From Thesaurus to Framenet

Sanni Nimb[1], Anna Braasch[2], Sussi Olsen[2], Bolette Sandford Pedersen[2], Anders Søgaard[3]

[1]The Society for Danish Language and Literature, Chr. Brygge 1, 1219 Copenhagen K
[2] University of Copenhagen, Centre for Language Technology, Njalsgade 136, 2300 Copenhagen S,
[3]University of Copenhagen, Department of Computer Science, Sigurdsgade 41, 2200 Copenhagen N
E-mail: sn@dsl.dk, braasch@hum.ku.dk, saolsen@hum.ku.dk, bspedersen@hum.ku.dk, soegaard@di.ku.dk

Abstract

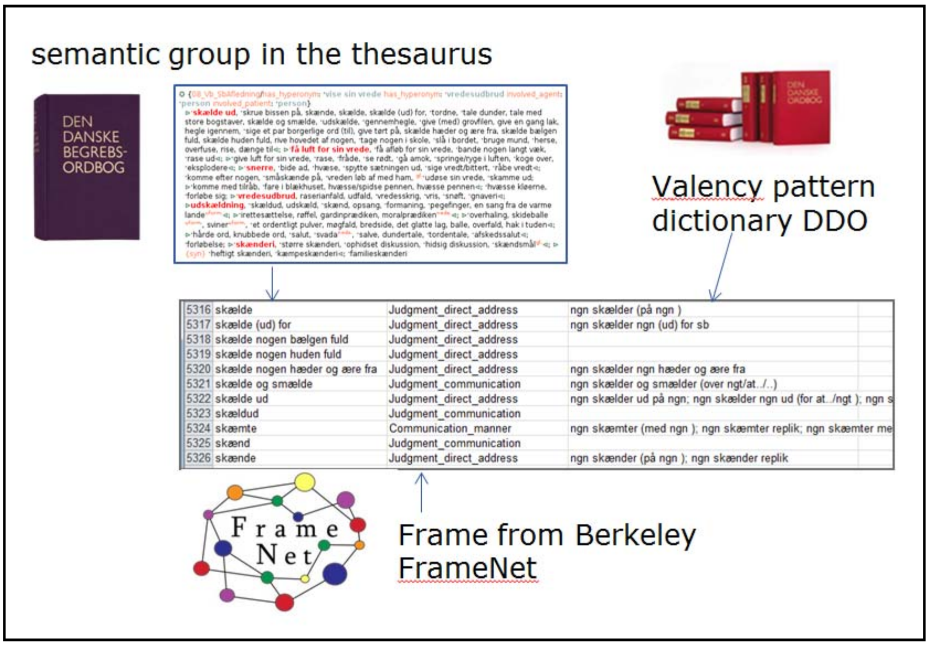
High-quality semantic data from a Danish thesaurus linked with valency information from a Danish dictionary allows us to compile a frame lexicon (Berkeley FrameNet style) for Danish in a very efficient way. In the paper we present the thesaurus as well as the dictionary and argue that they both represent valuable background information for assigning semantic frames to the Danish vocabulary. The resulting partial frame lexicon is tested in an annotation task where the semantic role inventory from English is directly transferred and made available for annotations of Danish. While simply aiming at reaching the highest possible frame coverage of the Danish vocabulary by reusing existing English frame and role inventories, we discuss the advantages and the drawbacks of the proposed method. The gained experiences from the work will be considered when scaling up the framenet resource to cover all verbs.

Keywords: Thesaurus; FrameNet; frame lexicon, Danish; annotation

1. Introduction

This article describes how we combine information from a monolingual Danish dictionary, Den Danske Ordbog (henceforth DDO) and a newly compiled Danish thesaurus, “Den Danske Begrebsordbog” (‘The Danish Concept dictionary’, Nimb et al., 2014a, henceforth the thesaurus), in order to compile standardized lexical-semantic data in the form of a partial Danish Frame lexicon compliant with the Berkeley FrameNet (BFN). The partial lexicon is tested in an annotation task carried out on already sense-annotated corpus data. The results from the pilot test are used to provide feedback to our method before we scale up the frame lexicon to cover all verbs in the thesaurus (financed 2016–2017 by the Carlsberg Foundation). We ask ourselves the following questions: How satisfying is the coverage of the generated frame lexicon based on thesaurus data, and how well can the roles described for English cover the semantics of Danish sentences?

Figure 1 illustrates the inter-linked background data.


semantic group in the thesaurus
DEN DANSKE BEGREBS-ORDBOG
DEN DANSKE ORDBOG
Valency pattern dictionary DDO
Frame Net
Frame from Berkeley FrameNet

○ {08_Vb_SbAfledning/has_hyperonym: ·vise sin vrede has_hyperonym: ·vredesudbrud involved_agent: ·person involved_patient: ·person}
⊳·**skælde ud**, ·skrue bissen på, skænde, skælde, skælde (ud) for, ·tordne, ·tale dunder, tale med store bogstaver, skælde og smælde, ·udskælde, ·gennemhegle, ·give (med) grovfilen, give en gang lak, hegle igennem, ·sige et par borgerlige ord (til), give tørt på, skælde hæder og ære fra, skælde bælgen fuld, skælde huden fuld, rive hovedet af nogen, ·tage nogen i skole, ·slå i bordet, ·bruge mund, ·herse, overfuse, rise, dænge til◁; ⊳·**få luft for sin vrede**, ·få afløb for sin vrede, ·bande nogen langt væk, ·rase ud◁; ⊳·give luft for sin vrede, ·rase, ·fråde, ·se rødt, ·gå amok, ·springe/ryge i luften, ·koge over, ·eksplodere◁; ⊳·**snerre**, ·bide ad, ·hvæse, ·spytte sætningen ud, ·sige vredt/bittert, ·råbe vredt◁; ·komme efter nogen, ·småskænde på, ·vreden løb af med ham, ·udøse sin vrede, ·skamme ud; ⊳·komme med tilråb, ·fare i blækhuset, hvæsse/spidse pennen, hvæsse pennen◁; ·hvæsse kløerne, ·forløbe sig; ⊳·**vredesudbrud**, raserianfald, udfald, ·vredesskrig, ·vris, ·snæft, ·gnaveri◁; ⊳**udskældning**, ·skældud, udskæld, ·skænd, opsang, ·formaning, ·pegefinger, en sang fra de varme lande◁; ⊳·irettesættelse, røffel, gardinprædiken, moralprædiken◁; ⊳·overhaling, skideballe, sviner, ·et ordentligt pulver, møgfald, bredside, det glatte lag, balle, overfald, hak i tuden◁; ⊳·hårde ord, knubbede ord, ·salut, ·svada, ·salve, dundertale, ·tordentale, ·afskedssalut◁; ·forløbelse; ⊳·**skænderi**, ·større skænderi, ·ophidset diskussion, ·hidsig diskussion, ·skændsmål◁; ⊳ (syn) ·heftigt skænderi, ·kæmpeskænderi◁; ·familieskænderi

5316	skælde	Judgment_direct_address	ngn skælder (på ngn)
5317	skælde (ud) for	Judgment_direct_address	ngn skælder ngn (ud) for sb
5318	skælde nogen bælgen fuld	Judgment_direct_address	
5319	skælde nogen huden fuld	Judgment_direct_address	
5320	skælde nogen hæder og ære fra	Judgment_direct_address	ngn skælder ngn hæder og ære fra
5321	skælde og smælde	Judgment_communication	ngn skælder og smælder (over ngt/at../..)
5322	skælde ud	Judgment_direct_address	ngn skælder ud på ngn; ngn skælder ngn ud (for at../ngt); ngn s
5323	skældud	Judgment_communication	
5324	skæmte	Communication_manner	ngn skæmter (med ngn); ngn skæmter replik; ngn skæmter me
5325	skænd	Judgment_communication	
5326	skænde	Judgment_direct_address	ngn skænder (på ngn); ngn skænder replik

Figure 1: Linked data: The word groups in a Danish thesaurus combined with the valency information in a Danish dictionary constitute the background for the framenet.

We first introduce the research project of which the pilot frame lexicon project is a subpart, including a presentation of the sense-annotated SemDaX corpus that has been established in the project and which guides the choice of semantic coverage of the lexicon we compile. In Section 3 we discuss how role semantic information supplements the semantics of sense annotations and argue that the BFN model is well-suited for our purpose. In Section 4 we present the lexical data we use from the dictionary and the thesaurus and present our method for compiling Danish frame data. We furthermore describe how we tested the frame lexicon in an annotation task. In Section 5 we discuss the results: Finally we draw an overall conclusion and outline future plans in Section 6.

