

From Standalone Thesaurus to Integrated Related Words in *The Danish Dictionary*

Sanni Nimb, Nicolai H. Sørensen, Thomas Troelsgård

Society for Danish Language and Literature

E-mail: sn@dsl.dk, nhs@dsl.dk, tt@dsl.dk

Abstract

This paper presents a method of integrating Danish standalone thesaurus data automatically into a monolingual dictionary of modern Danish (*Den Danske Ordbog*, ‘The Danish Dictionary’) and discusses the results, including some of the problematic cases. The method draws on the detailed semantic grouping with two types of keywords in a well-structured XML-manuscript of a recently published thesaurus of Danish (*Den Danske Begrebsordbog*, ‘The Danish Concept Dictionary’) and on the fact that the two resources are linked on sense level, allowing for the automatic identification of semantically related thesaurus extracts for any given sense in the dictionary. The paper also presents a study of similar integrations of thesaurus data in four online English dictionaries, namely the *Oxford English Dictionary*, the *MacMillan Dictionary*, the *Merriam-Webster English Dictionary* and the *Oxford Dictionaries: English*, which we carried out in order to compare the structure of the underlying English thesaurus data as well as the resulting dictionary presentations with the Danish case.

Keywords: thesaurus, dictionary, linked data, synonyms

1 Introduction

This paper presents and discusses the automatic extension of a monolingual Danish online dictionary (*Den Danske Ordbog*, henceforth the DDO dictionary) with more synonyms and semantically related words based on the automatic identification and extraction of data from a Danish thesaurus (*Den Danske Begrebsordbog*, henceforth the DDB thesaurus), a dictionary describing concepts of modern Danish published in print in 2014. The two dictionaries are both compiled at the Society for Danish Language and Literature (DSL) and closely related since the vocabulary of the DDB thesaurus is based on and linked to sense descriptions of the DDO dictionary (see Nimb et al. 2014). Both resources are continuously being extended with new words.

The underlying XML data of the DDB thesaurus manuscript is annotated with coarse-grained semantic types allowing for the identification of persons, animals, artifacts, acts, events, etc. These annotations have already proved to be useful when compiling different types of formal lexicons (Nimb & Pedersen 2012, Nimb et al. 2013, Nimb et al. 2017). The future plan is to publish an online DDB thesaurus based on the printed book where it is possible to browse through the hierarchy of named chapters and sections, and to study the full vocabulary of each section with direct access from the words to the definitions and entries in DDO. But in this paper we describe the first online dictionary use of the linked data the other way around: excerpts of words from the DDB thesaurus sections inserted into the entries of the online DDO dictionary in order to supply the user of DDO with more information on related words. This type of function in online dictionaries is not new, and in Section 2 we study how a number of different English dictionaries have incorporated related words from thesaurus data, before in Section 3 we turn to the detailed presentation of the structure and content of the DDB thesaurus. In Section 4 we present the method used to identify and extract the relevant thesaurus data. In Section 5 we discuss some of the problematic results and draw some conclusions. We start by giving an overall presentation of the DDB thesaurus and the task.

1.1 The DDB Thesaurus

The DDB thesaurus is not written bottom up with groups of synonymous words as the starting point, but top down with a thematic structure in the form of named chapters and sections as the starting point, inspired by Dornseiff (2004), and based on the vocabulary in the DDO dictionary. However, the XML-structure of the manuscript was from the very beginning intended to facilitate the transfer of semantically related data back to the same dictionary. Due to this, the thesaurus presents words in a semantic order to the lowest subgroup level in the structure, meaning that the nearest other word to both sides of a word in the manuscript is most likely to be either the most similar synonym or near-synonym. The printed manuscript is in many ways comparable to the well-known English work *Roget's Thesaurus* (2002). Like the DDB thesaurus, *Roget's Thesaurus* divides the vocabulary in a series of named chapters and sections which contain semantically ordered groups of words which are either synonyms, or co-hyponyms or otherwise semantically related, although always belonging to the same word class. Likewise, *Roget's Thesaurus* also marks certain words as keywords based on semantics (i.e. a hypernym or the most common word in a group of synonyms). To illustrate the semantic structure of the Danish DDB thesaurus, we transfer its structuring principles to a group from *Roget's Thesaurus*, namely to the group of soft drinks. The group is in *Roget's Thesaurus* initiated by the word *soft drink*, underlined as a keyword in the text with italic letters. Semicolons divide subgroups of (sometimes synonymous) words within the group. All words are semantically, not alphabetically ordered:

