

The SemDaX corpus – sense annotations with scalable sense inventories

Bolette S. Pedersen¹, Anna Braasch¹, Anders Johannsen¹, Héctor Martínez Alonso¹, Sanni Nimb², Sussi Olsen¹, Anders Søgaard¹, Nicolai Hartvig Sørensen²

¹University of Copenhagen & ²The Society for Danish Language and Literature

E-mail: bspedersen@hum.ku.dk, braasch@hum.ku.dk, ajohannsen@hum.ku.dk, alonso@hum.ku.dk, sn@dsl.dk, saolsen@hum.ku.dk nhs@dsl.dk

Abstract

We launch the SemDaX corpus which is a recently completed Danish human-annotated corpus available through a CLARIN academic license. The corpus includes approx. 90,000 words, comprises six textual domains, and is annotated with sense inventories of *different granularity*. The aim of the developed corpus is twofold: i) to assess the reliability of the different sense annotation schemes for Danish measured by qualitative analyses and annotation agreement scores, and ii) to serve as training and test data for machine learning algorithms with the practical purpose of developing sense taggers for Danish. To these aims, we take a new approach to human-annotated corpus resources by double annotating a much larger part of the corpus than what is normally seen: for the all-words task we double annotated 60% of the material and for the lexical sample task 100%. We include in the corpus not only the adjudicated files, but also the *diverging annotations*. In other words, we consider not all disagreement to be noise, but rather to contain valuable linguistic information that can help us improve our annotation schemes and our learning algorithms.

1. Annotating with different sense granularity

With the general aim of increasing the level of semantic corpus resources for Danish that can be used for training of semantic systems, the SemDaX corpus has been developed as a collaboration between Centre for language Technology, University of Copenhagen, and The Society for Danish Language and Literature. In the project “Semantic Processing across Domains” 2013-2016 granted by the Danish Research Council we build on previous experiences regarding combined lexical-semantic resources, and we experiment with annotation at different levels of sense granularity.

Starting in the late 90-ties with the English SemCor corpus based on the Princeton WordNet sense inventory (Landes et al. 1998), sense annotated corpora have been developed for a series of languages. Most of the corpora are annotated with fine-grained sense inventories based on existing wordnets (cf. Petrolito & Bond (2012) for an overview of these), and the resources can roughly be divided into those where all content words are annotated (so-called *all-words* corpora) and those where only selected content words are annotated (so-called *lexical sample* corpora).

As discussed in Ide & Wilks (2007), Kilgarriff (2007) and others, defining appropriate sense inventories for annotation and word sense disambiguation tasks is however a very hard task. The need for coarser and more manageable sense inventories has emerged, partly driven by poor sense annotator agreement scores in the aforementioned annotations. This has resulted in a series of annotation experiments applying manually and automatically clustered senses, as seen in Agirre & Lacalle (2003), Palmer et al. (2007), Passonneau et al.

(2012) Brown et al. (2010), de Melo et al. (2012), and others.

The need for “light weight” semantic annotations has led researchers to focus also on *very* coarse word sense annotation applying so-called supersenses that are derived from the list of WordNet’s *first beginners* or *lexicographical* files. This approach is becoming a de facto standard in recent years (Ciaramita & Johnson 2003, Qiu et al. 2011, Schneider et al. 2012).

In the SemDaX corpus we include both supersense annotations and lexical sample annotations with fine-grained and automatically clustered senses for a selected set of highly ambiguous nouns. All annotations in the corpus rely on the combined wordnet and dictionary resources: DanNet (cf. Pedersen 2009 et al.) and a comprehensive monolingual, corpus-based dictionary of modern Danish, Den Danske Ordbog (DDO, Hjorth et al. 2005), which share sense identifies. The aim of the corpus is twofold: i) to assess the reliability of the different sense annotation schemes in terms of different levels of granularity measured by qualitative analyses and annotation agreement scores, and ii) to serve as training and test data for machine learning algorithms with the practical purpose of developing sense taggers for Danish.