2. The Danish FrameNet in a broader context

Our method has evolved within a research project on semantic processing ("Semantic Processing across Domains", financed by the Danish Research Council 2013–2017) where several annotation tasks were carried out and used in machine learning experiments (Pedersen et al., 2014; 2016). The project focuses on Danish as a relatively low-resourced language and aims at increasing the level of semantic resources available for the Danish HLT community. A primary project goal is to provide semantically-annotated text corpora of Danish and to let these serve as training data for advanced machine learning algorithms which particularly address data scarcity and domain adaptation as central focus points. A corpus of 100,000 words has been sense-annotated with so-called supersenses (cf. Martínez Alonso et al.,

2016) and a smaller part of this has been annotated with semantic roles (frame elements) based on the frame lexicon that we describe below. The supersense annotations guided the first selection of relevant corpus data for our pilot frame semantics study on cognition and communication events.

2.1 The SemDaX corpus

The supersenses used to annotate the SemDaX corpus are based on the Princeton Wordnet lexicographical classes[1] which have become an international standard in coarse-grained sense tagging. The number of annotated sentences in SemDaX is 3,300, of which 60% have been annotated by two or more annotators, based on which a gold standard was developed. The SemDaX corpus[2] consists of various textual domains: newswire, blogs, chat, forum, magazine and written Parliament debates (Martínez et al., 2015; Olsen et al., 2015).[3]

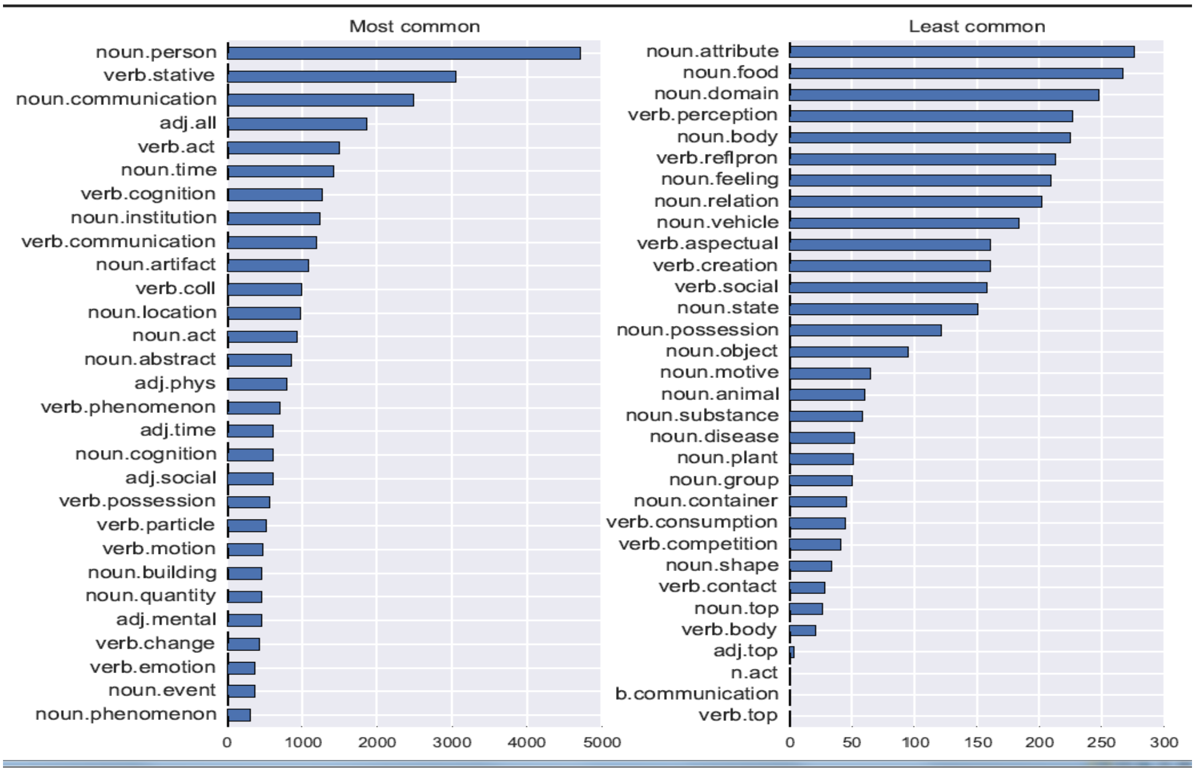


Figure 2: Most and less frequent supersenses in the complete annotated corpus (cf. Olsen et al., 2015)

[1] Cf. https://wordnet.princeton.edu/man/lexnames.5WN.html.

[2] Available for research at https://github.com/coastalcph/semdax.

[3] The texts have been extracted from the CLARIN reference Corpus, Asmussen 2012.

The most frequent supersenses in the corpus across word classes are ‘noun.person’, ‘noun.communication’ and ‘verb.stative’ (mainly constituted by the verb *være* (to be)), followed by supersenses for act, time, cognition and communication. It is interesting that the supersenses have a very different distribution across the various textual domains, revealing to a certain degree what the texts are mostly about. The supersense ‘noun.person’ is the most frequent in newswire and magazines, but much less frequent in chats, where the most frequent supersense instead is ‘verb.stative’ mainly constituted by the verb *være* (to be). Abstract supersenses such as ‘noun.abstract’ and ‘noun.act’ are much more frequent in Parliament debates than in the other text types. The least frequent supersenses in the corpus are either very specific ones, e.g. ‘verb.body’, ‘verb.competition’, ‘noun.plant’ and ‘noun.disease’, or abstract supersenses that the annotators, judged by the low inter-annotator agreement, found difficult to understand, such as ‘noun.attribute’, ‘noun.relation’ and ‘noun.domain’.

A point of great interest to our lexicon project is the frequency of the verb supersenses. Apart from the supersenses ‘stative’ and ‘act’, ‘verb.cognition’ and ‘verb.communication’ are the most common, and put together these two categories are as frequent as the most frequent verb category, ‘verb.stative’.

2.2 Selecting the frame lexicon vocabulary from the thesaurus

The supersense annotations in SemDaX enabled us to focus directly on very frequently occurring events describing communication and/or cognition. This choice was based on a comparison of the most likely supersenses of verbs in the thesaurus chapters, see Table 1, with the frequency of the different supersenses in SemDaX as illustrated in Figure 2.

