

A Thesaurus-Based Sentiment Lexicon for Danish: The Danish Sentiment Lexicon

Sanni Nimb¹, Sussi Olsen², Bolette S. Pedersen², Thomas Troelsgård¹

The Society for Danish Language and Literature¹, Centre for Language Technology, NorS, University of Copenhagen²

Christian Brygge 1, 1219 København K¹, Emil Holms Kanal 2, 2300 Copenhagen S²

{sn,tt} @dsl.dk, {saolsen,bspedersen} @hum.ku.dk

Abstract

This paper describes how a newly published Danish sentiment lexicon with a high lexical coverage was compiled by use of lexicographic methods and based on the links between groups of words listed in semantic order in a thesaurus and the corresponding word sense descriptions in a comprehensive monolingual dictionary. The overall idea was to identify negative and positive sections in a thesaurus, extract the words from these sections and combine them with the dictionary information via the links. The annotation task of the dataset included several steps, and was based on the comparison of synonyms and near synonyms within a semantic field. In the cases where one of the words were included in the smaller Danish sentiment lexicon AFINN, its value there was used as inspiration and expanded to the synonyms when appropriate. In order to obtain a more practical lexicon with overall polarity values at lemma level, all the senses of the lemma were afterwards compared, taking into consideration dictionary information such as usage, style and frequency. The final lexicon contains 13,859 Danish polarity lemmas and includes morphological information. It is freely available at <https://github.com/dsldk/danish-sentiment-lexicon> (licence CC-BY-SA 4.0 International).

Keywords: sentiment lexicon, linked data, thesaurus

1. Introduction

The Danish Sentiment Lexicon was published in 2021, and is freely available at <https://github.com/dsldk/danish-sentiment-lexicon> (licence CC-BY-SA 4.0 International). The lexicon contains approx. 14,000 Danish polarity lemmas and includes morphological information. This paper describes how the lexicon was compiled by use of lexicographic methods and based on linked lexical data in line with other formal lexicons compiled in collaboration between The Society for Danish Language and Literature (DSL) and Centre for Language Technology at University of Copenhagen (CST) the last two decades. In this case the linked data constitute The Danish Thesaurus *DDB* (*Den Danske Begrebsordbog*, Nimb et al. 2014) linked at sense level to the comprehensive monolingual Danish dictionary, *Den Danske Ordbog* (DDO) and representing approx. 100,000 lemmas and more than 120,000 word senses. The linked dataset consisted of lemma senses listed in semantic order from the manually identified polarity sections in the thesaurus, combined with information from DDO on style, usage and frequency.

In the same way as word senses have either neutral, positive or negative polarity, the titles of thesaurus sections convey a certain polarity. The idea was to identify the positive and negative thesaurus titles as the starting point for a sentiment lexicon for Danish with a much higher lexical coverage than existing Danish sentiment lexicons, and based not only on the subjective judgment of the word itself, but also on the comparison of its synonyms and near-synonyms in the thesaurus, as well as on the study of its sense description in the dictionary. By comparing synonyms and near-synonyms, they are able to better calibrate the degree of polarity within sets of words denoting very similar concepts.

The dataset was annotated manually by two lexicographers in several rounds. On the way, we obtained a dataset with polarity information at sense level which can be directly transferred back to the dictionary as supplementary user information on the senses. However, our main goal was to publish a lexicon which is practically applicable in automatic text analysis without requiring prior

identification of the word sense. Once again, the dictionary played an important role since the information on other senses of the same lemma was taken strongly into consideration when a generalized sentiment value at lemma level was decided upon. In the final dataset, morphological information from the DDO was added to the lemmas. The work is part of a research project of which the overall goal is to study how the DDB thesaurus can be used to facilitate the task of establishing computational semantic lexicons for Danish. In Pedersen et. al. (2019) we describe how we extend the number of adjective synsets in the Danish WordNet based on the thesaurus. Nimb et al. (2017) presents earlier use of the DDB thesaurus in the compilation of the Danish Framenet Lexicon, and Pedersen et al. (2022) presents the project COR where we compile a new Danish formal lexicon, and where the identification of relevant vocabulary is based on the keyword status in the DDB thesaurus.