To these aims, we take a new approach to human-annotated corpus resources by double annotating a much larger part of the corpus than what is normally seen¹, and by including in the corpus not only the adjudicated files, but also the *diverging annotations*. In other words, we do not consider all disagreement to be

¹ A well-established rule of thumb is to provide double annotations for at least 2% of the corpus in order to assess the annotation scheme and train the annotators.

noise, but rather to contain valuable linguistic information that we can use to improve our annotation schemes as well as in our learning algorithms. In the following, we describe how the corpus is composed of six textual domains in order to ensure a broad sense representation and how it has been annotated (Section 2) using the WebAnno annotation tool. We continue with an account of each annotation task and scheme: Section 3 describes the all-words task with the coarse-grained supersense inventory, whereas Section 4 describes the lexical sample task applying a finegrained and a clustered sense inventory. Finally, in Section 5 we conclude.

2. The corpus: numbers and tool

The SemDaX corpus is a sense-annotated extract of the 45 million words CLARIN Reference Corpus (Asmussen 2012).

Textual Domain	Description
Blog	Blog written by a woman in her forties
Chat	Chat forum mostly for young people
Magazine	Celebrity gossip magazine
Parliament	Speeches from the Danish Parliament written down by professionals
Forum	Chat forum for manga fans
Newswire	Danish newspaper

Table 1: The textual domains included in SemDaX.

SemDaX contains six domains (Table 1) and comprises two subcorpora SemDaX-Coarse (all words) and SemDaX-LexicalSample, both covering all the textual domains above. SemDaX-Coarse comprises 86,786 running words of which 34,421 words (all nouns, verbs and adjectives) have been annotated with supersenses. 60 % of these have been doubly annotated and adjudicated.

In SemDaX-LexicalSample the number of annotated sentences for each selected noun varies according to the number of senses of the noun ($100 + 15 \cdot \text{no. of senses}$), thus spanning from 177 to 535 sentences per noun. All have been doubly annotated.

The manual annotations have been carried out in WebAnno (Yimam et al., 2013), a web-based annotation tool designed for different linguistic annotations, which facilitates monitoring of the annotation progress, measuring the quality and curation of the annotated files. The annotators were a group of three advanced students assisted by the researchers in the project.

3. Annotating with supersenses

In this section we describe the task of annotating all content words of the corpus with supersenses (*SemDaX-Coarse*). Previous to the annotation, all DanNet synsets were automatically mapped onto the supersense inventory, i.e. the 44 wordnet lexicographical

classes², using a transfer scheme which projects the top-ontological types in DanNet (in terms of Vossen et al 1999) to supersenses. The set of supersenses was further specified in a few cases where a substantial part of the Danish synsets under a given supersense represented its own top-ontological type; this was the case for buildings, vehicles, containers, diseases and a few others. For an overview of the extension of the supersense inventory, we refer to Martinez Alonso et al. (2016).

Further a small set of satellite tags were established to enable annotation of lexical units that consist of more than one word. These comprise phrasal verbs (PART), reflexive verbs (REFL), and verbal collocations (COLL), and ensure that all particles, pronouns or other elements of a collocation are understood as a lexical unit in conjunction with its preceding verb, as illustrated in Figure 1.

The big vodka drinks seemed to go directly into the blood

Figure 1: Phrasal verbs with more than one particle (*se ud til* ('seem')) are annotated as collocations with the sense label (here: verb.cognition) on the lexical kernel (*se*).

Previous to the manual annotation, all corpus data were automatically pre-annotated (based again on DanNet) in order to speed up and ensure more consistent annotation. In the ambiguous cases, the annotator chose between the pre-annotated ones, however, the annotator was free to overrule a pre-annotated sense and assign an alternative more adequate sense. Unknown words (mostly compounds) were obviously not pre-annotated; in these cases the annotators selected the most appropriate sense from a pick list.