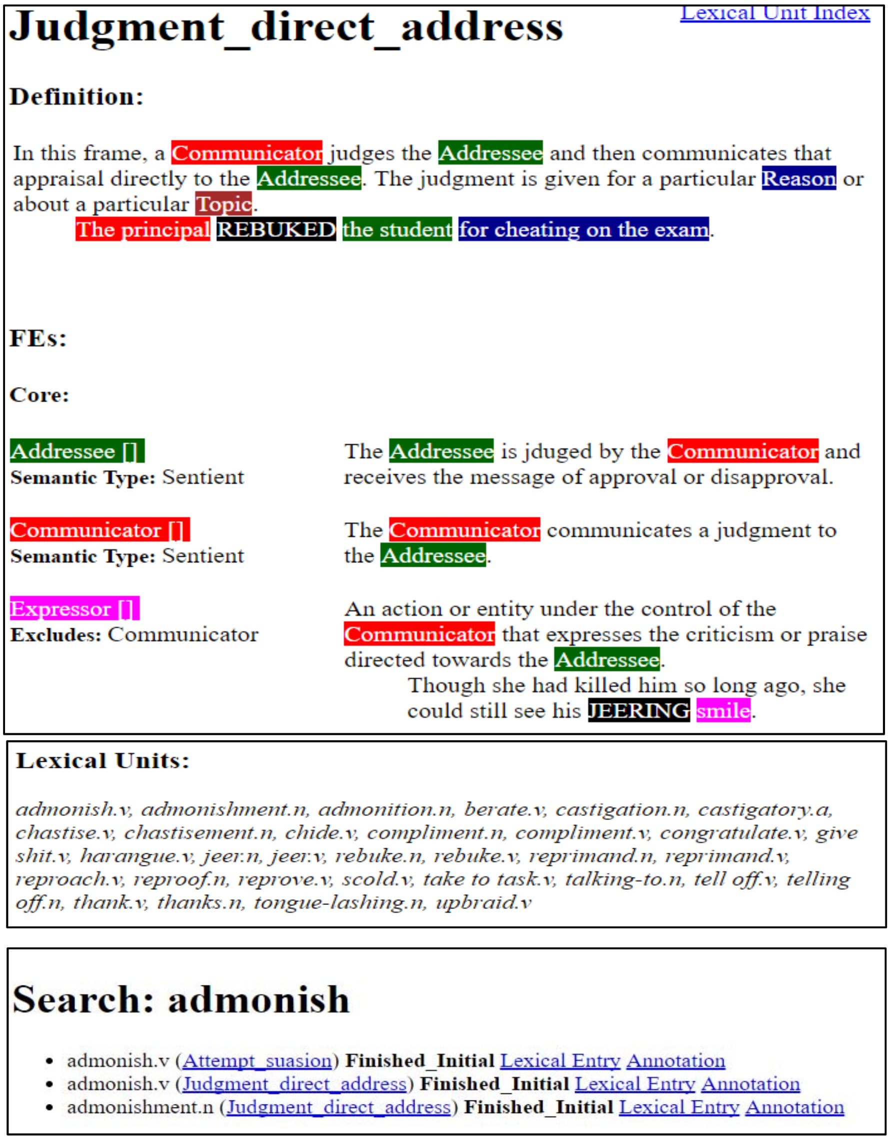
The chapters which contain a rather high number of verbs and verbal nouns compared to the average of 2% are the following: ‘5 Relation, property’, ‘8 Location, motion’, ’9 Volition, act’, ‘10 Emotions’, ’11 Thinking’, ’12 Communication’,’15 Social life’, and ‘21 Economy, finances’. A comparison with the most frequent supersenses of verbs in Figure 2 (‘stative’, ‘communication’, ‘cognition’, and ‘act’) led to the decision that in order to obtain enough sentences to annotate, the best choice would be the chapters ’11 Thinking’, ’12 Communication’ and parts of chapter ’13 Science´and ’15 Social life’ which we, based on our detailed knowledge of the thesaurus, estimate to contain mainly the very frequent supersenses ‘cognition’ and ‘communication’. Although ‘act’ verbs are typically found in chapter ‘9 Volition, act’, they are likely to also occur in a large variety of other chapters and therefore not suitable for our task. The chapters ‘8 Location, motion’, ‘10 Emotions’ and ‘20 Economy, finances’ were discarded because the corresponding supersenses ‘verb.motion’, ‘verb.emotion’ and ’verb.possession’ are not among the most frequent in Figure 2. Chapter 5 was discarded even though it contains many stative verbs which are frequent in texts, simply due to the fact that the BFN model focuses on the part of the vocabulary describing human activity.

Chapter in thesaurus	**Percentage of all verbs and verbal nouns**	**Expected to contain verbs with the following supersense:**
1 Natur og miljø (nature, environment)	1,4 %	phenomenon, act
2 Liv (life)	5 %	phenomenon, stative, body
3 Rum, form (space, form)	2,5 %	change, stative, contact
4 Størrelse, mængde, tal, grad (size, amount, number, degree)	4 %	change, quantity, relation
5 Forhold, egenskab (relation, property)	6,6 %	stative, phenomenon, relation, change, aspectual
6 Tid (time)	2,7 %	Time
7 Sanseindtryk, tilstandsformer (sense impression, material state)	4,1 %	Perception
8 Sted og bevægelse (location, motion)	9 %	Motion
9 Vilje og handling (volition, act)	11,8 %	Act
10 Følelser (emotions)	8.4 %	Emotion
11 Tænkning (thinking)	7 %	Cognition
12 Tegn, meddelelse, sprog (communication)	6 %	Communication
13 Videnskab (science)	1,4 %	Cognition
14 Kunst og kultur (arts, culture)	1,7 %	Creation
15 Socialt liv (social life)	8,6 %	social, competition, communication
16 Mad og drikke (food and drinks)	1,7 %	Consumption
17 Sport og fritid (sports and leisure)	3,6 %	body, creation, motion, competition
18 Samfund (society)	5,1 %	Social
19 Apparater, teknik (artifacts/instruments, technique)	3 %	creation, communication
20 Økonomi, finans (economy, finances)	7,1 %	possession, social
21 Ret, etik (law court, ethics)	2,1 %	Social
22 Religion (religion)	0,5 %	Cognition

Table 1: Number of verbs and verbal nouns in the 22 thesaurus chapters, and their estimated supersense types. They constitute a total of 44,607 word and expressions (=20% of whole thesaurus)

3. FrameNet as semantic model

While supersense annotations supply us with very coarse-grained semantic information at sense level, role-oriented semantic annotations are needed if we want to label in a formalized way who does what, where and when. An ongoing discussion in the Danish group has been whether to adopt a deep-syntactic approach to role-labeling as taken in PropBank (Palmer et al., 2005) and VerbNet (Schuler 2005) or a more semantically-driven, frame-based approach to roles as provided by BFN, where both the frame inventory and the frame elements describe verb semantics at quite a detailed level: what kind of act (of about 1,000 possible) is carried out, and who are the participants (e.g. speaker and addressee). Figure 3 shows the BFN interface with descriptions of frames, English lexical units and search facilities.



Judgment_direct_address

Lexical Unit Index

Definition:

In this frame, a Communicator judges the Addressee and then communicates that appraisal directly to the Addressee. The judgment is given for a particular Reason or about a particular Topic.

The principal REBUKED the student for cheating on the exam.

FEs:

Core:

Addressee [] **Semantic Type:** Sentient	The Addressee is jduged by the Communicator and receives the message of approval or disapproval.
Communicator [] **Semantic Type:** Sentient	The Communicator communicates a judgment to the Addressee.
Expressor [] **Excludes:** Communicator	An action or entity under the control of the Communicator that expresses the criticism or praise directed towards the Addressee. Though she had killed him so long ago, she could still see his JEERING smile.

Lexical Units:

admonish.v, admonishment.n, admonition.n, berate.v, castigation.n, castigatory.a, chastise.v, chastisement.n, chide.v, compliment.n, compliment.v, congratulate.v, give shit.v, harangue.v, jeer.n, jeer.v, rebuke.n, rebuke.v, reprimand.n, reprimand.v, reproach.v, reproof.n, reprove.v, scold.v, take to task.v, talking-to.n, tell off.v, telling off.n, thank.v, thanks.n, tongue-lashing.n, upbraid.v

Search: admonish

- admonish.v (Attempt_suasion) **Finished_Initial** Lexical Entry Annotation
- admonish.v (Judgment_direct_address) **Finished_Initial** Lexical Entry Annotation
- admonishment.n (Judgment_direct_address) **Finished_Initial** Lexical Entry Annotation

Figure 3: Frame description from BFN (the frame Judgment_direct_address), including lexical units and also the search facility where different frames of the same verb, here *admonish*, are presented, one of which is the above frame. Cf. Berkeley FrameNet

In recent years, frame-semantic *parsing* has received increased interest in the NLP community, and in spite of BFN's relatively fine-grained inventory of frames and frame elements, this approach has also proven manageable in practical tasks (cf. Section 2a). Frame-semantic parsing was introduced to the NLP community in SemEval 2007, with the introduction of a standard bench-marking corpus for English. Parsing models, such as the two-stage parsing model of Dipanjan Das et al. (2014), have been applied to various tasks, both within research and industry. Two examples of tasks that benefit greatly from frame-semantic parsing are knowledge base population (Søgaard et al., 2015) and document summarization (Schluter & Søgaard, 2015). Frame-semantic parsing is also likely to instigate break-throughs in question answering, relation extraction, and dialogue systems. Consequently, framenets are currently being built for a number of languages since it is seen as an important resource in a particular language's composite set of HLT resources. However, one major bottleneck for the application of frame-semantic parsers is still the lack of resources for many languages. Johannsen et al. (2015) therefore discusses cross-lingual adaptation of frame-semantic parsing models induced from the English corpus, to other languages such as Danish, German and Greek. While such work can potentially make the above technologies available for languages other than English, the models developed in Johannsen et al. (2015) were evaluated by using datasets that were not adjudicated, and where annotators did not have access to associations between trigger words and frames in the target languages. In comparison, our method suggests that annotators are presented with a list of the most *likely* frames to choose from.

Taking both the BFN as well as a semantic resource of the target language as starting points for the development of a new framenet, is not in itself a novel approach. Swedish FrameNet (Heppin & Gronostaj, 2012; 2014) applies BFN as the initial structural backbone of the resource but bases the sense inventory on a monolingual Swedish resource, SALDO. In contrast, other framenets like Japanese FrameNet (Ohara 2014) and French FrameNet (Candito et al., 2014) rely more solely on a lexical mapping from BFN, enriching and supporting the resource subsequently with corpus data in the target language.