“*soft drink*, teetotal d., nonalcoholic beverage; water, drinking w., filtered w., eau potable, spring water, fountain; soda water, soda, cream s., soda fountain, siphon; table water, carbonated w., mineral w., Perrier (tdmk), tonic water, barley w., squash, low calorie drink, mixer; energy drink; iced drink, frappé; milk, milk shake; ginger beer, ginger ale, Coca Cola or Coke (tdmk); fizz, pop, lemonade, orangeade, bitter lemon; cordial, fruit juice, orange j., apple j., tomato j., vegetable j.; juice box; coconut milk; tea, iced t., lemon t., herbal t., char, pekoe, orange p., Indian t., China t., green t., black t., Russian t., herb t., maté, coffee, café au lait, café noir, black coffee, white coffee, decaffeinated coffee, decaf, Irish coffee, Turkish c., espresso, cappuccino, latte, cocoa”

Based on the structuring principles of the DDB thesaurus, the group would contain not only one, but two types of keywords: *soft drink* as a keyword at first level (in bold letters), and a number of words in the text marked as keywords at second level (in bold, italic letters), namely (at least) *water*, *soda water*, *tea* and *coffee*, maybe also *fruit juice*. Furthermore, there would be supplementary divisions and therefore more subgroups (e.g. between *maté* and *coffee*). We also find that the keywords as well as the subgroups (divided by ●) are more transparent in the Danish thesaurus. Here we present the *Roget's Thesaurus* data as they would look in the DDB thesaurus structure:

“**soft drink**, teetotal d., nonalcoholic beverage ● **water**, drinking w., filtered w., eau potable, spring water, fountain ● **soda water**, soda, cream s., soda fountain, siphon ● table water, carbonated w., mineral w., Perrier (tdmk), tonic water, barley w., squash, low calorie drink, mixer ● energy drink ● iced drink, frappé ● milk, milk shake ● ginger beer, ginger ale, Coca Cola or Coke (tdmk) ● fizz, pop, lemonade, orangeade, bitter lemon ● cordial, fruit juice, orange j., apple j., tomato j., vegetable j. ● juice box ● coconut milk ● **tea**, iced t., lemon t., herbal t., char, pekoe, orange p., Indian t., China t., green t., black t., Russian t., herb t., mate ● **coffee**, café au lait, café noir, black coffee, white coffee, decaffeinated coffee, decaf, Irish coffee, Turkish c., espresso, cappuccino, latte, cocoa”

Our goal is to be able to present the users of the DDO dictionary with a small extract of the most related words and expressions from the DDB thesaurus next to the sense definition, and eventually already existing synonyms (without having to activate a link), such as some of the other types of coffee

when the word *espresso* is looked up. But we also want to allow the user to be able to activate a link to boxes presenting a much larger variety of words and expressions from the DDB thesaurus, namely any noun which is related to *espresso* in the sections where the word (in that specific sense) occurs, in this case it would be a box with the entire group of *soft drinks*, since this is the keyword (explanatory headline) of the coffee keyword.

The challenge we deal with is that the automatic extraction must cover any type of word (e.g. keywords at first or second level, or words which are not a keyword) as well as any size of the word group – from one to maybe 30-40 words. While some words have no direct synonymous or near-synonymous neighbors (e.g. *energy drink*), others have far more than we want to show in the small extract directly in the entry (e.g. *espresso*). Furthermore, we have no indication in the DDB thesaurus of whether the subgroup (or whole group) consists of synonyms (like “*Coca Cola or Coke*”), or words which are semantically related in other ways, e.g. being co-hyponyms as in the case of soft drinks above, or just somehow thematically related. The thesaurus contains no information on direct synonymy, neither on co-hyponymy. Our task is furthermore complicated by the fact that some of the DDO sense descriptions, but far from all, already contain a few manually selected synonyms, near-synonyms and/or antonyms that we do not want to repeat.