Guidelines describing and exemplifying the supersenses were established and these also included annotation practices for idiomatic expressions and phrasal and reflexive verbs. As mentioned, 60% of the all-words files has been annotated by more than one annotator, and have received acceptable agreement results with 0.62 (Krippendorff's α) as average³. All doubly annotated data have been adjudicated, establishing hereby a gold standard. Not all domains scored equally high in agreement. In fact, newswire and parliament speeches proved to have a lower agreement than blogs and chat, probably because the language of these text types is intrinsically more complex and contains more abstract concepts (for a detailed study on domain differences in the annotations see Olsen et al. 2015). Figure 2 shows how each noun supersense is represented disagreement-wise in the corpus.

The rows in the disagreement plot are sorted after the size of the diagonal value. Rows with many large, spread boxes indicate supersenses with low agreement which need a closer examination or more precise guidelines. For

² <http://wordnet.princeton.edu/wordnet/man/lexnames.5WN.html>

³ Cf. Artstein & Poesio 2008 for discussion of agreement scores in computational linguistics.

instance, the supersenses *n.person* and *n.institution* seem to be hard for the annotators to distinguish from each other, whereas *n.disease* has proven easy to identify.

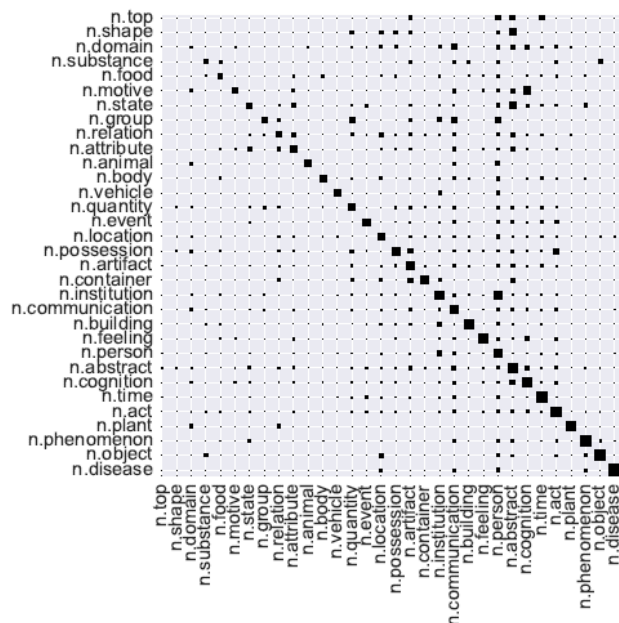


Figure 2: Dispersion plot for noun supersenses

SemDax-Coarse has been used for training and testing of a sense tagger, both in randomized cross-validation in a cross-domain setup. The sense tagger has been implemented following the state-of-the-art model of Johannsen et al (2014) and adapted to Danish. The tagger uses a structured perceptron on an extensive set of features such as word embeddings, wordnet structure and morphology.⁴ The tagger achieves an overall F1 score (including the out-of-span class "O") of 0.82 on heldout data; considering only the F1 of supersense labeling, our micro-averaged score is ~0.65. For more details on the tagger, cf. Martínez Alonso et al. (2015b, 2015c).

4. Fine-grained and clustered annotation of a set of highly ambiguous nouns

The *SemDax-LexicalSample* subcorpus contains sense annotation of 20 highly ambiguous nouns (11 senses on average) which at the same time had a high frequency in Danish corpora. The senses are derived from DDO and DanNet, combining the main and subsenses from the dictionary and the top-ontological types from the wordnet. The main idea is to automatically cluster subsenses (in DDO) that are of the same ontological type (in DanNet) and to compare the agreement scores for these annotations with the agreement scores of the full, fine-grained sense annotations.

DDO was compiled by a staff of 15 editors on the basis of single annotated corpus samples (typically 100-200 sentences, in cases of very polysemous words 1.000) and

⁴ https://github.com/coastalcp/dsl_semtagger

a set of precise guidelines defining the sense structuring principles. In rough terms, these principles were based on the distinctions between core and subsenses as defined by Cruse (2000:110ff). Among several types of regular relations between senses, Cruse defines four types where the senses might be of the same ontological type (and therefore candidates of clustering):

- i) autohyponymy (broad main sense and narrow subsense of the same hypernym);
- ii) autosuperordination (opposite case),
- iii) automeronymy (the main sense constitutes a whole, the subsense a part of it), and
- iv) autoholonymy (the opposite case)

In some cases, however, DDO deviates from Cruse's sense structuring principles. Apart from the general uncertainty about how far a meaning can be extended before it is appropriate to describe it as a new, main sense, the structuring principles were typically neglected when the editors wanted to avoid deep, user-unfriendly substructures and therefore established a new main sense instead.