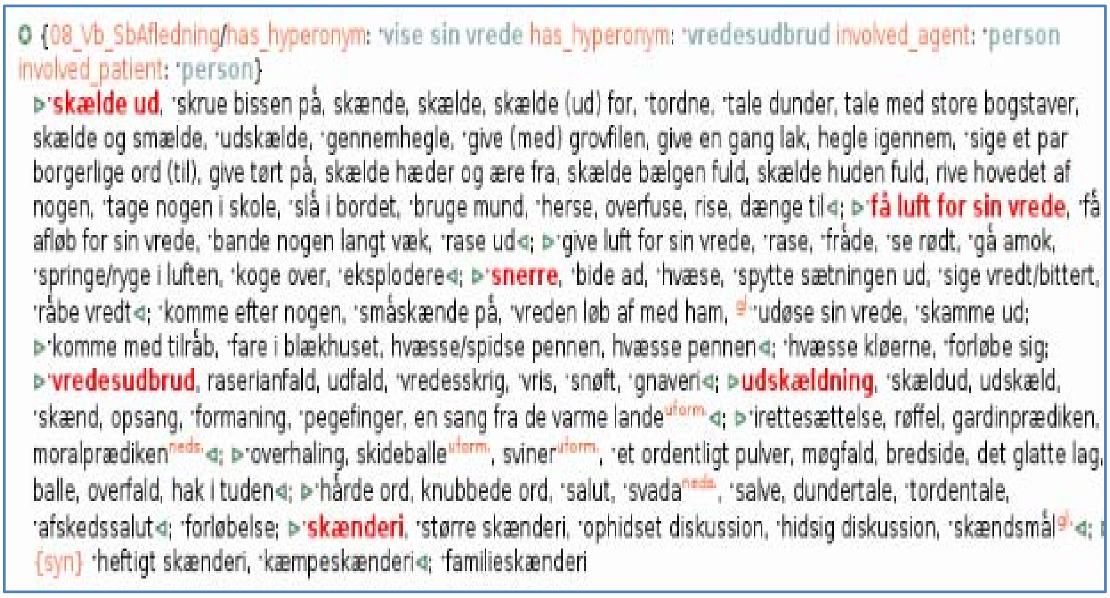
4. Compilation of a Danish Frame Lexicon

The thematic divisions in the thesaurus allow us to identify and extract large groups of near synonymous verbs within our "pilot" fields, communication and cognition. The thesaurus covers approx. 200,000 words and expressions, covering 80% of the approx. 136,000 senses described in DDO (Nimb et al., 2014b). DDO was compiled as a printed dictionary in the 90s. Today the dictionary is online and continuously extended with new words and expressions.

The thesaurus is divided into 22 named chapters and 888 named sections inspired by the division in Dornseiff (2004), but adjusted to the Danish language community of today. Each section arranges the DDO vocabulary according to semantics in lists of synonyms and near synonyms. In the source document (not in the printed book) the lists of synonyms and near synonyms are clustered in 8,300 coarse-grained semantic groups across word classes in an annotated XML structure, making it possible to identify and extract large semantic groups of words of the type ‘person’, ‘artifact’, ‘event’ etc. in each named section. By the use of these formal annotations we extracted all groups described with the semantic relation ‘involved agent’ in the chapters ‘11 Thinking’, ’12 Communication’ and furthermore some sections in ‘13 Science’ (concerning studies and science) and ‘15 Social life’ (Sections like ’15.19 Acknowledgement’, ’15.20 Flattery’, and ’15.24 Scolding’ with many communication verbs). We assumed that to a large extent these sections together would cover the verb vocabulary of cognition and communication, and thereby also the verbs annotated with these supersenses in SemDaX.

The ‘involved agent’ groups in the thesaurus include both verbs and verbal nouns, but since verbal nouns are annotated with a broad supersense ‘noun.communication’ covering both the act sense and the result, as well as semiotic artifacts in SemDaX, they are not automatically identifiable in the corpus, and we chose not to include them in the annotation task. In the lexicon, the verbal nouns are assigned frames corresponding to the verbs from which they are derived.

In Figure 4 we present an ‘involved agent’ group from the XML document.



O {08_Vb_SbAfledning/has_hyperonym: ˈvise sin vrede has_hyperonym: ˈvredesudbrud involved_agent: ˈperson involved_patient: ˈperson}

▷ˈ**skælde ud**, ˈskrue bissen på, skænde, skælde, skælde (ud) for, ˈtordne, ˈtale dunder, tale med store bogstaver, skælde og smælde, ˈudskælde, ˈgennemhegle, ˈgive (med) grovfilen, give en gang lak, hegle igennem, ˈsige et par borgerlige ord (til), give tørt på, skælde hæder og ære fra, skælde bælgen fuld, skælde huden fuld, rive hovedet af nogen, ˈtage nogen i skole, ˈslå i bordet, ˈbruge mund, ˈherse, overfuse, rise, dænge til◁; ▷ˈ**få luft for sin vrede**, ˈfå afløb for sin vrede, ˈbande nogen langt væk, ˈrase ud◁; ▷ˈgive luft for sin vrede, ˈrase, ˈfråde, ˈse rødt, ˈgå amok, ˈspringe/ryge i luften, ˈkoge over, ˈeksplodere◁; ▷ˈ**snerre**, ˈbide ad, ˈhvæse, ˈspytte sætningen ud, ˈsige vredt/bittert, ˈråbe vredt◁; ˈkomme efter nogen, ˈsmåskænde på, ˈvreden løb af med ham, $^{\text{gl}}$ ˈudøse sin vrede, ˈskamme ud; ▷ˈkomme med tilråb, ˈfare i blækhuset, hvæsse/spidse pennen, hvæsse pennen◁; ˈhvæsse kløerne, ˈforløbe sig; ▷ˈ**vredesudbrud**, raserianfald, udfald, ˈvredesskrig, ˈvris, ˈsnøft, ˈgnaveri◁; ▷**udskældning**, ˈskældud, udskæld, ˈskænd, opsang, ˈformaning, ˈpegefinger, en sang fra de varme lande$^{\text{uform}}$◁; ▷ˈirettesættelse, røffel, gardinprædiken, moralprædiken$^{\text{neds}}$◁; ▷ˈoverhaling, skideballe$^{\text{uform}}$, sviner$^{\text{uform}}$, ˈet ordentligt pulver, møgfald, bredside, det glatte lag, balle, overfald, hak i tuden◁; ▷ˈhårde ord, knubbede ord, ˈsalut, ˈsvada$^{\text{neds}}$, ˈsalve, dundertale, ˈtordentale, ˈafskedssalut◁; ˈforløbelse; ▷ˈ**skænderi**, ˈstørre skænderi, ˈophidset diskussion, ˈhidsig diskussion, ˈskændsmål$^{\text{gl}}$◁; {syn} ˈheftigt skænderi, ˈkæmpeskænderi◁; ˈfamilieskænderi

Figure 4: ’Involved agent group’ from the Danish Theaurus. The header contains annotations and introduces a large list of verbs and verbal expression with the sense ‘skælde ud’ (’to scold’, initiated by ‘skælde ud’) followed by a list of verbal nouns with the same sense (initiated by ‘vredesudbrud’)

As stated above, each word and expression in the thesaurus is linked to a DDO sense via a common identification number; this opens up a large variety of combined lexical data across the two resources, one of which we exploit here by transferring valency patterns from DDO to the verbal groups in the thesaurus.