Before we describe the thesaurus structure and the chosen transfer method in detail, we take a closer look into how other dictionaries have integrated thesaurus data, taking into consideration also which type of thesaurus data they had at their disposal in order to see whether or not it resembles the DDB thesaurus data. We chose to study a number of English dictionaries since the language is closely related to Danish.

2 Thesaurus Data Integrated in English Dictionaries

We studied four English dictionaries which have either integrated thesaurus data in the entries, or present links to thesaurus data, namely the *Oxford English Dictionary* (OED), *MacMillan Dictionary*, the *Merriam-Webster English* online dictionary, and *Oxford Dictionaries:English*¹.

In the comprehensive *Oxford English Dictionary* (OED) there is a link from each sense description to thesaurus data taken from the standalone *Historical Thesaurus of the Oxford English Dictionary* but presented in OED style. The thesaurus is based on the information in the OED and linked to its senses. It organizes English words throughout the history into detailed hierarchies of meaning. In the OED the user is presented with a list of related words (including the headword itself) ordered historically, i.e. based on information on year of first recorded use with the oldest word first. The list is initiated by a headline presenting the hierarchy of chapters and section levels to which the word belongs in the historical thesaurus. For example, in the case of the adjective *high* in the sense ‘high-necked’, the user is presented by the headline ‘the world > textiles and clothing > clothing > types or styles of clothing > [adjective] < having specific parts < neckline’ to the list of adjectives: *high*, *low*, *low-necked*, *décolleté*, *semi-high* and *turtle-necked*. In the case of *happy* which has many synonyms, these are initiated by the headline ‘the mind > emotions > pleasure > happiness > happy’. The list of words is in this case long, and due to the historical order the most common words of modern English meaning ‘happy’ are spread in between rare synonyms and words which are not used any more.

In the *MacMillan Dictionary* there is again a link to a thesaurus presenting related words which are, also in this case, introduced by explanatory headlines. But the thesaurus behind the data consists in fact only of named groups of synonyms which are not organized in meaning hierarchies. But in

¹ We chose to leave out Dictionary.com which sometimes, but not always presents synonyms directly in the entries (e.g. for *hard* and *happy*, but not for *décolleté*, and, more surprisingly maybe, not for *stupid*). The dictionary links to more synonyms in Thesaurus.com, however the synonyms in the dictionary itself do not seem to come from Thesaurus.com.

contrast to the OED, the headline and a small excerpt of the data (three words) are presented directly in the dictionary entry itself. In the case of *décolleté*, the headline ‘Words used to describe clothes’, and the adjectives *A-line*, *backless*, *baggy* are presented, ordered alphabetically. Had the words been ordered on the basis of semantics, it would have been possible to present instead the most closely related words of *décolleté* (e.g. *low* with almost the same sense). The dictionary instead chose to simply present the first three words of the alphabetically ordered list. However, we do find cases in *MacMillan* where the synonyms are listed in semantic order, both in the integrated excerpt and when we click on the thesaurus link, e.g. in the case of *happy* where the adjectives ‘*happy*, *glad*, *alive*’, are presented directly in the entry, introduced by the headline ‘Feeling happy’.

The Merriam-Webster English online dictionary, which presents a semasiological dictionary and a thesaurus on the same homepage, contains one, maybe two precise (manually inserted) synonyms in the text with links to their sense description. The top lines of the thesaurus data are directly visible at the bottom of the whole entry (not at sense level), in the form of a few close synonyms with the headline ‘Synonyms’. When the link in the field is activated, the rest of the box becomes visible. The related words are presented in groups with the headlines ‘Synonyms’, ‘Antonyms’, ‘Near antonyms’, and ‘Related words’. In each of these, the word order is alphabetic, not semantic. The list of the synonyms of *happy* consists for example of *blissful*, *delighted*, *glad*, *gratified*, *joyful*, *joyous*, *pleased*, *satisfied*, *thankful*, *tickled*, thereafter a group of antonyms is presented before the near-synonyms are visible. Also in this case, the thesaurus data are not organized hierarchically based on meaning. As the only dictionary, *Merriam-Webster* gives a very detailed description of the relation between the different synonyms of the word and how they are used for different purposes at the bottom of the entry.