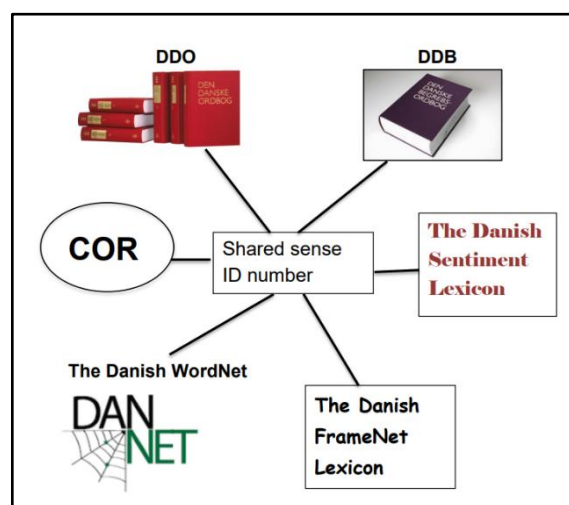


Figure 1: Four computational lexicons linked to a monolingual dictionary DDO as well as to the thesaurus DDB via the sense ID numbers of DDO

It is worth underlining that all these formal lexicons (the WordNet, the FrameNet lexicon, the COR lexicon and the

sentiment lexicon) are linked to the sense inventory of the DDO. So is the thesaurus vocabulary (see figure 1), allowing us today to combine formally as well as informally expressed lexicographic information on Danish lemmas and word senses in many different ways, see also Pedersen et al. (2022).

In the following section we start by introducing sentiment lexicons and describe related work before we in section 3 turn to a detailed description of the lexicographic method behind the sentiment lexicon. In section 4 the lexical content of the lexicon is presented. In section 5 we discuss our method and further work, and conclude.

2. Introduction to sentiment lexicons and related work

Sentiment lexicons contain information on polarity values of words and are used in the automatic analysis of textual data where texts are analyzed and classified as negative, positive, or neutral, e.g. in product reviews, assessment of customer service, monitoring of social media, study of political attitudes (for Danish, see Enevoldsen and Hansen, 2017). The automatic analysis based on such lexicons can also be used to show mood or attitude development in literary works, see Liu (2015). The texts are typically classified based on the presence of unique polarity-bearing words but in some cases also on the assignment of probable polarity to arbitrary words. Hybrid methods that combine the information in sentiment lexicons with statistical methods are also used (Jacobs, 2019).

Sentiment lexicons have been created for many languages by use of different methods, however rarely based on the re-use of existing dictionary information.

Lower-resourced languages like Danish often have only a small handful of sentiment lexicons available, typically compiled over a period of time and with quite varying approaches and coverage. For several years, the most widespread Danish sentiment lexicon has been AFINN (Nielsen, 2018). The glossary was translated from an English sentiment lexicon of the same name (Nielsen, 2011) and includes more than 3000 unique lemmas marked with polarity value (negative / positive) as well as a polarity degree from -5 to +5. A newer sentiment lexicon for Danish, SENTIDA (Lauridsen et al., 2019) is based on a list of the 10,000 most frequent Danish words published by DSL¹. Nouns, adjectives, verbs, adverbs and interjections were marked with AFINN's polarity degree from -5 to +5 (incl. 0 for neutral) by three annotators, a common value was found and also in this case purely neutral words were removed. The glossary was then compiled and supplemented with words from AFINN in case of lack of lexical coverage. Next, the glossary was subjected to automatic 'tuning', where each lemma is reduced to "the part of the lemma which does not change, but which still retains the meaning of the word" (Lauridsen et al., 2019), e.g. *spillende* 'playing', *spiller* ('player') and *spille* ('to

play') is reduced to *spil* ('play'). The result is a list of approx. 5,200 entries, which according to the authors correspond to approx. 35,000 wordforms. The authors have performed various experiments which show that SENTIDA gives better results than AFINN. Because we dispose of comprehensive and well-structured lexical resources for Danish with to some extent already existing information on polarity, as well as easily identifiable polarity data from a thesaurus, we found it well worth the effort to compile a Danish lexicon with an even higher number of entries.