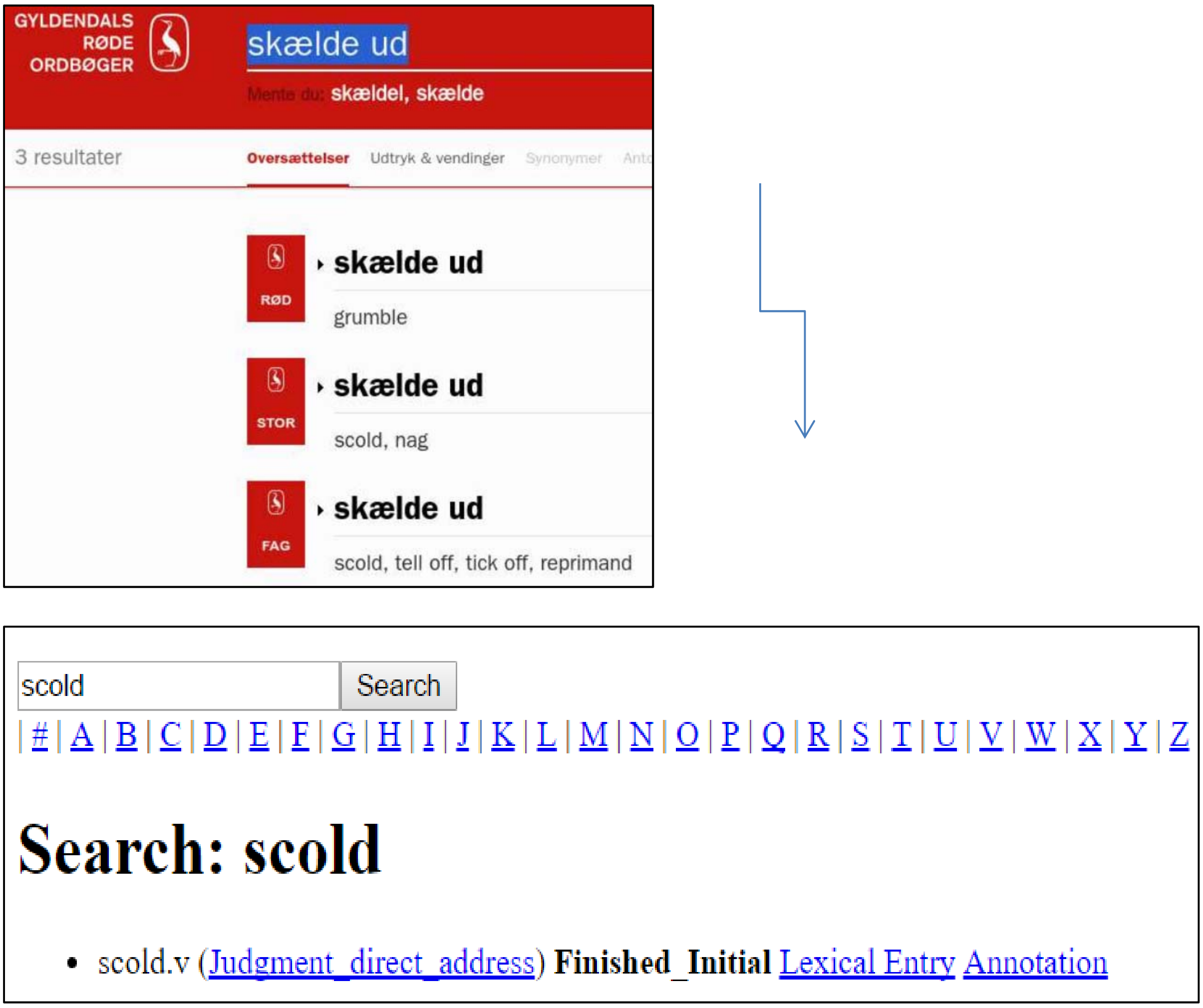
We extracted approx. 7,000 words and expressions, constituting about 16% of all verbs and verbal nouns in the thesaurus XML document (see Table 1 above). This indicates that we find many synonymous and near synonymous words and expressions within the semantic areas of cognition and communication. There seems to be some kind of parallel between frequency in Danish texts and frequency in the Danish lexicon, also when we compare other chapters in Table 1 with the supersense frequencies in Figure 2. When we often talk about a theme or concept it seems to influence the variety of words and expressions that we use in order to do it.

In Table 2 we see the extract of the same data, now supplied with valency patterns from DDO via the shared identification numbers, and supplied with the information on the corresponding frame in BFN.

section title, word/expression (= 'to scold') from the thesaurus		**shared ID number**	**valency pattern from DDO**	**frame from BFN**
Skælde ud	*skælde ud*	21074700	ngn skælder ud på ngn; ngn skælder ngn ud (for at/ngt); ngn skælder (ngn) ud over ngt/at	Judgment_direct_address
Skælde ud	*skrue bissen på*	21074701	NGN skruer bissen på (over for NGN)	Judgment_direct_address
Skælde ud	*Skælde*	21010806	ngn skælder (på ngn)	Judgment_direct_address
Skælde ud	*skælde (ud) for*	21033375	ngn skælder ngn (ud) for sb	Judgment_communication
Skælde ud	*skælde nogen bælgen fuld*	21034458	NONE	Judgment_direct_address
Skælde ud	*skælde nogen huden fuld*	21074699	NONE	Judgment_direct_address
Skælde ud	*skælde nogen hæder og ære fra*	21090433	ngn skælder ngn hæder og ære fra	Judgment_direct_address
Skælde ud	*skælde og smælde*	21074701	ngn skælder og smælder (over ngt/at)	Judgment_communication

Table 2: Lexical units from the thesaurus linked to valency patterns from DDO and supplied with frames from BFN

By focusing on one semantic area at a time (made possible via the chapter grouping in the thesaurus), the lexical data considered are likely to be assigned the same frame, or at least a closely related frame, from BFN. In the work process, the Danish word or expression is translated to an English equivalent (via Gyldendal's Danish English Dictionary), and the equivalent (or a more common synonym) is searched for in the lexical unit index of BFN, leading to one or more frame possibilities, see Figure 5. The frame description is studied carefully before it is assigned, see Figure 3 above. It has to be verified whether it covers the Danish lexical unit *skælde ud*, e.g. by comparing the Danish valency pattern and the role inventory of the frame.



skælde ud

Mente du: skældel, skælde

3 resultater

Oversættelser
Udtryk & vendinger
Synonymer
Anto
RØD
STOR
FAG

skælde ud

grumble

skælde ud

scold, nag

skælde ud

scold, tell off, tick off, reprimand

scold

Search
| # | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z

Search: scold

- scold.v (Judgment_direct_address) **Finished_Initial** Lexical Entry Annotation

Figure 5: Translation of the Danish verb *skælde ud*. Equivalent 'to scold' used as input to manual search for a relevant frame in BFN (Judgment_direct_address, see Figure 3)

To cover approx. 3,500 words and expressions describing the semantic area of communication (Chapter 12 and part of 15), we used the following 52 BFN frames: Be_in_agreement_on_action, Be_in_agreement_on_assessment, Attempt_suasion, Attention, Become_silent, Bragging, Chatting, Commitment, Communicate, Communication_manner, Communication_noise, Communication_response, Contacting, Deny_permission, Discussion, Education_teaching, Encoding, Gesture, Going_back_on_a_commitment, Grant_permission, Hearsay, Intentional_deception, Judgement_communication, Judgment_direct_address, Justifying, Label, Linguistic_meaning, Manipulate_into_doing, Mention, Name_conferral, Permission, Prevarication, Publishing, Quarreling, Questioning, Reading_aloud, Reassuring, Reporting, Request, Respond_to_proposal, Response, Reveal_secret, Silencing, Spelling_and_pronouncing, Statement, Suasion, Summarizing, Telling, Text_creation, Translate, Verification, Warning.

To cover approx. 2,600 words and expressions describing the semantic area of cognition (Chapter 11 and part of 13), we used the following 54: Adding_up, Adducing, Annoyance, Attention, Awareness, Becoming_aware, Categorization, Certainty, Cogitation, Coming_to_believe, Coming_up_with, Correctness, Creating, Differentiation, Education_teaching, Estimating, Evoking, Examination, Expectation, Experiencer_focus, Experiencer_obj, Experimentation, Feigning, Grant_permission, Grasp, Intentional_deception, Intentionally_act, Judgment, Just_found_out, Linguistic_meaning, Make_cognitive_connection, Manipulate_into_doing, Memorization, Memory, Mental_property, Opinion, Perception_active, Purpose, Questioning, Reading_activity, Reading_perception, Reasoning, Regard, Reliance_on_expectation, Remembering_experience, Remembering_information, Remembering_to_do, Research, Resolve_problem, Reveal_secret, Scrutiny, Sign, Topic, Trust.

In both cases the number of used frames constitute only about 5% of the 1,073 frames described in BFN. By focusing on only one semantic area at a time—first communication, then cognition—we made it possible for the lexicographer to gain confidence in the different frame descriptions, enabling her to distinguish between semantically closely related frames and to carry out a more homogenous assignment of frames. The information on valency patterns from DDO was crucial when it came to the lexicographer's clarification of the scenario in question in Danish, and her choice of exactly the one English frame which would cover the sense and the connected constituents as described in the valency pattern in the best way.

4.1 The annotation task

For the annotation task, sentences in SemDaX with verbs already annotated with the supersenses cognition and/or communication were extracted and assigned frames.