Finally, we take a look at the online **Oxford Dictionaries:English**. In the entry of a lemma, right below the definition, there is a link to synonyms. When it is activated, synonyms (or near-synonyms in a narrow sense) are presented in a box. Other types of semantically related words, such as co-hyponyms or antonyms, are not included in the presentation. Due to this, there is no data for *décolleté* – it has no synonyms – but a lot of data for *happy*: *contented*, *content*, *cheerful*, *cherry*, *merry*, *joyful*, *joyous*, *jolly*, *joking* and so on. There might also be more than one group of synonyms presented. In the case of *stupid*, for example, there are two: one initiated by the keyword *unintelligent*, another initiated by the keyword *foolish*. There is a link from the box to the thesaurus homepage itself, but like in the case of both *MacMillan Dictionary* and *Merriam-Webster*, the thesaurus data consists only of groups of related words which are not organized hierarchically, initiated by a bold headword (e.g. **contented**). Table 1 gives an overview of the different dictionaries, and compares the thesaurus data with the DDB thesaurus.

Two of the English dictionaries present a small extract of thesaurus data directly in the dictionary entry before linking to a larger group of related words. In one case the extract is introduced by a headline. Two dictionaries have at least to some degree ordered the thesaurus words semantically, but none as detailed as in the DDB thesaurus. Only one English dictionary contains manually selected synonyms as part of the dictionary sense description itself like the DDO dictionary, and only one, the OED, bases the integration on a thesaurus with meaning hierarchies similar to the ones we find in the DDB thesaurus, allowing for the automatic extraction of precise headlines of related words. In DDB however, we have no labels at the lowest grouping levels (for *décolleté* corresponding to ‘types or styles of clothing < having specific parts < neckline’). Two of the dictionaries, the *Merriam-Webster English Dictionary* and the *Oxford Dictionaries:English*, make a clear distinction between presenting synonyms (in a broad sense) and other related words, allowing them to give the thesaurus data in either separate groups, or to leave out completely words not being either a synonym or a closely related near-synonym. The data of the DDB thesaurus lacks the information which is needed to present synonyms and near-synonyms in a narrow sense, and less related words such as near-synonyms in a broad sense (for example co-hyponyms and thematically related words).

Table 1: Overview of compared thesaurus integrations.

Based on thesaurus with:	DDO	OED	MacMillan	Merriam-Webster	Eng. Oxford Dic
Meaning hierarchy	Yes	Yes	No	No	No
Distinction in thesaurus data between synonyms and other related words	No	No	No	Yes	Yes (only synonyms are shown)
Semantic order at lowest group level	Yes	No, historical	Sometimes	No, alphabetic	Sometimes
The solution in the dictionary contains:					
Small extract of directly shown data	Yes	No	Yes (with headlines)	Yes	No
Link to large extract of data in box	Yes	Yes	Yes	Yes	Yes
Large extract in box has explanatory headlines	Yes	Yes	Yes	No	No (but headword in bold)
Already manually selected related words to be considered	Yes	No	No	Yes	No

The solution we chose in the case of the integration of DDB thesaurus data into the DDO dictionary is most similar to the *MacMillan Dictionary*, although without an explanatory headline of the small, directly shown extract of closely related words. But in contrast to *MacMillan* we have to consider the already manually inserted synonyms, near-synonyms and antonyms in DDO: these must be left out from the small extract. Furthermore our headlines are most likely to be less precise, since they are transferred automatically from the title of the section in which the word group is only one of several groups. In *MacMillan*, each group of synonyms and near-synonyms has probably been assigned headlines manually. We will instead profit from the many keywords in the thesaurus data, which in most cases function quite well as headlines of the following group of words, as in the case of *Oxford Dictionaries:English*. In some cases, however, they do not, and we will discuss this in Section 5. In the next two sections we give a detailed presentation of the structure of the DDB thesaurus and the method of extracting data from the thesaurus to the dictionary DDO.