The Swedish sentiment lexicon SenSALDO (Rouces et al., 2018a and b), is based on a thesaurus, *SALDO*, following a similar method. Based on the information that can be deduced from the topological placement of the words and their placement in semantic groups in *SALDO*, the words that were judged to contain polarity were selected. Only single words of at least two letters were included and the list was limited to contain only particularly frequent nouns, verbs, adjectives, and interjections based on frequency measurements in a large text corpus for modern Swedish. The result was a glossary of almost 2,000 words with polarity. The latest version of SenSALDO has been expanded with far more word meanings, and unlike the first version and the AFINN lexicon, it also includes many neutral ones. It contains 12.287 lemmas of which 6,386 convey polarity, being either negative (68%) or positive (32%) while the rest is neutral. Like our lexicon, SenSALDO is published in a version including inflectional information, a total of approx. 85.000 words².

Better resourced languages, like English, tend not surprisingly to have developed a much larger selection of alternatives. Following however in essence the different methodologies described above, such are either manually constructed (including crowdsourcing), bootstrapped from a set of seed words, semi-transferred from other domains, or generated from machine learning based on human sentiment coding. One of the most well-known for English is the hand-crafted SentiWordNet 3.0 (Baccianella et al., 2010) relying on the wordnet framework (Fellbaum (ed), 1998); this lexicon seems however to be in competition with other (some more recent) sentiment lexicons such as the KWWSI Sentiment Lexicon (Khoo & Johnkhan, 2017), MPQA (Multi-perspective Question Answering) Subjectivity Lexicon (Wilson et al., 2005), and National Research Council Canada (NRC) Word-Sentiment Association Lexicon³. Khoo & Johnkhan (2017) perform a comparative evaluation of six sentiment lexicons for English and give recommendations as to which resources are better suited for which tasks.

3. Compiling the lexicon

Like the Swedish sentiment lexicon SenSALDO described above, our approach is ontological, however in our case the DDB thesaurus that we use is linked to a comprehensive monolingual dictionary at sense level. The thesaurus

¹ <https://korpus.dsl.dk/resources/details/freq-lemmas.html>

² see <https://spraakbanken.gu.se/resurser/sensaldo>.

³ <https://nrc.canada.ca/en/research-development/products-services/technical-advisory-services/sentiment-emotion-lexicons>

presents the vocabulary and the senses in the DDO dictionary in semantic order organized in groups of synonyms and near-synonyms. The groups are presented in 888 named sections, divided in 22 named chapters. The section and chapter division is inspired by the divisions in a German thesaurus (Dornseiff 2004, Nimb et al. 2014). The DDB covers approx. 95% of the DDO lemmas and senses, some of which are included in more than one of the thematic thesaurus sections. Since the DDO dictionary is corpus based, the DDB also covers to a very high degree the word senses found in modern Danish texts.

3.1 Selection of lexical data to be annotated

The starting point was the manual annotation of polarity relevant sections in the DDB. 24% of the 888 sections were judged by two lexicographers to contain polarity words based on the section name. 57% were judged to contain negative words (e.g. the sections betitled (translated into English) 'Unimportant' and 'Sadness'), 37% to contain positive words (e.g. the sections 'Important', 'Admire' and 'Friendship'), while 6% were identified as sections from which at least some of the words probably would be relevant to include in a sentiment lexicon (e.g. 'Reputation', and 'Protest, Rebellion'). The next step was to transfer the negative and/or positive values at section level to each word in the section and combine them with the relevant information from DDO.

3	45	Harmonisk	idyllisk	fredfyldt, harmonisk og fri for problemer,	2	harmonisk, lige
3	45	Harmonisk	ligevægtig	med samme betydning el. vægt og derfor l	2	harmonisk, bala
3	45	Harmonisk	musikalsk	iorefaldende; harmoniskiorefaldende	2	vellydende, vell
3	45	Harmonisk	organisk	karakteriseret ved naturlig, harmonisk sammenhæ	2	harmonisk, bala
3	45	Harmonisk	regelmæssig	som er bygget el. formet på en ensartet el.	2	harmonisk, bala
3	45	Harmonisk	regulær	med regelmæssig form (og derfor ordentlig	2	harmonisk, bala
3	45	Harmonisk	rolig	enkel, harmonisk, afdæmpet e.l.harmonisk	2	2 harmonisk, regu
3	45	Harmonisk	rytmisk	som foregår med regelmæssige mellemrumforegå	2	vellydende, vell
3	45	Harmonisk	smagfuld	som vidner om el. bærer præg af god æstetisk	2	harmonisk, ega
3	45	Harmonisk	stabil	som er konstrueret solidt uden store muligheder	2	2 harmonisk, sym
3	45	Harmonisk	stilfuld	med elegant, imponerende, smukt el. harmonisk	2	harmonisk, bala
3	45	Harmonisk	symmetrisk	som udviser el. er kendetegnet ved symmetrisym	2	harmonisk, bala

