from his tobacco box.
9. The water temperature gauge will install with its *pigtail* dangling – the *pigtail* will be tied later. Be sure to dress the water temperature gauge lead so it doesn't short out to the ammeter.
10. The lid is connected to the scanner by its *pigtail*.
11. I want everything to look OEM but the switch I bought didn't come with a *pigtail*. Where can I get a *pigtail* for the fog light switch?
12. I have a friend who owns a *pigtail*. I love that monkey, but he can be very unpredictable, his instincts will over ride if anything at all seems out of the norm.
13. The Flamingo Flower (herbaceous epiphytes) displays striking color and shape. Still using the Minolta 70-210 beer can, I zoomed in on the distinctive *pigtail*.
14. I'm still trying to figure out how to use these *pigtails* in a flower arrangement.
15. Basically, on the inside lines of my LF Havoc, the knot on the kite lines' *pigtail* is relatively small.

It is obvious that for several of the dictionaries mentioned above, the user runs into meaning lacunas when trying to find the meaning of the word in these examples, and there is an obvious meaning lacuna in all of the dictionaries when reading text example 1. In this example, the writer is neither talking about a hairstyle, a connector, tobacco nor a flower. Instead, the word refers to the little, curly tail placed on the hindmost part of a pig. It is a typical treatment of this type of word (compounds with an obvious "literal" meaning), that is, that the concrete and, what one may assume, obvious meaning is not incorporated[2]. But if a person comes across example 1 when read-

1 We are aware of a potential problem concerning the spelling, that is, whether it should be spelled as one word, two words without a hyphen or two words with a hyphen. Several dictionaries lemmatize it as one word, but the texts in which it occurs as one word are comparable to the texts in which it occurs as two words, though with a difference in frequency. Therefore, this discussion is irrelevant for this contribution.

2 In other examples of similar words, e.g. *lobster claw*, some dictionaries provide the definition 'a lobster's claw', in which case the lexicographer simply duplicates the words of which the lemma is composed. Nothing is mentioned about what characterizes a *lobster claw*, e.g. what it looks like and in what contexts the word normally occurs (e.g. cooking). This definition does not help the user understand what the word really means. Though this is a relevant discussion, it is not the theme of this contribution.

ing or listening to a text (i.e. reception), then none of the dictionaries fulfill their function, that is, helping the user understand the meaning of an unknown word, nor do they help the user understand a word whose meaning s/he is unsure of, but may already know.

All five dictionaries contain the meaning 'hairstyle' for the word *pigtail*, and four of the five dictionaries contain the meaning 'tobacco'. Two of the dictionaries also include meaning items about electrical devices, while The American Heritage Dictionary of the English Language has provided a synonym corresponding to the meaning 'flower'. With this dictionary, it is, to some degree, possible to understand the meaning of the word in text example 14. However, with only a synonym, the user will not learn anything about what type of flower it is or its characteristics. This type of meaning description in the form of a synonym is useful if the user only needs to confirm what the word means (cf. user type 1 above), that is, that the word in the text refers to a flower; but if the user is someone who does not already know the word or who is an interested layman, this meaning description is inadequate. For the other dictionaries, this use of the word is a case of meaning lacuna.

If we focus on the meaning which all the dictionaries share, 'hairstyle', it is phrased in the following ways:

1. plaited lock of hair worn singly at the back or on each side of the head
2. a length of hair that is tied at the back of the head or at each side of the head, sometimes in a plait (= twist)
3. a tight braid of hair
4. a braid of hair hanging down the back of the head
5. a plait of braided hair