3 The Structure of the DDB Thesaurus

The DDB thesaurus is organized into 22 named chapters and 888 named sections. It contains approximately 200,000 words and expressions, almost all of which are linked to a specific sense in the DDO dictionary² by shared ID numbers, in the case of collocational expressions to at least one of the included word senses (see Nimb et al. 2014). The overall semantic grouping principles in each section of the thesaurus are based on the distinctions between 1st, 2nd and 3rd order entities (Lyons 1977) and between semantic types and relations (Pustejovsky 1995). In each of the 888 sections a number of subgroups of words, typically sharing the same semantic type, are listed one after the other in five word class groups ('noun', 'verb', 'adjective', 'adverb' and 'other'). The semantic order of the words in a subgroup is based on linguistic characteristics such as prototypicality, frequency, broad versus narrow meaning, style, etc., based on the descriptions in DDO as well as the subjective judgment of

² DDO contains approximately 140,000 sense descriptions

the lexicographer who weighs up the different characteristics against one another when the order is established. The subgroups might consist of co-hyponyms or maybe just thematically related words, not only synonyms. Some of the subgroups are initiated by a keyword, as described in Section 1.1. First level keywords indicate large shifts in meaning between the semantic types in the subgroups, for instance between artifacts and persons. Second level keywords indicate meaning shifts within the same semantic type, e.g. between different types of artifacts (like *tea* and *coffee* as described above). First level keywords function as a kind of ‘headline’ until the next one, while second level keywords do the same, but only until the next keyword, no matter whether it is a first or second level one. In Figure 2 we illustrate the structure in a formal way. The first word, ‘A’, functions as the headline of all the words until ‘X’. The words in the first group, ‘A, b, c, d’ are more closely related to one another than they are to the words in the next group, ‘e, f, g, h’. Likewise, ‘j’ and ‘k’ are more closely related to ‘i’ and ‘m’ in their own group than to ‘g’ and ‘h’ in the preceding group, although still closer to ‘g’ and ‘h’ than to ‘s’ and ‘t’ in the preceding group, which starts with a keyword, ‘r’.

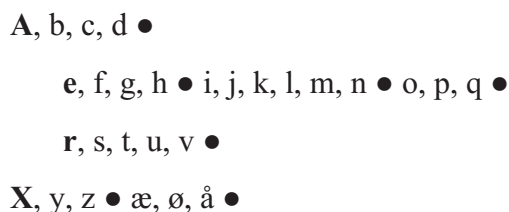


Fig. 2: The structure of the DDB thesaurus data. Each letter represents a word.

During the dictionary-making process of the thesaurus we found the structure very flexible. It allowed us to group words without having to specify the degree of synonymy, since this demands a rather detailed study of the use of the words. Instead we aimed at including as many words as possible from the DDO dictionary, i.e. also those which do not have any real synonyms. We focused on the communicative purposes of the word order, trying to make a fluent ‘text’ with as slight a change in meaning as possible from one word to the next. Whenever there was a big meaning change, the word was changed into a keyword. The drawback of this flexible structure is that we cannot identify the synonyms and closest near-synonyms from groups of other related words in the manuscript, and we are therefore not able to present automatically extracted very precise synonym groups in the same way as e.g. *Merriam-Webster English* online dictionary. The related vocabulary that we extract from the DDB thesaurus is most likely to consist of either synonyms, near-synonyms or co-hyponyms, but might also contain a broader variety of related words than any of the English dictionaries do.

4 Extraction of Data from the Thesaurus to the DDO Dictionary

In order to automatically identify exactly which DDB thesaurus data to extract, the semantic relevance of the surrounding words in a section where the DDO sense in question is represented is dynamically calculated. The calculation is based on the semantic order of the words as well as on keywords and subgroup structure. We benefit from the fact that the closest words in the DDB thesaurus structure within the same subgroup are most likely to be also the semantically closest word, and that the first keyword to the left is almost certainly closely related, since it was chosen as a kind of headline of the word.