Irrespective of these exceptions to the general principles, the clustering method results in a reduction of senses of 23.5 % on average. For the annotation purpose, all files are annotated with both the full and the reduced sense inventory by more than one annotator. In 68% of the cases the annotator agreement (using Krippendorff's α) was improved for the clustered senses⁵. The average agreement score for full senses is 0.52, compared to an average agreement on 0.56 for clustered senses. However, each ambiguous noun tells its own very individual story, some being fairly easy to annotate with or without clusters and others being more or less impossible, agreement scores spanning from 0.048 for *plade* (plate, sheet, disc, etc.) to 0.84 for *kurs* (course, exchange rate, price, track, etc.). Likewise, Figure 3 shows how each noun differs regarding to which extent sense reduction is helpful for the annotation.

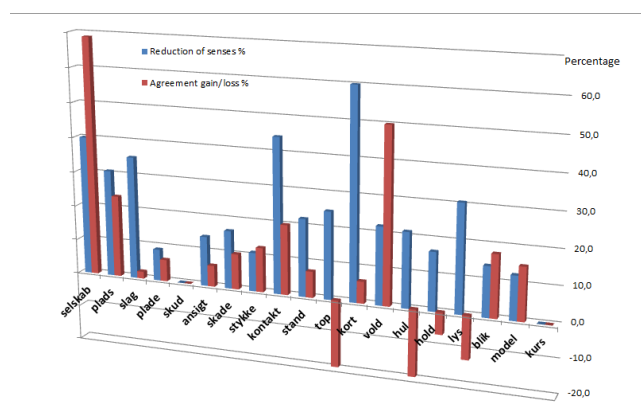


Figure 3: The sense reductions for each noun as well as the gain/loss in inter-annotator agreement when annotating with full sense tag sets compared to clustered sense tags.

⁵ Note that Krippendorff's α is a chance corrected measure.

A closer look into the data sheds some light on the particular lexicographical problems encountered by the annotators. When investigating the cases of high and low inter-annotator agreement scores, respectively, we observe the following tendencies: Agreement is highly dependent on the sense structure of the word as captured in DDO, even in case of annotations using the clustered tag-sets. Not surprisingly, agreement scores are high in cases of relative simple sense structures with clear distinctions between main senses and subsenses, and with only few subsenses, as in the case of *model* (model, form, pattern, person, example..) with six main senses, two subsenses and one multiword expression (MWE) listed. The agreement is lower in cases of complex sense structures with less transparent sense distinctions and with a larger number of subsenses and figurative senses, e.g. *plads* (place, space, square, envelope, rank, job ...): seven main senses, six sub- and figurative senses, and 12 multiword expressions.

The degree of agreement within annotations of a single word varies in the following ways. High agreement is observed in case of concrete objects or easily observable phenomena denoted as in *skade* (magpie (bird); skate (fish)), and other, clearly distinct core senses, e.g. *hold* (team (a group of persons) vs. pain (myalgia)).

The annotation of such multiword expressions that are present in the provided MWE-list of the word achieves the highest agreement because the MWEs are not individually interpreted by the annotators, as they just have to mark them up according to the list. However, occasional syntactic and lexical co-incidence between a free construction and a fixed expression makes up a difficulty for an appropriate identification of the lexical vs. phraseological sense.

Low agreement is generally observed in annotation of subsenses with a high degree of semantic similarity, especially when the disambiguating context is very short or ambiguous, also figurative senses seem to be difficult to agree on. Examples (1) and (2) of inter-annotator disagreement illustrate the most frequent divergences in sense identification:

- (1) *Gennem hårdt arbejde og omfattende **kontakter** fik Udenrigsministeriet lokaliseret danskeren..*
(Through hard work and extensive **contacts** the Ministry for Foreign Affairs localized the Dane..)
- (2) *I dette **lys** er det ikke betryggende at opleve repræsentanter [...] tage afstand fra menneskerettighederne...*
(In this **light**, it is not adequate to experience representatives [...] being discountenancing the human rights...)