In all of the dictionary articles, it is described as a way of making one's hair, but the amount of help provided by these dictionary descriptions varies. For text examples 2 and 3, it is easy to understand the meaning especially with dictionary articles 1 and 2 as these describe the fact that a *pigtail* may either consist of one or two braids, and dictionary article 2 also mentions that when there are two, these are not necessarily braided. If a text reads *many times she would finish with a ribbon tied around the bottom of each pigtail*, dictionary article 2 can help the reader understand that this is a hairstyle in which a bunch of hair sticks out on either side of the head. For text example 4, *the Manchu emperor ordered Chinese men to grow a pigtail and braid it otherwise you were disloyal*, it becomes more difficult to understand the use of the word if the dictionary user has to rely on the definitions provided above. If the user is told that this is a tight braid of hair, this will not clearly explain what the word actually refers to. In this text example, it is not only a matter of hairstyle. Instead, this is a waist-long braided ponytail – often only made by braiding the hair at the back of the neck and often also combined with the hair on the front of the head being shaved off – which was required to be worn by men in ancient China as a symbol of obedience and submission. Simply described as braided hair, the user will not be able to understand what *pigtail* actually means and refers to in text example 4, and therefore a dictionary with this definition is not a tool that can be used for solving reception problems. Producing definitions for an information tool is not a matter of finding one broadly described meaning item that combines as many meaning items as possible, but it is a matter of finding out what the word refers to in the world and with this information identify and formulate all of these meaning items. For *pigtail* as a hairstyle it is possible to distinguish between two physically different items in the world. Furthermore, it is possible to provide a linguistic argument for the identification of two meaning items with the use of collocations (though with a certain number of overlaps) and synonyms:

> Meaning item 1
> **Collocations:** make a pigtail, her girlish pigtails, wear pigtails, wear your hair in pigtails, long braided pigtails, a long pigtail
> **Synonyms:** plait; braid

Meaning item 2
Collocations: grow a pigtail, his long pigtail, cut his pigtail, a long pigtail
Synonyms: queue; cue

As mentioned for the lemma *calcium*, the first step in the lexicographic process for the production of meaning items for reception involves identification and selection followed by the formulation of meaning items. However, the formulation may lead to another round of identification and selection. In the case of the meaning 'hairstyle' for *pigtail*, we find that after the initial identification and selection, a further meaning item can be identified, which is a result of the formulation as well as a reshuffling of the identified collocations and synonyms. In the formulation, it is essential to describe the physical design of the item as well as the function of it. The lexicographer could for example also mention in what age this hairstyle was worn and what it meant if a man did not wear a *pigtail*, but this would more likely be information to be provided in a lexical note. There is no clear boundary between the data to incorporate in the definition field and the data to be applied in a potential lexical note, but as a rule of thumb the lexicographer should focus on the things that are spoken about or occur in the context when the word is used with a specific meaning. It is therefore not a matter of what from a linguistic viewpoint may be considered the nucleus and the only correct meaning of the word. The lexicographer identifies and selects information about a word based on what is normally said about the word and how it is used, and this is the information to incorporate in the definition.

For text examples 5 and 6, we find a related use of the word, but neither in these two examples are the dictionary definitions able to explain well enough what the meaning is if we consider what the word refers to in the world. In these examples, the meaning of the word is clearly closely related to the other meaning items that have been mentioned, but the word points towards something else in the world. Therefore, this use results in the identification of one – actually two – additional meaning items based on the following sets of collocations:

Meaning item 1
a black Chinese hat with attached pigtail
a Chinese hat with pigtail
the pigtail on the hat

Meaning item 2
knit a hat with pigtails
a knitted beanie with pigtail
a knitted red Tibetan hat with woven pigtails

As indicated with two sets of collocations, it is possible to identify two different types of hats (or rather, the items on these hats): one of them is a Chinese round hat, onto the backside of which a long artificial braid resembling the Chinese queue is attached. The other type of hat is often knitted and it has one, two or several pom-poms or strings attached to it. Thus, the lexicographer is able to identify two different meaning items. But for these meaning items, the lexicographer also needs to consider whether or not to select them for incorporation in a dictionary. In the current case, the lexicographer may choose to select one or the other, select both of them or deselect both of them.