The annotation tool by Johannsen presents the annotator with the corresponding frame of the verb which has to be confirmed or rejected. In case of more than one frame for a given verb, the set of frames are listed and the annotator selects the right one after having checked the lexicon (which often presents the verb with different collocates, e.g. the verb *indsamle* ('to collect') with the noun *viden* ('knowledge') in Table 3) or/and BFN.

Danish lexical unit	**Frame from BFN**	**Danish valency pattern**
indsamle viden (lit. 'collect knowledge' ('study'))	Scrutiny	ngn indsamler ngt
indse ('understand')	Be_in_agreement_on_assessment	ngn indser at sætn/ngt
indse ('realize')	Coming_to_believe	ngn indser at sætn/ngt
indse ('realize')	Coming_to_believe	ngn indser ngt
indskole ('do introductory schooling')	Education_teaching	
indskrive ('register/inscribe')	Text_creation	ngn indskriver ngn/ngt
indskyde ('add')	Mention	ngn indskyder ngt/at sætn

Table 3: Alphabetic extract from the lexical unit index of Danish words and expressions

Once the most appropriate frame is selected, its role inventory (transferred from BFN to the annotation tool) is studied in the BFN descriptions of the frames (in case of doubts) and used for annotation, based on the assumption that the inventory covers the set of Danish roles as well due to the relative similarity between the two languages and linguistic communities.

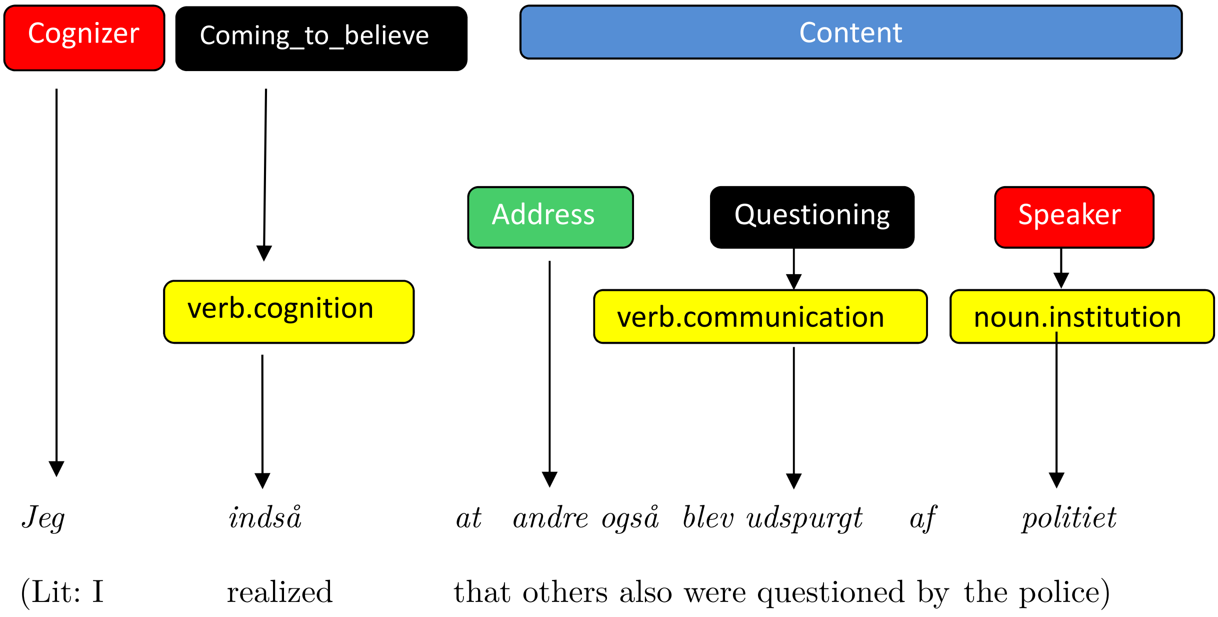


Figure 6: BFN frames and roles annotated on top of Danish supersense annotations

Regarding the sentence in Figure 6: "*Jeg indså at andre også blev udspurgt af politiet*" ('I realized that others were also questioned by the police'), the annotator is presented to two options (via the annotation tool) for *indse* ('realize'/supersense verb.cognition), namely Be_in_agreement_on_assessment and Coming_to_believe. The latter is chosen (after having checked the lexical unit index in Table 3 in case of doubt), and the core roles of the frame are studied in BFN and annotated in the sentence as well, in this case Cognizer ("Jeg" ('I')) and Content (the complement clause "*at andre også blev udspurgt af politiet*" ('that others were also questioned by the police')). Furthermore the main verb *udspurgt* ('to question', 'to pump', verb.communication) is annotated with the frame Questioning, of which the present roles in the phrase are Speaker (*politiet* ('the police')) and Addressee (*andre* ('others')).

In total, 440 cognition and communication verbs in SemDaX were annotated and will later be used in different machine learning experiments.

5. Discussion of method

We argue that the very fact that DDO is corpus based—as is the thesaurus since it uses DDO as its lexical backbone—makes both resources well qualified as background resources for creating lexical frames. But the method also has some pitfalls, as we will demonstrate.

5.1 The advantages and disadvantages of using the DDO valency patterns

DDO is corpus based. This includes the description of the valency patterns which is established on the study of a set of randomly chosen concordance examples, typically 100–200 sentences, for high-frequent verbs with many senses, up to 1,000 examples. One could thereby claim that the valency patterns function as a sort of condensed extract of the verbs' linguistic behavior in real text, including the semantic roles they typically occur with, similar to that for which we would expect to seek and annotate in the SemDaX corpus. They contribute with very important information when the frame lexicon is compiled. But a drawback is the differences between SemDaX and the corpus used to compile DDO in the 90s. The sentences we annotate constitute newer texts (2008–2011) and cover a wider range of (new) text domains than does DDO, such as blogs and chat from the Internet.

The valency patterns in DDO describe to the dictionary user whether the verb in the same sense also might be construed as a phrasal verb with a particle (presented in brackets), whether the constituents of the verb are facultative (presented in brackets) or not, whether they are introduced by an obligatory or facultative preposition, selectional restrictions such as 'person' or 'not person', or maybe instead a phrase or an infinitive construction. Sometimes additional selectional restrictions are mentioned, e.g. 'animal'. The sense of a verb might even have several valency patterns, each of them with facultative complements or particles. The patterns aim at making the dictionary user able to construct well-formed sentences in Danish with the verb in question, but they are not described by an unambiguous, formalized pattern; they depend on human interpretation.

When used in combination with the semantic grouping from the thesaurus to compile the frame lexicon, the valency patterns function as a clear indicator of which type of frame to assign from BFN. The different patterns of a semantic group also support one another, making the picture even clearer. The constituents in the patterns are strongly connected to the (core) roles described for each frame in BFN. Altogether the exact scenario evoked by the Danish word in question becomes quite clear through the comparison of valency descriptions and the frame.

Valency pattern in DDO	**Lit.**	**English equivalent**	**Frame from BFN**
NOGEN skælder (NOGEN) ud over NOGET/at	somebody scolds (somebody) out over something / that = somebody scolds (somebody) because of something/ because he/she	somebody scolds somebody because of something/ because he/she	Judgment_direct_adress
		somebody nags about something/that somebody	Judgment_communication
NOGEN skælder ud på NOGEN	somebody scolds out at somebody	somebody scolds somebody	Judgment_direct_adress
NOGEN skælder NOGEN ud (for at/NOGET)	somebody scolds somebody out (for that /for something	somebody scolds somebody (for doing) (for something)	Judgment_direct_adress

Table 4. The valency patterns of *skælde ud* (‘scold’) in DDO is complex, involving several facultative complements, and it is therefore likely that the Danish verb is to be assigned more than just one BFN frame

It is important to underline that there is no one-to-one correspondence between senses and valency patterns in DDO on the one side, and frames in BFN on the other side. The same sense of a verb in DDO might be assigned more than one frame in our lexicon. It complicates the process that the choice of frame might depend on whether or not facultative elements of the valency pattern correspond to semantic roles. The phrasal verb skælde ud (lit. ‘scold out’ (‘scold’ as in ‘scold somebody for something’)) is such a case, as shown in Table 4. BFN distinguishes between scenarios where somebody is criticizing a person directly in front of him or her. In this case the frame is Judgment_direct_adress. Scenarios where somebody is talking negatively about something, e.g. what a person who is not present did (= nagging about somebody) the frame is instead Judgment_communication. The Danish verb skælde ud covers both senses (as seen in Table 4, Gyldendal translation), and only the presence of specific semantic roles clarifies the sense in question. It is not clarified in the definition of the word sense in DDO that this is the case. In many cases the thesaurus presents such ambiguous senses in more than one section. E.g. skælde ud is also mentioned in chapter ‘10 Emotions’ in the section ‘10.26 Unsatisfied’ together with other verbs with the sense ‘to complain’, ‘to nag’, and would have been assigned the frame Judgment_communication if words from this section had been included in our frame lexicon vocabulary.