In (1) three different sense tags are used covering (a) relationship, link between two parties; (b) party, connection; (c) contact; liaison. These three senses are partly overlapping, thus very difficult to differentiate.

In (2) we see an example of partial disagreement which appears frequently in case of nouns with several figurative senses. In this case the annotators agree on the figurative feature, but interpret the meaning differently. The three sense tags used cover (a) attention (Fig.); (b) knowledge, insight, 'light of recognition' (Fig.); (c) MWE, as a variant of the fixed expression 'see sth. in a new/an other light'. Abstract and figurative senses are very often closely related to each other and thus difficult to distinguish from each other, and even large dictionaries, such as DDO, show inconsistencies in their sense distinctions and linguistic labeling.

Further, annotators seem to make mistakes in cases when a free construction and a fixed expression occasionally coincide as seen in example (3).

- (3) *Den 1. maj så Szhirleys første barn **lyset**, og fødslen forløb [...] fint*
(First of May **saw** the first child of Szhirley **the light**, and the birth proceeded well)

In (3) Two sense tags were used covering (a) light (physical phenomenon); (b) MWE: 'see the light'. In this particular case, an additional problem arises in the sense that the MWE-list does cover the expression, however, only with a very different sense, viz. 'to get unexpected insight in sth. (with a touch of irony)', and not with the sense 'to be born'. Presumably, this is the reason for the annotation with the sense tag (a). Further, the 'born' expression usually is *se dagens lys* ('see the light of the day').

5. Conclusions and future work

We have presented the SemDaX corpus which contains approx. 90,000 words and includes sense annotations with sense inventories of varying granularity. The corpus comprises 6 different domains and includes files with all-words annotations (supersenses on all content words) as well as more fine-grained annotations of 20 highly ambiguous nouns.

We have taken an approach where we doubly annotate a large amount of the corpus and consider disagreement as valuable information. To this end, some preliminary conclusions regarding the soundness of the annotation schemes can be made⁶. The annotations show that the coarse-grained supersense scheme is quite manageable to the annotators resulting in an acceptable agreement of 0.62 applying Krippendorff's α . However, as shown in the dispersion plot in Figure 1 the scheme leaves room for improvements and adjustments; i.e. some particular supersenses prove very hard to agree upon. Further, the considerable information loss in the coarse annotations should be addressed in future extrinsic evaluations; for instance, it can be questioned to which extent we actually capture the practically relevant ambiguities with this coarse scheme; see also Martínez Alonso et al. 2015 for a first attempt of inducing a supersense tagger from our

⁶ Note that the supersense scheme is not directly comparable to the fine-grained schemes since the annotation tasks differ (all-word vs. lexical sample).

supersense annotations.

This leads us to the finer-grained dictionary-driven annotations of highly ambiguous nouns that we described in Section 4. Here we can conclude that a clustered annotation scheme based on an ontologically driven collapsing of subsenses performs substantially better than a fully fine-grained scheme (disregarding here the better chance of agreeing on few tags than on many). However, it is remarkable how each individual noun exposes its own pattern, and how some very ambiguous nouns prove almost impossible to annotate – with or without clusters.

This goes especially for nouns with many subsenses, and in particular when a large part of them are abstract senses. A closer look at the organisation of the lexical source (DDO-DanNet) therefore seems indispensable. In corpus-based dictionary making, the problem of deciding when to describe a set of usages as a separate sense is often resolved on an ad-hoc basis, resulting in many cases of overlapping senses where instances fall under more than one sense category at the same time (Pustejovsky & Rumshisky 2010). This goes for abstract and figurative senses in particular, as we see in the described cases above. Even though DDO is based on corpus annotations and linguistic guidelines, inconsistencies occur due to the many ad hoc decisions needed to be taken in the working process. Furthermore, the sense inventory of DDO has been established with the focus on the human dictionary user, and not on reflecting formal distinctions or logical relations between senses of the word in question.