For text example 7, none of the dictionary definitions is able to provide an adequate explanation of the word. In this example, the word does not refer to a braid of hair, but instead to people who wear this braid. This is an example of creative language use, comparable to potentially (and derogatorily) calling a person a *rug* or saying that *the coat came walking down the street with a grocery bag in each hand*. Because we only find very few examples of the word used in this way in the chosen corpus, and because as language users we cannot imagine this use of the word occurring in another corpus, the lexicographer should not include the meaning item in the dictionary; this meaning item can be deselected. Only if a dictionary user contacts the dictionary project and requests a definition of this meaning item should it be incorporated – because a dictionary is

an information tool that serves to fulfill its users' needs. In an electronic dictionary, the meaning item will then occur in the next update, while in a paper edition it will occur in the next edition. In this sense, lemma selection and meaning selection are very similar processes: When we identify meaning items connected to a specific lemma, these meaning items should be thought of as different words – they simply happen to be called the same thing orthographically speaking. As mentioned, these meaning items could, in theory, have been connected to each their different linguistic expression since the content is different. Therefore, each meaning item should be considered a unique word. But the way that this unique word is semantically connected to other words with the same linguistic expression is not important from a user perspective.

Had it been the case that the meaning of *pigtail* corresponding to a Chinese man's hairstyle only rarely occurred in a corpus, or that the word used with this meaning were considered incorrect, then the lexicographer should still include (i.e. select) such identified meanings in a dictionary for reception because it is possible to find the word used with this meaning in texts in a corpus, whether this use is correct or not. Thus, a user should be able to find a definition of a word that reflects how it is used in a text, even though the writer has used the word incorrectly, because the user is looking for an answer, which s/he would not be given if the definition were left out of the reception dictionary. In a production dictionary, the lexicographer should explain, e.g. in a usage note, that the use of the word with this (incorrect) meaning would be considered a mistake.

As already emphasized, the five dictionaries with the lemma *pigtail* have one (somewhat) shared meaning item, but for the rest of the meaning items they vary. The explanation behind the incorporation and exclusion of meaning items is not obvious. Why does The American Heritage Dictionary of the English Language choose to include a meaning item (only explained with a synonym) which has a low frequency and has been deselected in the other dictionaries? And why has none of the dictionaries included the meaning items referring to electrical devices as these are highly frequent in a corpus? We suggest that for the lemma *pigtail*, formulations corresponding to all of the 15 uses in the text examples above should be made for a reception dictionary, but it must be emphasized that these text examples are only excerpts from a corpus; there may be additional meaning items to include in a dictionary. If a sufficient number of examples of a meaning item can be found in a corpus (approx. >10), this meaning item should be included in a dictionary. Below, we present our suggested meaning formulations of the meaning items of *pigtail* that have been discussed in this contribution:

1. the small, curly body part on the hindmost part of a pig
2. the fleshy tail from a pig used for cooking, e.g. in soups, or prepared as a snack for dogs
3. a length of hair that is tied at the back of the head or at each side of the head, sometimes in a braid; mainly worn by women
4. a waist-long braided ponytail made by braiding the hair at the back of the neck, often also combined with the hair on the front of the head being shaved off, which men in ancient China were demanded to wear as a symbol of submission and obedience
5. item of artificial material resembling braided human hair, which is attached to the back of a Chinese hat and often worn as part of a costume imitating a Chinese man from ancient China
6. one, two or several pom-poms or strings of yarn or other fabric attached to a knitted hat, which often resemble a braid of hair or the tail of a pig

It should be noted that all of the meaning items are placed at the same level. The structuring of polysems in dictionaries varies as some apply a linear structure, describing all senses as same-level senses, which the lexicographer either calls homonyms or polysems, or as hierarchical structures, discriminating between different layers of senses, which all could be called polysems or be labelled either as polysems, subpolysems or subsubpolysems. Therefore, whether a particular word has been described as monosemous, polysemous or homonymous varies according to the specific dictionary approach. Some dictionaries also include phrases and idioms as unique mean-