Initially we determine the ‘best’ section in the case of multiple occurrences in the DDB thesaurus, based on the number of words within the scope of the first keyword to the left: the higher the number,

the better (i.e. the more likely to contain most synonyms and near-synonyms of the word sense). We then select one to six words from the immediate surroundings of our headword (first the first one to the left, then the first one to the right, then the second one to the left and so forth), include the keyword to the left but exclude the headword itself and eventually manually inserted DDO synonyms. We present the words as a list integrated directly in the DDO entry, in the DDO style. Secondly, we present a link which gives access to an even larger extract of words, presented in a box initiated by the corresponding section name in DDB and with the search word highlighted in red. In the case of multiple occurrences of the specific sense of the search word, data from different DDB sections is presented in separate boxes which are listed according to their section number in the DDB thesaurus. See Figure 3.

The image shows a screenshot of the Danish DDO dictionary entry for the verb *slentre*. On the left, the entry includes the word, its part of speech (verbum), inflection (BØJNING), pronunciation (UDTALE), and origin (OPRINDELSE). Below this, the section 'Betydninger' (meanings) is shown, with a red oval highlighting a link 'vis mere' (show more). To the right, three boxes from the DDB thesaurus are displayed, each with a red oval around its title. The first box, 'BEVÆGELSE', lists synonyms like *gå*, *trave*, and *spadsere*. The second box, 'LANGSOM BEVÆGELSE', lists synonyms like *bevæge sig langsomt*, *snegle sig*, and *gå langsomt*. The third box, 'FORNØJELSER OG FRITIDSAKTIVITETER', lists synonyms like *gå*, *gå lange ture*, and *trave*. Each box also includes a reference to 'Den Danske Begrebsordbog, kapitel 8'.

Fig. 3: To the left, the verb *slentre* ('stroll around (in a relaxed manner)') in the DDO dictionary, extended with the direct presentation of seven near synonyms from DDB, including a link '*vis mere*' ('show more'). To the right, the link has been activated, showing three boxes with more related words, based on the three occurrences of the sense in the DDB thesaurus: 1) words from the section "*Bevægelse*" ('movement'), 2) words from the section "*Langsom bevægelse*" ('slow movement'), and 3) words from the section "*Fornøjelser og fritidsaktiviteter*" ('leisure').

The extract from each box is calculated dynamically, based on an algorithm telling us how far to the left and right of the search word we should go in DDB if we intend to present the words from the same word class which are most similar in meaning. Keywords on the first level always constitute the borderline, indicating a shift to a new semantic type. Keywords on the second level, including the words within their scope, are sometimes, but not always, included in the extract, depending on the type of search word, which may itself be a keyword on the first or the second level. The boxes include synonyms already presented in the DDO entry.

5 Results and Conclusion

In most cases, the method results in very useful extracts of related words, as for example seen in Figure 3, where we get three boxes of words altogether reflecting very well the different aspects of the verb *slentre*. We have received a lot of positive user feedback since the thesaurus function was released at the beginning of January 2018. But there are also some problems, especially when it comes to the small automatically inserted extract. One challenge was to decide whether to include the headword in the extract or not, and we discussed at length how to define the small extract (how many words to show, how to find the best box to extract them from, etc.). When comparing to the English dictionaries which mostly leave out the small extract when a few synonyms have already

been described manually, it is maybe worth reconsidering whether to leave it out, and link directly to the boxes.

The many cases of collocations in the DDB thesaurus also cause problematic results. As an example, the expression *glad dreng* ('happy young man') in the DDB thesaurus is linked to the adjective *glad* ('happy') in the XML-structure, not to the noun *dreng* ('young man') in the DDO dictionary. This results in a box in the entry of *glad* with words that are in fact near synonyms of 'young man' (next to three boxes presenting very useful related words of *glad*, we should mention). Negated senses in the thesaurus also cause problems. The DDO dictionary sometimes describes the 'positive' sense of a word which rarely occurs without a negated context, then mentioning this as a constructional comment. But in the thesaurus the word is presented negated, since it is hard to understand it without the negated context. This leads of course to the extraction of antonyms instead of synonyms. We solved these problems by presenting the expression (e.g. *glad dreng*) in the box headline, but in some cases we might consider changing the underlying data