Figure 2 : Extracted data from the thesaurus. 12 lemma senses from *Chapter 3, section 45 Harmonisk* ('harmonious'), translated into English: 'idyllic', 'balanced', 'musical', 'organic', 'regular', 'calm', 'rhythmic', 'tasteful', 'stable', 'stylish', 'symmetrical'; the corresponding definitions from the DDO, the estimated polarity value '2', then the value '2' from AFINN for two of the lemmas, *rolig* and *stabil*, and finally the entire semantic group in in the thesaurus (only partly visible).

We also added polarity information from AFINN (at lemma level, i.e. inserted for all senses of the lemma in the dataset). Furthermore, the group of synonyms and near synonyms of the word was inserted, identified by calculating the scope of the nearest keyword in the thesaurus as described in Nimb et al. (2018). Fixed expressions were omitted from the dataset. The full dataset consisted of 25,000 word senses, some of which were represented more than once (being extracted from different sections in the thesaurus), see figure 2.

3.2 Polarity annotation of the lexical data

The 25,000 senses were annotated by two highly trained lexicographers, half each (however, many senses were double represented in the dataset and thereby annotated 2828

more than once in different contexts). To justify the method, we measured beforehand the inter-annotator agreement based on the double-annotation of 400 words also present in AFINN. We annotated only negative and positive values, not the degree of polarity. The inter-annotator agreement based on Cohen's Kappa was 0.83. The cases of disagreement were discussed, and a set of guidelines was established. However, some cases of sense polarity depend on more personal beliefs and political observance. For others, the polarity depends on exterior factors, typically expressed by the object of the verb (verbs like 'prevent' and 'interrupt'). Other examples on complicated cases were adjectives which in a basic sense are negative but in a derived (often adverbial and informal use) are positive (*skør* (crazy), *vild* ('wild'), *fed* ('fat')). We also observed a few disagreements with AFINN, for example *lyve* ('to tell a lie') and *streng* ('severe') which we consider to be negative, and *undskyldte* ('to apologize') which we consider to be positive.

The annotation task consisted of several rounds (see table 1 for an overview). In the first round of the annotation process, the automatically transferred section values were simply to be confirmed, when not replaced with another value by the two lexicographers.

In the second round, values representing the degree of negativity or positivity (a scale from -3 to +3) were added. All lemma senses with AFINN values higher than 1 were used as the starting fix point in the dataset. Both the lemma sense and the surrounding (polarity conveying) near synonyms and synonyms, typically not being represented in AFINN, were annotated, always inspired by the corresponding degree in AFINN. However, we merged the three highest values (3,4,5) in AFINN. AFINN only contains a few of these, many of which are interjections and swearwords. At a later stage, information on negative interjections and swearwords can be transferred directly into the sentiment lexicon based on information in the DDO, so can higher degrees of negativity on some of the most negative lemmas. Instead, we put an effort into comparing the semantically related words to one another with respect to polarity degree: Which of the grouped synonyms and near synonyms was the most negative one, which ones were less negative, or maybe neutral? Also morphologically derived lemmas in the dataset were compared with AFINN and the AFINN value was expanded to these, i.e., if the noun *apati* ('apathy') has the negative value -2 in AFINN, the derived adjective *apatisk* ('apathetic'), which is not included in AFINN, also got the value -2.