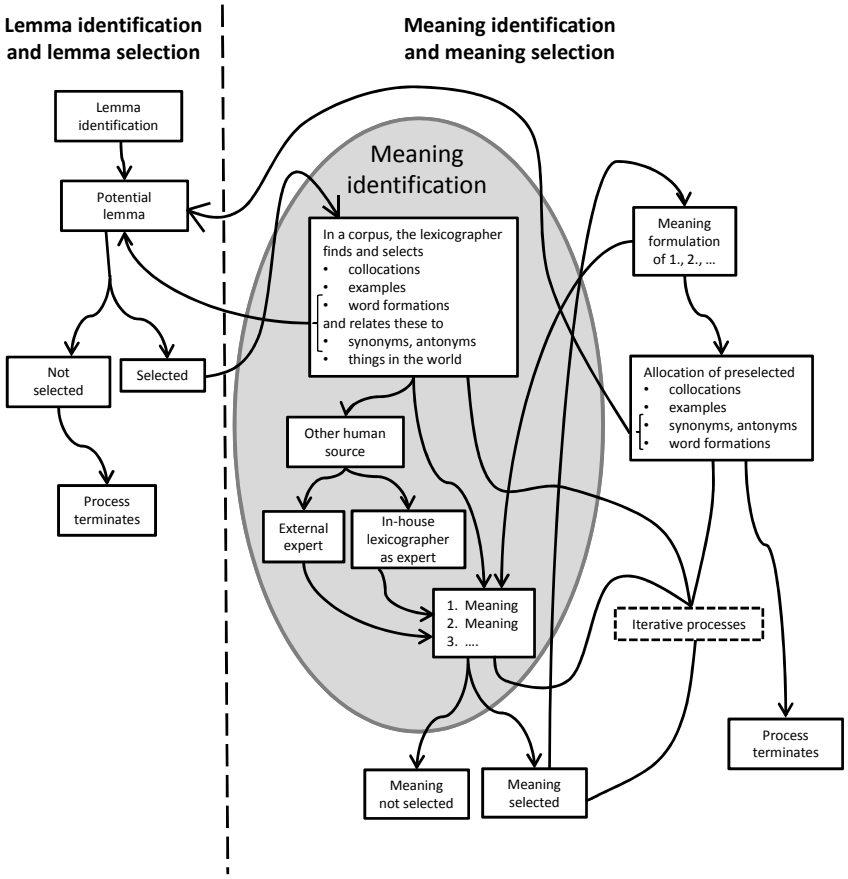
ing items whereas others add them as part of another dictionary article. Hence, quantifying and labelling senses as monosemous, polysemous or homonymous in dictionaries is not a matter of either-or, but of choice. For this reason, we do not consider *polysemy* a relevant term in lexicography. When lexicographers apply this term, they neglect the important focus: the use of the definitions, i.e. what the purpose of the definition is and what the function of the information tool in which it occurs is. Instead, they mistakenly tend to focus on the semantic relation between the different word senses. Thus, instead of talking about *polysemy* and *polysemous senses* (or *word senses*), we recommend applying the terms *meaning* and *meaning items* (as we have done throughout most of this contribution). How these meaning items are identified, formulated and structured is not based on how closely they are connected semantically but instead on the use of the dictionary, e.g. whether it is for reception or production, or e.g. whether it is for an expert, a layman or an interested layman.