In these cases, a closer look at the organisation of the lexical source (DDO-DanNet) seems indispensable. Some of the differences in word behaviour could probably be explained by the fact that the sense inventory of DDO has been established with the focus on the human dictionary user, and not on reflecting formal distinctions or logical relations between senses of the word in question.

Not only does doubly annotated data provide valuable feedback regarding the annotation schemes, we also think that it can help us improve our learning algorithms. Our corpus is about one fifth of the size of SemCor. However, as mentioned, a large part of the data has been doubly annotated and later adjudicated. We make available both the final adjudicated version and the individual annotations in order to facilitate research that deals with the linguistic information that resides in agreement variation. In Plank et al. (2014) and Martínez Alonso et al. (2015a) we present an algorithm that learns regularizers from small seeds of doubly annotated data. In future work we will apply SemDaX for further experiments along the same lines.

Finally, our project includes a pilot study on the compilation of a Danish framenet (similar to the well-known Berkeley FrameNet, cf. <https://framenet.icsi.berkeley.edu/fndrupal/>). This part of the project has been embarked recently by focusing on the approx. 1/3 of the sentences of our corpus where cognition and communication verbs are present (identified via the previously mentioned supersense annotation). A comprehensive frame-lexicon for Danish cognition and communication verbs has been compiled,

based on the extraction of groups of verbs with these senses from a Danish thesaurus which is also based on, and linked to, the sense inventory of DDO and DanNet (Nimb et al. 2014). The frame lexicon allows us to present the annotators to a very reduced and precise set of possible frames to choose from. In these annotations we investigate whether a frame lexicon based on a thesaurus in fact covers a satisfactory part of the corpus occurrences, and whether the role inventory of each frame can be transferred successfully from English.

References

- Artstein, R. & Poesio, M. (2008). Inter-Coder Agreement for Computational Linguistics. In: *Computational Linguistics*. 34. 4.
- Asmussen, J. (2012). CLARIN-Referencekorpus. *Sprogteknologisk Workshop October 31, 2012*, University of Copenhagen. <http://cst.ku.dk/Workshop311012/sprogtekno2012.pdf>
- Agirre, E. & de Lacalle, O.L. (2003). Clustering word senses. In: *Proceedings of the Conference on Recent Advances on Natural Language Processing (RANLP '03)*. Borovets, Bulgaria.
- Brown, S.W., Rood, T., & Palmer, M. (2010). Number or Nuance: Which factors restrict reliable word sense annotation? *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC'10)*. Valletta, Malta.
- Ciaramita, M., Johnson, M. (2003). Supersense Tagging of Unknown Nouns in WordNet. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*.
- Cruse, D.A. (2000). *Meaning in Language: an introduction to semantics and pragmatics*. Oxford University Press.
- Hjorth, E. & Kristiansen, K. (Eds.) (2005). Den Danske Ordbog (DDO). Det Danske Sprog- og Litteraturselskab & Gyldendal, Copenhagen, Denmark. Online: ordnet.dk/ddo
- Ide, N. & Wilks, Y. (2007). Making Sense About Sense. In: Agirre, E. & Edmonds, P. (Eds): *Word Sense Disambiguation - Algorithms and Applications*. Springer.
- Johannsen, A., Hovy, D., Martínez Alonso, H., Plank, B., & Søgaard, A. (2014). More or less supervised supersense tagging of Twitter. *Proceedings of *SEM, 1-11*.
- Kilgarriff, A.(2007). Word senses. E. Agirre & P. Edmonds (Eds.): *Word Sense Disambiguation - Algorithms and Applications*. Springer.
- Landes, S., Leacock, C., Tengi, R. (1998). Building Semantic Concordances. Fellbaum (ed.) *WordNet – an*