In the third round, the annotations of double represented senses (many senses are represented in more than one thesaurus section) were checked and harmonized, including the degree values, which in some cases varied for the same lemma sense across the dataset since they depended on the surrounding synonyms and near synonyms in the specific thesaurus section. 2,250 lemmas of which none of their senses conveyed polarity of any kind, were in this round sorted out (as in the AFINN and SENTIDA lexicons, but

	Annotation task	The judgement is based on
Step 1	sense polarity: 0, positive or negative?	AFINN Comparison of synonymy and near synonymy in the DDB thesaurus Information in DDO
Step 2	degree of sense polarity: -3, -2, -1, +1, +2, +3?	AFINN Comparison of synonymy and near synonymy in the DDB thesaurus. Information in DDO
Step 3	harmonizing values across the dataset	Lemma / sense represented more than once in the dataset? (Due to multiple representations in thesaurus)
Step 4	deciding upon polarity at lemma level	Information in DDO. Conflicting polarities of lemma senses? Rare sense to be ignored? Or lemma to be left out?
Step 5	validation	1/3 of annotated data + comparison of all lemmas with same high degree (e.g. all +3 lemmas)

Table 1: The five steps in the polarity annotation task

opposite to SenSALDO). A last check of the total lexical coverage was carried out. First, 168 lemmas having at least one sense labeled 'derogatory' in DDO, but which however were not included in any of the polarity conveying thesaurus sections and thereby not in our dataset, were added and annotated with respect to degree of negativity. Secondly, we checked that all lemmas represented in AFINN were included (cf. SENTIDA's method). This led to the extraction and annotation of lemma senses from 3 more sections in DDB, for example from the section *Ophør* ('End, to cease'). The result was a dataset consisting of 14,271 lemmas with at least one sense conveying either negative or positive connotation, in total 17,883 annotated senses. This information is ready to be transferred to the DDO sense descriptions, where it can serve as useful user information in the online dictionary <https://ordnet.dk/ddo>.

The final goal, however, is a sentiment lexicon that indicates only one unique polarity value for each lemma and thereby is more practically applicable in automatic text analysis since it's use does not require prior identification of the word sense according to the DDO sense inventory. In the fourth and last annotation round, all polysemous lemmas of which the sense polarity values diverged or directly conflicted were therefore reviewed one again. The goal was to either remove polarity ambiguous lemmas completely from the dataset, or resolve the ambiguity (i.e., neglect one of the senses due to low frequency). This was done in two steps:

- The 0,1% of the lemmas in the sense lexicon that had both a positive and negative polarity value were studied. Half of these were rejected due to the ambiguity (e.g. *frelst* 'saved', *sej* ('tough'), *skarp* ('sharp'), *overlegen* ('superior') og *glat* ('smooth'). The other half was kept in the lexicon, since we estimated that the polarity sense was by far the most frequent sense of the lemma. For example, the negative sense of *vigtig* ('important') "who has excessively high thoughts about himself", had no influence on the overall lemma polarity level because it is marked as rare in DDO. In this way *vigtig* could be preserved in the lexicon with its positive polarity, otherwise we would have had to delete the lemma from the lexicon.
- Lemmas that, in addition to their polarity sense, also had a frequent neutral meaning, were discarded from the dataset. E.g., the noun *gås* 'goose' was not included, despite the negative figurative sense 'stupid woman', cf. SENTIDA's approach.

In a last and final round, approx. 1/3 of the annotated data was validated by an experienced lexicographer. Lemmas with identical high polarity degrees, for example all -3 lemmas, were also compared and validated a last time.

The result is a sentiment lexicon with 13,859 lemmas having a negative or positive polarity value with varying degrees from -3 to +3. Inflectional forms of all lemmas were added to the dataset before it was released.

4. The distribution of polarity values in the lexicon

Of the 13,859 lemmas, 8575 (62%) have negative polarity and 5284 (38%) positive polarity, which is roughly the same distribution as in SenSALDO (see section 2), i.e. a clear predominance of negative lemmas. This phenomenon seems to apply generally to sentiment lexicons according to e.g. Devitt & Ahmad (2013) who also compare it with the distribution in texts where the relationship is thought to be reversed. Interestingly, there is also a clear predominance of negative DDB thesaurus sections and more or less the same distribution between negative and positive sections in it (see section 3.1 above).

Of the 13,859 lemmas, 4% are annotated to be extremely positive (+3), e.g. *glædesudbrud* ('outburst of joy') and *gennemsolid* ('completely solid'), while almost the double, 7%, are annotated as extremely negative (-3), e.g. *lortevejr* ('shit weather'), *ækelhed* ('disgust') and *løgnhals* ('liar'). 12% of the lemmas are annotated to be very positive (+2), e.g. *suveræn* ('superb'), *strålende* ('brilliant') and *supergod* ('super good'), while more than double, 27%, are annotated to be very negative (-2), e.g. *fattigdom* ('poverty') and *ukultiveret* ('uncultivated'). See figure 3. The share of negative lemmas rises as the polarity degree rises, see figure 4.