4. The process of meaning identification and meaning selection

The lexicographic method used in the identification and selection of meaning items is, as described for *calcium* and *pigtail*, a process which is both holistic and iterative. This means that the order of each individual step as demonstrated in the numbered list below should be followed in the same way as the figure following this list demonstrates: At any one of the steps, the current step may be repeated; you may have to go back to a previous step; or you may have to start again from step 2:

1. A search for and selection of collocations, examples and word formations in a corpus is made. These help the lexicographer relate the word to specific things in the world, and they also help him or her infer potential synonyms. The selected word formations and synonyms may be lemmatized if they do not already occur in the lemma list.
2. An assessment of this selection is made, and this will lead to the identification of one or more meaning items. In this part of the process, the lexicographer may add more synonyms and make a first attempt to formulate meanings (see Bergenholtz/Agerbo 2014a).
3. Subject specific terms may be sent to an in-house lexicographer, as in the Danish project Den Danske Netordbog. In this project, the lexicographers are experts in different specialized fields such as law and economics (Bergenholtz 2013). In special cases, (unknown) external experts are contacted, but these experts never provide the formulations; they only act as consultants, i.e. they either confirm or explain the meaning of a word, which is then formulated by the lexicographer.
4. More of the preselected collocations, examples and word formations are allocated to the identified and selected meaning items. Sometimes the lexicographer will make another search in the corpus to find more collocations, examples, word formations and synonyms that are connected to the selected meaning items. The selected word formations and synonyms may be lemmatized if they do not already occur in the lemma list.
5. The formulation of meaning items is finished. This step may also lead to an additional identification of meaning items.

This process clearly demonstrates that lemma selection and the selection of identified meaning items are similar. In most cases, knowledge about the language history can explain why different meanings are expressed with the same word. But in principle, most meaning items could have had each their orthographic form – which they often get when translated into a foreign language. The process can also be described in the following way:



This process is rather complex as there are many iterative subprocesses. These repetitions are included in the figure, though not as a step of their own because the repetitions may occur at each individual step, which could be repeated here or repeated from one of the preceding steps. In order to simplify this account, we have only shown connecting lines going from the dotted box to the initial step in each main process of meaning identification and selection.

It is crucial to understand that the lexicographer does not assume that there is one meaning which could be split into two or more meanings; it is not a matter of lumping or splitting. Instead, it is correct to assume that any kind of meaning identification, meaning formulation and also meaning selection is made by human beings. Lexicographers describe language use in a way that makes the description useful in a specific dictionary with a specific genuine purpose. Consequently, we disagree with the English tradition, in which *sense division* is viewed in terms of either *lumping* or *splitting*, cf. Kilgarriff (1992 and 1997), Atkins and Rundell (2008) and Lew (2013), illustrated with the following quote:

> ”Any particular dictionary is written [….] with a particular editorial philosophy in relation to debates such as lumping vs. splitting” (Kilgarriff 1997, 100).

It is neither a matter of lumping nor splitting when for the lemma *pigtail* we distinguish between a tail on a pig, a flower or the hairstyle worn by Chinese men in ancient times. It is a matter of identifying different meaning items based on language use and the relation between language and phenomena in the world (concrete or abstract). In this contribution, we have focused on concrete nouns, i.e. nouns that refer to concrete things in the world. In future contributions, we will elaborate on meaning items connected to both adjectives, verbs and nouns related to abstract phenom-

ena. The process of identification and selection will be based on the suggestions provided in the current contribution, though with a number of elaborations.

5. References

5.1. Dictionaries and Other Information Tools

Cambridge Advanced Learner's Dictionary & Thesaurus, British English. Cambridge Dictionaries Online. http://dictionary.cambridge.org/ (Accessed 13 November 2013).

Collins English Dictionary, British English. Collins English Dictionary. http://www.collinsdictionary.com/ (Accessed 29 March 2013).

Den Danske Netordbog. Ordbogen.com – Danmarks Største Online Ordbog. http://www.ordbogen.com/ (Accessed 13 November 2013).

Dictionary.com Unabridged. Dictionary.com – Free Online English Dictionary. http://dictionary.reference.com/ (Accessed 13 November 2013).

Merriam-Webster Dictionary. Dictionary and Thesaurus – Merriam-Webster Online. http://www.merriam-webster.com/ (Accessed 13 November 2013).

Oxford British & World English Dictionary. Oxford Dictionaries Online. http://oxforddictionaries.com/ (Accessed 13 November 2013).