5.2 The advantages and disadvantages of using the thesaurus as input to a framenet

The thesaurus presents a large variety of lexical data in the form of extensive lists of near synonymous words and multiword units. It often displays the same sense of DDO in different, more or less fixed expressions. Thereby the thesaurus supplies us with far more multiword units than does DDO. E.g. in the case of facultative particles in the valency patterns, the thesaurus presents two lexical units where DDO only provides us with one. The DDO verb sense of *printe* (‘to print’/’to print out’) with the valency pattern “NGN printer NGT (ud)” (‘somebody prints something (out)’) thereby results in two synonymous lexical units, corresponding to print and print out in English, listed together in the thesaurus in the same semantic group with other synonymous verbs (*udprinte*, *udskrive* and *skrive ud*).

Given that DDO is corpus-based, the lexical data represents a small ‘summary’ of the behavior of the verb in real text, in line with the valency patterns but more focused on the lexical semantic restrictions.

Add to this that the thesaurus very often covers several aspects of a DDO sense by presenting it in more than just one section or chapter. Thereby it also sums up the different aspects of a word quite similar to that which we would probably discover by annotating large amounts of text (as it is done in the BFN project).

Furthermore, BFN is in many ways similar to a thesaurus as also stated in Ruppenhofer et al. (2016): “*Each lexical unit is linked to a semantic frame, and hence to the other words which evoke that frame. This makes the FrameNet database similar* *to a thesaurus, grouping together semantically similar words*". But it is important to underline that, although we find some consistencies between section divisions across the two resources, the thesaurus and BFN are profoundly very different in their way of dividing the vocabulary into sections and chapters, and frames, respectively. BFN has 'scenarios' and the core role inventory of these as the overall division criteria. As stated in Ruppenhofer et al. (2016), "*The frames represent story fragments, which serve to connect a group of words to a bundle of meanings; for example the term avenger evokes the Revenge frame, which describes a complex series of events and a group of participants*". As an example, BFN does not distinguish between negative and positive directly expressed judgments: *to compliment* and *to scold* both evoke the frame Judgment_direct_address. The 'story', or 'scenario', as well as the participants are the same; in both cases we deal with a judgment scenario. As a consequence, it distinguishes between scenarios where the participants are not the same: when a person complains about somebody who is not present and thereby not constituting the role of the addressee, the evoked frame is Judgment_communication, but when the person complained about at the same time is the addressee in the scenario, the evoked frame is Judgment_direct_adress. Likewise, antonymous words describing the same type of cognitive event, such as the verbs 'to forget' and 'to remember', are also considered to belong to the same frame. In other words, the same frame is evoked by lexical units no matter whether these are negated or not in the phrase.

In contrast, the thesaurus divides the vocabulary according to domains (football, food, movies), but also according to traditional sense division criteria. Antonomy is an important aspect, and in some chapters most of the sections could be seen as having opposite meanings to one another, covering concepts of 'thin' as opposite to 'thick', 'angry' opposite to 'happy', 'early' to 'late', 'strong' to 'weak' etc. This is also the case for cognition and communication verbs in Chapters 11, 12, 13 and 15. E.g. the Danish lexical units of Judgment_direct_adress are found in different sections such as '15.19 Approval', '15.20 Flattering' and '15.24 Scolding'. Likewise, the thesaurus contains the two sections '11.37 Remembering' and '11.38 Forgetting' while Framenet, as stated above, has only one frame covering both, namely Remembering_information. Table 5 describes the different division criteria in the two resources. Svendsen (2017: 26) proposes that we should consider adopting the method suggested by the Swedish FrameNet project (Friberg Heppin & Gronostaj 2012) who split up such frames according to positive and negative meanings, due to the fact that the Swedish lexical resource (SALDO), just like the thesaurus distinguishes clearly between such senses.

Not surprisingly we had some cases of Danish verbs that were difficult to assign an English frame. The verbs *misforstå* ('misunderstand'), *mistolke* ('misinterpret') and near synonymous words are some of these cases. Svendsen (2017) points out other problems of the language transfer method, e.g. caused by reading too much meaning into the BFN frames when they are assigned to the Danish vocabulary. We will not study and discuss in this paper whether the problems are due to differences between the Danish and English vocabulary, or rather to the fact that BFN is still being developed and therefore does not cover all possible scenarios yet. In general, we found that the Danish semantic areas we chose were in fact surprisingly well-covered in BFN, but we also found a certain lack of frames concerning what you could describe as ‘acts one does not carry out’, like *undlade* (‘to leave undone’) or ‘acts one does not succeed with’, like *overvurdere* (‘to overestimate, to overrate’). Also frames for domain terms were missing, like *anonymisere* (‘to anonymize’); naturally BFN does not yet cover all types of domains and terminology. When the whole thesaurus has been assigned frames, we will study the vocabulary left without frame assignment. Likewise it will be necessary to look at BFN frames which have not been applied to any Danish verbs.

Criteria to division in the thesaurus → Criteria to division in BFN ↓	15.19 Anerkendelse (‘approval’) only positive includes both talking about and talking directly to the person	15.24 Skælde ud (‘to scold’) only negative Includes both talking about and talking directly to the person
Judgement_communication Both positive and negative Not directly to judged person	*berømme* (‘to praise’)	skælde og smælde (’to nag’), *bande langt væk* (‘curse somebody up and down’)
Judgement_direct_adress Both positive and negative Directly to judged person	*komplimentere* (‘to compliment’)	*overfuse* (‘heap/pour abuse on’), *gennemhegle* (‘to dress someone down’)

Table 5: BFN and DT use different criteria when dividing into frames and sections respectively

5.3 Frame and role coverage in the annotation task

Before initiating the annotation task, we studied the list of the approx. 1,600 verbs in SemDaX which are annotated as either cognition or communication. By doing so, we found that approx. 20% words at a first glance did not seem to belong to any of these semantic classes. Some had a much broader sense which was used with a communication or cognition sense in the corpus while depending on a very specific context; others were ad hoc figurative senses. Such cases are typically neither represented in DDO nor in the thesaurus vocabulary. Dictionaries do not fully cover all senses of words as they are represented in corpora. When lexicographers describe the senses of a lemma, they focus on prototypical word use and normally discard senses with very low frequency. In the case of a set of quite similar, but rare sense instances, they try to merge them into one overall sense description whenever possible, and they normally discard ad hoc figurative use. Framenet projects like BFN and the Japanese FrameNet project which annotate texts instead of focusing on lexical units from a resource, do not encounter this problem. The cognition and communication verbs in SemDaX were by far the most cases described in DDO and the thesaurus, but some were presented in sections in the thesaurus which we did not consider to be relevant in the first place when we extracted communication and cognition groups. Chapter 19 of the thesaurus which covers artifacts and devices, and therefore also the sections on telephones and computers, is one such case. It describes an important part of the communication vocabulary which in BFN corresponds to the frames Communication_means and Contacting. These words will, in a future digital version of the thesaurus, be included in Chapter 12 on communication, and in this way, our project also gives feedback to the thesaurus project. The words and their corresponding frames were added to our lexicon before we initiated the annotation task.

If we turn to the results of the annotation task, the assignment was, in by far the most cases, easy to carry out and clearly facilitated by the reduced set of possible frames suggested by the annotation tool via the lexicon data. But approx. 20% of the cases gave us some interesting challenges. E.g. it turned out that some of the possible frames of the most frequent verbs in Danish were missing due to the fact that not all verb senses of highly frequent verbs with many senses in DDO are covered by the thesaurus. When the thesaurus was compiled, the aim was to include the highest number of different lemmas as possible and not to cover all senses of the same lemma as described in DDO. This has apparently led to a too narrow representation of some of the very frequent cognition and communication verbs in our pilot frame lexicon. When we expand it, these verbs will be assigned a bigger variety of frames according to their many senses in DDO. The thesaurus will once again benefit from the study: some highly polysemous verbs will have to be added to extra sections.