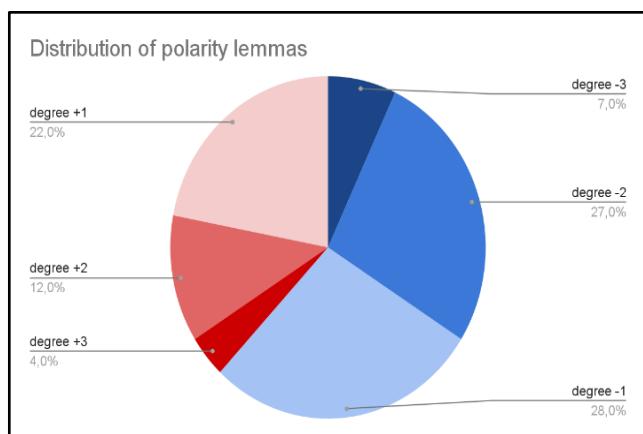


Figure 3: Share of positive (red, to the left) and negative (blue, to the right) lemmas and their degree subgroups

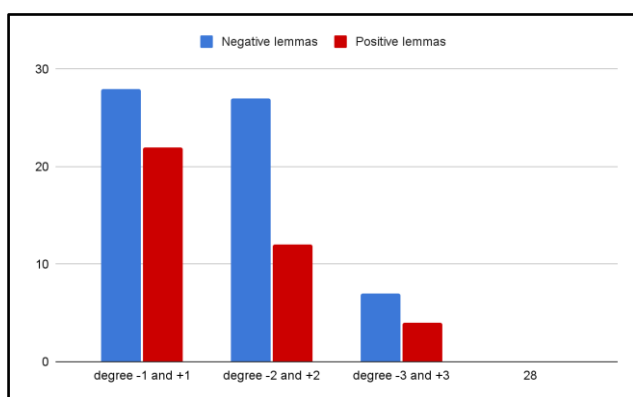


Figure 4: No matter the degree of polarity, we always find most negative lemmas, but the higher the degree, the more negative lemmas. The vocabulary contains around double the amount of ‘very negative’ lemmas (-2) compared to ‘very positive’ ones (+2). This is also the case when we compare the number of ‘extremely negative’ lemmas (-3) with the number of ‘extremely positive’ ones (+3).

5. Discussion of method

A challenging part of the task of establishing a sentiment lexicon is to find an objective way to include scalable values in the polarity marking as a supplementary information to the pure negative and positive values. A most challenging task is to perform a calibration of polarity values between word classes and between different semantic fields. It is hard for an annotator to be consistent, i.e., is *good* more positive than *favor*, or *importance* more positive than *idealism*? And what is most negative: *resentful*, *heartbreaking*, or *crooked*? In the preparation of the first version of the Swedish sentiment lexicon SenSALDO, a method called Best-Worst Scaling was used. Out of four random words with polarity, not only one but four annotators had to select the two words that were most value-laden, i.e., most at each end of a positive-negative scale, before determining the final value in the lexicon.

As described in section 3.2. above, we have instead compared lemmas in our dataset from the DDB thesaurus section with the degree of polarity of the same word in

AFINN. Only a few words in each DDB section had an enhanced value since AFINN has a significantly lower lexical coverage (3000 words), but since we have maintained the semantic order from the thesaurus, we were able to either expand the enhanced value in AFINN to all synonyms and near synonyms or modify it when necessary. We find it highly relevant to compare words denoting the same concept with respect to the nuances of polarity since the language user’s conscious choice of e.g. a highly negative lemma in the many cases where there are in fact neutral or less negative alternatives, clearly must reflect a highly negative attitude. The calibration across semantic fields was carried out by comparing afterwards the list of e.g., all ‘extremely positive’ lemmas, where a few polarity degrees were balanced when compared to other lemmas of the same degree. In a future task, it might be worth considering whether crowdsourcing could be used to validate the values.