Stedman's Medical Dictionary. mediLexicon. http://www.medilexicon.com/medicaldictionary.php (Accessed 17 October 2013).

The American Heritage Dictionary of the English Language. The Free Dictionary by Farlex. http://www.thefreedictionary.com/ (Accessed 13 October 2013).

5.2. Other literature

Atkins, Sue 1993: Theoretical Lexicography and its Relation to Dictionary Making. In *Dictionaries, Journal of the Dictionary Society of North America* 14, 4-43.

Atkins, Sue/Rundell, Michael 2008. *The Oxford Guide to Practical Lexicography.* New York: Oxford University Press.

Bergenholtz, Henning 2003: Ordbogskritik i LexicoNordica. In *LexicoNordica* 10, 7-26.

Bergenholtz, Henning 2005: Falsche und richtige lexikographische Definitionen. In Gottlieb, Henrik/Mogensen, Jens Erik/Zettersten, Arne (eds.), *Symposium on Lexicography XI. Proceedings of the Eleventh International Symposium on Lexicography May 2-4, 2002 at the University of Copenhagen.* Tübingen: Niemeyer, 125-132.

Bergenholtz, Henning 2012: What is a Dictionary? In *Lexikos* 22, 20-30.

Bergenholtz, Henning 2013: The role of linguists in planning and making dictionaries in modern information society. In Kwary, Deny/Wulan, Nur/Musyahda, Lilla (eds.), *Lexicography and Dictionaries in the Information Age. Selected papers from the 8th ASIALEX Conference*. Surabaya: Airlangga University Press, 1-10.

Bergenholtz, Henning/Agerbo, Heidi 2014a: Extraction, selection and distribution of meaning elements for monolingual information tools. In *Lexicographica* 30 (in print).

Bergenholtz, Henning/Agerbo, Heidi 2014b: There is no need for the terms polysemy and homonymy in lexicography. In *Lexikos* 24 (in print).

Bergenholtz, Henning/Bergenholtz, Inger 2013a: A dictionary is a tool, a good dictionary is a monofunctional tool. In Fuertes-Olivera, Pedro A./Bergenholtz, Henning (eds.), *e-Lexicography: The Internet, Digital Initiatives and Lexicography.* London, New Delhi, New York and Sydney: Bloomsbury Academic, 187-207.

Bergenholtz, Inger/Bergenholtz, Henning 2013b: One Database, Four Monofunctional Dictionaries. In *Hermes* 50, 119-125. Bergenholtz, Henning/Johnsen, Mia 2005: Log Files as a Tool for Improving Internet Dictionaries. In *Hermes* 34, 117-141.

Bergenholtz, Henning/Johnsen, Mia 2007: Log files can and should be prepared for a functionalistic approach. In *Lexikos* 17, 1-20.

Bergenholtz, Henning/Kaufmann, Uwe 1996: Enzyklopädische Informationen in Wörterbüchern. In Weber, Nico (ed.), *Semantik, Lexikographie und Computeranwendungen.* Tübingen: Niemeyer, 168-182.

Haiman, John 1980: Dictionaries and Encyclopedias. In *Lingua* 50, 329-357.

Kilgarriff, Adam 1992: *Polysemy.* Unpublished PhD thesis, University of Sussex, UK.

Kilgarriff, Adam 1997: I Don't Believe in Word Senses. In *Computers and the Humanities* 31, 91-113.

Lew, Robert 2013: Identifying, Ordering and Defining Senses. In Jackson, Howard (ed.), T*he Bloomsbury Companion to Lexicography.* London: Bloomsbury Publishing, 284-302.

Wierzbicka, Anna 1985: *Lexicography and Conceptual Analysis.* Ann Arbor: Karoma Publishers.

Wierzbicka, Anna 1993: What Are the Uses of Theoretical Lexicography? In *Dictionaries. Journal of the Dictionary Society of North America* 14, 44-78.