The groups in the DDB thesaurus which consist of co-hyponyms or thematically related words do not always give good results when they are presented as related words in the DDO dictionary. This was foreseen to be a problem since we lack information on whether the group consists of co-hyponyms or rather of synonyms. Especially when keywords were introduced in the thesaurus in very large word groups, not because they were good headlines, but simply in order to facilitate the look up-process from the index in the printed book, we have a problem of odd 'headlines' of a group of words. This problem cannot be solved without going through all the cases manually.

Finally, we would like to present the boxes not in the numeric order of sections in the thesaurus, but rather from a calculation of semantic relevance, for example by giving prominence to sections where the search word itself is a keyword at level one, or where it has the highest number of direct near-synonyms in the DDB thesaurus. We plan to improve the presentation order as part of the hopefully future project of presenting the entire DDB thesaurus manuscript, including the hierarchies of meaning, as an online thesaurus at the DSL dictionary site www.ordnet.dk.

References

Online dictionaries

Macmillan English Dictionary Online. Accessed at: <http://www.macmillandictionary.com> [27/03/2018].

Merriam-Webster Accessed at: <http://www.merriam-webster.com> [27/03/2018].

Oxford English Dictionary (OED) Accessed at: <http://www.oed.com> [27/03/2018].

The Historical Thesaurus of English Accessed via: <http://www.oed.com> [27/03/2018].

Dictionary.com Accessed at: <http://www.dictionary.com> [27/03/2018].

Oxford Dictionaries: English. Accessed at: <https://en.oxforddictionaries.com/english> [27/03/2018].

The dictionary DDO: Den Danske Ordbog (the Danish Dictionary) Accessed at: <https://www.ordnet.dk/ddo>.

Dornseiff, Franz (2004). *Der deutsche Wortschatz nach Sachgruppen*, 8. Auflage, Berlin/New York: Walter de Gruyter.

Lyons, John (1977). *Semantics. Volumes 1-2*, Cambridge, University Press.

Nimb, Sanni, Anna Braasch, Sussi Olsen, Bolette Sandford Pedersen, Anders Søggaard (2017). From Thesaurus to Framenet. In: (Eds. Kosem, I., Tiberius C., Jakubíček, M., Kallas, J., Krek, S., Baisa, V.): *Electronic lexicography in the 21st century. Proceedings of eLex 2017 conference, Leiden, the Netherlands, 19–21 September 2017*, s. 1-22.

Nimb, Sanni, Lars Trap-Jensen, Henrik Lorentzen (2014). The Danish Thesaurus: Problems and Perspectives. In: Andrea Abel, Chiara Vettori & Natascia Ralli (eds.). *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19 July 2014. Bolzano/Bozen: EURAC Research, s. 191-199.

- Nimb, Sanni, Henrik Lorentzen, Liisa Theilgaard, Thomas Troelsgård, Lars Trap-Jensen (2014). *Den Danske Begrebsordbog*, Det Danske Sprog- og Litteraturselskab, Copenhagen.
- Nimb, Sanni, Bolette S. Pedersen, Anna Braasch, Nicolai H. Sørensen & Thomas Troelsgård (2013). Enriching a wordnet from a thesaurus. In: *Workshop Proceedings on Lexical Semantic Resources for NLP from the 19th Nordic Conference on Computational Linguistics (NODALIDA)*. Linköping Electronic Conference Proceedings; Volume 85.
- Nimb, Sanni, Bolette Sandford Pedersen (2012). Towards a richer wordnet representation of properties – exploiting semantic and thematic information from thesauri. In: *LREC 2012 Proceedings*. Istanbul, Turkey.
- Pustejovsky, James (1995) *The Generative Lexicon*, MIT Press, Cambridge, Massachusetts.
- Roget, Peter Mark (2002). *Roget's Thesaurus, 150th anniversary edition* edited by George Davidson 2002. London: Penguin.