The newest version of SenSALDO includes neutral lemmas in the dataset, and we might consider doing so as well. We have annotated some of the data already, and the linked data method that we use would also allow us to identify a very large part of the neutral vocabulary in DDO, in this case based on the initial annotation of neutral thesaurus section titles.

6. Conclusions

As expected, the lexical coverage of the lexicon is high since the data is based on comprehensive monolingual lexical resources. But for the very same reason, it does not cover ad hoc composita of which there are many in Danish texts. We manually annotated a literary piece of text containing many polarity conveying lemmas (Pedersen et al. 2021), and the (few) words that were not covered by the lexicon, were all ad hoc composita. To be able to cover such cases, the sentiment lexicon should be extended with semantic morphemes with a high degree of polarity, and a module for automatic splitting of unknown compositions could be added. We might also consider adding neutral lemmas in the next version.

We are also aware that even though we are sure to have identified most of the negative and positive lemmas in Danish, there might still be undiscovered ones in the thesaurus, e.g. some negative synonyms of the derogatory DDO words found only in neutral thesaurus sections, e.g., the sections *Kvinde* (‘Woman’) and *Mand* (‘Man’) which were not selected for annotation in the first place. When we find such lemmas, we will consider including them in sentiment lexicon, and if relevant also in other thesaurus sections.

A feature of the lexicon which differentiates it from other sentiment resources for Danish is that the links to the DDO lemma and sense ids guarantees the linking to a large number of other Danish lexical resources, and also to the Linguistic Linked Open Data via the Danish wordnet, DanNet. This allows for experiments where the polarity

values are combined with syntactic and semantic properties.

We plan to transfer the polarity information at sense level (which is not part of the first release) not only into the DDO manuscript where it will be useful in the editorial work and might open for new presentations of the dictionary content, e.g. of synonyms, but also to DanNet, which already contains this information for a small part of the vocabulary. In this way, connotative knowledge about the synset (or rather the synset member) can be utilized in automatic analyzes based on the WordNet. The sentiment lexicon will also be integrated in the COR resource (Pedersen et al. 2022), but in the first place as a separate module linked to the COR register at lemma level since the COR-S sense inventory to some degree diverges from the DDO sense inventory.

The distribution of negative and positive polarity values in the sentiment lexicon gives us new insights into the Danish vocabulary, and it is very interesting that it is in line with the distribution in the Swedish SenSALDO lexicon. Since there is a much higher number of negative lemmas than positive ones in the vocabulary of comprehensive lexicons, sentiment analysis methods based on these are likely to identify an overweight of negative polarity lemmas in the text compared to methods using lexicons with a lower coverage. In the latter case, many of the polarity bearing lemmas in the text are probably not identified.