- Electronic Lexical Database*. MIT Press.
- Martínez Alonso, H., Plank, B., Kjærholt, A., Søgaard, A. (2015a). Learning to parse with IAA-weighted loss. *North American Chapter of the Association for Computational Linguistics (NAACL)*. Denver.
- Martínez Alonso, H., Plank, B., Johannsen, A., & Søgaard, A. (2015b). Active Learning for Sense Annotation. In *015 Proceedings of the 20th Nordic Conference of Computational Linguistics Nodalida 2015*. Linköping University Electronic Press.
- Martínez Alonso, H., Johannsen, A., Olsen, S., Nimb, S., Sørensen, N., Braasch, A., Søgaard, A., Pedersen, B. (2015c). Supersense tagging for Danish. In *Proceedings of the 20th Nordic Conference of Computational Linguistics Nodalida 2015*. Linköping University Electronic Press.
- Martínez Alonso H., Johannsen, A., Olsen, S., Nimb, S. and Pedersen, B. (2016). An empirically grounded expansion of the supersense inventory. *Proceedings of the 8th Global Wordnet Conference*, Bucharest, Romania
- Melo, G. de, Baker, C., Ide, N., Passonneau, R., Fellbaum, C. (2012). Empirical Comparisons of MASC Word Sense Annotations. *Proceedings from LREC 2012*, Istanbul, Turkey.
- Nimb, S., Lorentzen, H., Theilgaard, L., Troelsgård, T. (2014). *Den Danske Begrebsordbog*. Det Danske Sprog- og Litteraturselskab, Copenhagen, Denmark.
- Olsen, S., Pedersen, B., Martínez Alonso, H., Johannsen, A. (2015). Coarse-grained sense annotation of Danish across textual domains. *Proceedings of the Workshop on Semantic resources and Semantic Annotation for Natural Language Processing and the Digital Humanities at NODALIDA 2015*. Linköping University Electronic Press. Sweden.
- Palmer, M., Dang, H.T., Fellbaum, C. (2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Natural Language Engineering, Vol. 13:2*.
- Passonneau, R., Baker, C., Fellbaum, C., Ide, N. (2012). The MASC Word Sense Sentence Corpus. *Proceedings from LREC 2012*, Istanbul, Turkey.
- Petrolito, T. & Bond, F. (2014). A survey of WordNet Annotated Corpora, In *Proceedings of the Seventh Global WordNet Conference, Tartu, Estonia, January 25-29 2014*, pp. 236-243. Tartu, Estonia.
- Plank, B., Hovy, D., Søgaard, A. (2014). Learning part-of-speech taggers with inter-annotator agreement loss. *The 14th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. Gothenburg, Sweden.
- Pedersen, B.S., Nimb, S., Asmussen, J., Sørensen, N., 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. Volume 43, Issue 3*, pp.269-299.
- Pustejovsky, J. & Rumshisky, A. (2010). Mechanisms of Sense Extension in Verbs. G-M de Schryver (ed.) *A Way with Words: Recent Advances in Lexical Theory and Analysis. A Festschrift for Patrick Hanks*. Kampala: Menha Publishers.
- Qiu, L. Y.W. & Shao, Y. (2011). Combining Contextual and Structural Information for Supersense Tagging of Chinese Unknown Words. *Proceedings of the 12th International Conference on Computational Linguistics and Intelligent Text Processing, volume 6608 of Lecture Notes in Computer Science*, pp.15-28.
- Schneider, N., Mohit, B., Oflazer, & Smith, N. (2012). Coarse lexical semantic annotation with supersenses: an Arabic case study. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics*. Jeju Island, Korea.
- Yimam, S.M., Gurevych, I., Eckart de Castilho, R., & Biemann, C. (2013). WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. *Proceedings of ACL-2013, demo session*, Sofia, Bulgaria.
- Vossen, P. (ed). (1999). *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*. Kluwer Academic Publishers, The Netherlands.
- Vossen, P., Görög, A., Laan, F., van Gompel, M., Izquierdo, R., van den Bosch, A. (2011). DutchSemCor: Building a semantically annotated corpus for Dutch. *Proceedings of Electronic Lexicography in the 21st century: New Applications for new users (eLEX2011)*, Slovenia.