Interestingly enough, some verbs from the semantic area cognition in the SemDaX corpus turned out to have a communication sense. These verbs are not part of the communication vocabulary in the thesaurus since they depend so strongly on the linguistic context (they occur only together with direct speech/discourse), that it would be almost impossible to decode their communication sense for the user. One example is the verb *mene* ('to find', 'to think') as in "*Jo, vejret ser ud til at holde, mente han*" ('yes, the weather conditions seem to last, he found' (='he said'). We also find verbs from other semantic areas having communication senses in this context: *slutte* ('finish'), *fortsætte* ('to continue') and *begynde* ('to begin') as in "*Jeg har haft en drøm, begyndte han*" ('I had a dream, he started' (= 'started to say') and *gabe* ('to yawn') as in "*nu må vi se at få sovet lidt, gabte moren*" ('now we ought to sleep, the mother yawned' (='said while she yawned')). In order to significantly improve our lexicon, we must fully cover such verbs which we have completely disregarded in the first place, since we focused entirely on the thesaurus vocabulary. Most of them are in fact easily identifiable by their valency pattern in DDO which describes the possibility of direct speech.

Once the correct frame was selected, the English role inventory proved to fulfill our requirements and was in fact rather easy to apply. Most phrases, however, contained rather few realised roles, for instance, the addressee was often absent in communication phrases.

While annotating the corpus sentences, another question arose: should the annotator stick to the most 'literal' frame that the verb evokes, which is normally also integrated in the frame lexicon, or should she rather try to represent the underlying meaning in the phrase? One example is the phrase: "*Nogle personer kan du lære at leve med, andre ikke*" ('Some people you are able to learn to live with, others you are not'). Should *lære* ('to learn') in this case be annotated with the frame Grasp (with the roles Cognizer and Phenomenon) or rather with the frame Tolerating (with the roles Experiencer and Content)? In the frame lexicon, only the collocation *lære at kende* ('get to know') is described, but not *lære at leve med* ('learn to live with'). We find that this case illustrates very well why the many collocations in the thesaurus are well-suited as input to a frame lexicon; in this case *lære at leve med* is candidate to occur in the thesaurus in the same group as verbs like *tolerere* (to tolerate) and its synonyms in a future version.

6. Conclusion

Overall we can conclude from our method that any possible frame of a word that the lexicographer would even think of when assigning the frames from BFN, should better be included in the lexicon right away in order to provide the annotators with a maximal set of frames for a given word. When the full thesaurus data (that is, verbs from all 888 sections) has been assigned frames, we hope to have covered a very large variety of frame possibilities of the DDO senses. Our annotation tool did not give access to the full set of frames in BFN, only to the frames assigned to each verb in our lexicon. Even though it was very clear that the predefined and manageable set made the distinctions between frames much easier to grasp and thereby facilitated the annotation process, we soon understood that it is necessary to have access to the full set of frames in BFN in order to also be able to annotate the ad hoc language use often found in corpora but not described in dictionaries.

The frame annotations are used to train a semantic parser; however, the number of annotated sentences (440) is currently rather small for this task, and we therefore plan an extension. We also plan to look deeper into the frequency of the different frames and roles in the Danish texts, in order to compare frequency across text domains as has been done in the supersense annotation task.

7. Acknowledgements

The project “Semantic Processing across Domains” and the partial frame lexicon is financed by the Danish Research Council (2013-2017). The Carlsberg Foundation is funding the extension of the lexicon in 2016/2017.Thomas Troelsgård has carried out the extraction of XML data from the thesaurus and the dictionary. Anders Johannsen has developed the annotation tool.

8. References

Candito, M., Amsili, P., Barque L., Benamara, F., De Chalendar, G., Djemaa, M., Haas, P., Huyghe, R., Yannick Mathieu, Y., Muller, P., Sagot, B. & Vieu, L. (2014). Developing a French FrameNet: Methodology and First Results. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.

Das, D., Chen, D., Martins, A., Schneider, N. & Smith, N. (2014). Frame-semantic parsing. *Computational linguistics*, 40(1), pp. 9–56.

Friberg Heppin, K. & Toporowska Gronostaj, M. (2012). The rocky road towards a swedish framenet - creating SweFN. In *Proceedings of the Eighth conference on International Language Resources and Evaluation (LREC‘12)* Istanbul, Turkey, pp. 256–261.

Friberg Heppin, K. & Toporowska Gronostaj, M. 2014. Exploiting FrameNet for Swedish: Mismatch? *Constructions and Frames*, 6(1), pp. 52-72.

Johannsen, A., Martinez Alonso, H. & Søgaard, A. (2015). Any-language frame semantic parsing. In: Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: Association for Computational Linguistics, pp. 2062-2066.

Kipper Schuler, K. (2005). *Verbnet: a broad-coverage, comprehensive verb lexicon.* Ph.D. thesis, Philadelphia, PA, USA. AAI3179808.

Martínez Alonso, H., Johannsen, A., Nimb, S., Olsen, S. & Pedersen, B. (2016). An empirically grounded expansion of the supersense inventory. In *Proceedings of Global Wordnet Conference 2016.*

Martínez Alonso, H., Johannsen, A., Olsen, S., Nimb, S., Sørensen, N., Braasch, A., Søgaard, A. & Pedersen, B. S. (2015). Supersense tagging for Danish. In: *Proceedings of the 20th Nordic Conference of Computational Linguistics NODALIDA 2015. Vol. 109*, Linköping University Electronic Press, NEALT Proceedings Series, Vol. 23.

Nimb, S., Lorentzen, H. & Trap-Jensen, L. (2014b): The Danish Thesaurus: Problems and Perspectives. In: A. Abel, C. Vettori & N. Ralli (eds.) *Proceedings of the XVI EURALEX International Congress: The User in Focus.* 15-19 July

2014. Bolzano/Bozen 2014: EURAC Research, pp. 191-199.

Ohara, K. H. (2014). Relating Frames and Constructions in Japanese FrameNet. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland, pp. 2474-2477.

Olsen, S., Pedersen, B. S., Martínez Alonso, H. & Johannsen, A. (2015). Coarse-grained sense annotation of Danish across textual domains. In *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., Gildea, D. & Kingsbury, P. (2005). The Proposition Bank: A Corpus Annotated with Semantic Roles, *Computational Linguistics Journal, 31:1,* Association for Computational Linguistics.

Pedersen, B. S., Martínez Alonso, H., Braasch, A., Johannsen, A., Nimb, S., Olsen, S., Søgaard, A. & Sørensen, N. (2016). The SemDaX Corpus - sense annotations with scalable sense inventories. In: *Proceedings of the 10th edition of the Language Resources and Evaluation Conference, (LREC'16)*, Portorož, Slovenia.

Pedersen, B. S., Nimb, S., Olsen, S., Søgaard, A. & Sørensen, N. (2014). Semantic Annotation of the Danish CLARIN Reference Corpus. In: *Proceedings from isa-10, 10th Joint ACL - ISO Workshop on Interoperable Semantic Annotation* pp. 25-29, Reykjavik, Iceland.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F. & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice* (Revised November 1, 2016.) https://framenet.icsi.berkeley.edu/fndrupal/the_book.

Schluter, N.-E. & Søgaard, A. (2015). Unsupervised extractive summarization via coverage maximization with syntactic and semantic concepts. In: *The 53rd Annual Meeting of the Association for Computational Linguistics (ACL). Vol. 2* Association for Computational Linguistic, pp. 840-844.

Søgaard, A., Plank, B., Martinez Alonso, H. (2015). Using frame semantics for knowledge extraction from Twitter. In: *Proceedings of he twenty-ninth AAAI Conference on Artificial Intelligence: AAAI 2015.* Association for the Advancement of Artificial Intelligence, pp. 2447-2452.

Svendsen, M.M. (2017). *Constructing a FrameNet for Danish as a tool for lexicographers.* Unpublished thesis, Aarhus University, Denmark.

Websites:

Asmussen, J. (2012).The CLARIN Reference Corpus: Accessed at http://cst.ku.dk/Workshop311012/sprogtekno2012.pdf

Berkeley FrameNet: Accessed at: http://framenet.icsi.berkeley.edu.

Johannsen, Anders. FrameNet annotation tool: Accessed at https://github.com /andersjo/framenet-annotation.

Dictionaries:

DDO. Accessed at: http://ordnet.dk/ddo, Society for Danish Language and

Literature, Copenhagen, Denmark.

The Danish thesaurus: Nimb, S., Lorentzen, H., Troelsgård T., Theilgaard, L., Trap-Jensen, L. (2014a): *Den Danske Begrebsordbog*, Society for Danish Language and Literature, Copenhagen, Denmark.

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

Gyldendal Danish English Dictionary. Accessed at: ordbog.gyldendal.dk.