7. Acknowledgements

The work is funded by the Carlsberg Foundation 2019-2022

8. Bibliographical References

- Devitt, A. & Ahmad, K. (2013). Is there a language of sentiment? An analysis of lexical resources for sentiment analysis. In: *Language Resources & Evaluation* 47: 475–511. doi 10.1007/s10579-013-9223-6.
- Baccianella, S., Esuli, A., Sebastiani, F. (2010). SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. In: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. European Language Resources Association (ELRA), Valletta, Malta. http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf
- Dornseiff, F. (2004). *Der deutsche Wortschatz nach Sachgruppen*. W. De Gruyter, Berlin; New York
- Enevoldsen, K. C., & Hansen L. (2017). Analysing Political Biases in Danish Newspapers Using Sentiment Analysis. In: *Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift*, 2(2), pp.87–98.
- Jacobs, A. (2019). Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics. In: *Frontiers in Robotics and AI*. 6. 53. 10.3389/frobt.2019.00053.
- Lauridsen, G., Dalsgaard, J., & Svendsen, L. (2019). SENTIDA: A New Tool for Sentiment Analysis in Danish. In: *Journal of Language Works - Sprogvidenskabeligt Studentertidsskrift*, 4(1), 38-53. <https://tidsskrift.dk/lwo/article>
- Liu, B. (2015). *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139084789
- Nielsen, F. Å. (2018). *Danish resources*. <https://bit.ly/2NDHcbW>
- Nielsen, F. A. (2011). A New ANEW: Evaluation of a Word List for Sentiment Analysis in Microblogs. In: *Making Sense of Microposts (MSM2011)*, 93-98. http://ceur-ws.org/Vol-718/paper_16.pdf.
- Sanni Nimb, Sørensen, N.H., Troelsgård, T. (2018). From Standalone Thesaurus to Integrated Related Words in The Danish Dictionary. In: Čibej, J., Gorjanc, V., Kosem, I., Krek, S. (eds.). *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana University Press, Faculty of Arts, Slovenia.
- Nimb, S., Braasch, A., Olsen, S., Pedersen, B. S. & Søgaard, A. (2017). From Thesaurus to FrameNet. In: *Proceedings of eLex 2017*. Electronic Lexicography in the 21st Century - Proceedings of eLex 2017 Conference. s. 1-22. Leiden, The Netherlands.
- Nimb, S., Lorentzen H. & Trap-Jensen, L. (2014). The Danish Thesaurus: Problems and Perspectives In: Abel A., Vettori C. & Ralli N. (eds.). *Proceedings of the XVI EURALEX International Congress: The User in Focus*. 15-19 July 2014. Bolzano/Bozen 2014: EURAC Research, pp. 191-199.
- Pedersen, B. S., Sørensen, N. C. H., Nimb, S., Olsen, S., Flörke, I. & Troelsgård, T. (2022). Compiling a Suitable Level of Sense Granularity in a Lexicon for AI Purposes. The Open Source COR Lexicon. (Submitted for review). *LREC 2022*
- Pedersen, B. S., Nimb, S. & Olsen, S. (2021). Dansk betydninginventar i et datalingvistisk perspektiv. In: *Danske Studier*. 2021, pp. 72-106.
- Pedersen, B. S., Nimb, S., Olsen, I. R. & Olsen, S., 2019, Linking DanNet with Princeton WordNet. In: *Global WordNet 2019 Proceedings*, Wroclaw, Poland. Wroclaw: Oficyna Wydawnicza Politechniki Wroclawskiej, 10 pp.
- Pedersen, B. S., Nimb S., J. Asmussen, Sørensen, N. H., Trap-Jensen, L. & Lorentzen, H. (2009). DanNet – the challenge of compiling a WordNet for Danish by reusing a monolingual dictionary. *Language Resources and Evaluation, Computational Linguistics Series*, pp.269-299.
- Rouces J., Borin, L., Tahmasebi, N., Eide, S. R. (2018a). Defining a gold standard for a Swedish sentiment lexicon: Towards higher-yield text mining in the digital humanities. In: *CEUR Workshop Proceedings vol. 2084. Proceedings of the Digital Humanities in the Nordic Countries 3rd Conference Helsinki*, Finland, March 7-9, 2018. Edited by Eetu Mäkelä Mikko Tolonen Jouni Tuominen.
- Rouces J., Borin, L., Tahmasebi, N., Eide, S. R. (2018b). SenSALDO: Creating a sentiment lexicon for Swedish. In: *Proceedings of LREC 2018, Eleventh International Conference on Language Resources and Evaluation*, si. 4192-4198, Miyazaki, ELRA.

9. Language Resource References

- The Danish Sentiment Lexicon* : github.com/dsldk/danish-sentiment-lexicon, Det Danske Sprog- og Litteraturselskab & CST, University of Copenhagen, Copenhagen, Denmark
- DanNet*: Pedersen B. S., Nimb, S., Asmussen, J., Sørensen, N. H., Trap-Jensen, L. & Lorentzen, H. (2009). DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. In *Language Resources and Evaluation*, 43, pp. 269-299. Available at : wordnet.dk; <https://andreord.nors.ku.dk>.
- Den Danske Ordbog (DDO)* (2008-): online dictionary, ordnet.dk/ddo, Det Danske Sprog- og Litteraturselskab, Copenhagen, Denmark. <https://ordnet.dk/ddo>.
- DDB*, The Danish Thesaurus: Nimb, S., Lorentzen, H., Theilgaard, L., Troelsgård, T. (2014). *Den Danske Begrebsordbog*, Det Danske Sprog- og Litteraturselskab, Copenhagen, Denmark.
- SALDO*: Lars Borin, Markus Forsberg, Lennart Lönngrén (2013): SALDO: a touch of yin to WordNet's yang, in *Language resources and evaluation*, volume 47, issue 4, pp. 1191-1211. <https://spraakbanken.gu.se/en/resources/saldo>.