Propbank Annotation of Danish Noun Frames

Eckhard Bick *(University of Southern Denmark, eckhard.bick@gmail.com)*

Abstract

This paper presents a frame annotation scheme for Danish nouns, with VerbNet-derived frames and semantic roles covering both frame arguments and satellites. The scheme was implemented as a new module for a Danish frame tagger and applied to a 90,000-token Danish treebank with ongoing manual revision.

In addition to explicit frames, Constraint Grammar rules are used to map free semantic roles onto noun dependents without pre-defined frames, using general syntactic-semantic context clues. We discuss the annotation scheme and present a statistical breakdown and linguistic evaluation of the assigned noun frames and ad-nominal roles in the corpus.

1. Introduction

- Frame & role annotation are rooted in verb sense classification (e.g. Levin 1993) and the concept of semantic argument roles (Fillmore 1968)
- In a frame-based framework, these are seen as interdependent, and predications are annotated for both:
  - Lexicon: FrameNet (Baker et al. 1998, Ruppenhofer et al. 2010)
  - Corpus: PropBank (Palmer et al. 2005)
- Danish resources: Framenet.dk, frame tagger (DanGram, Bick 2011)
- **Problem:** Unlike larger languages, e.g. the German Salsa corpus (Rehbein et al. 2012), the Danish resources largely ignore nominal predications
- **Suggested solution:** Three-pronged approach with automatic corpus annotation
  - (a) systematic derivation of noun frames from verb frames
  - (b) lexicographical treatment of argument-carrying nouns
  - (c) free role-mapping rules based on semantic noun classes and syntactic triggers

2. De-verbal noun frame derivation

- Hypothesis: Frames have a sufficient level of semantic abstraction to work across not only syntactic, but also morphological/POS variants.
- Therefore: Import the frame inventory of the Danish FrameNet (~ 500 categories) *as is* for noun frame annotation.

Verbs can mimic other POS classes, while still retaining their arguments, with parallel constructions in Danish and English:

Participles and gerunds in adjective and adverb slots:

The new book, [published by Elsevier in 2012]
TH/theme   AG/agent   LOC-TMP/temporal location

Infinitives for noun slots:

[To visit Paris   without visiting the Louvre] is a weird thing to do
LOC/place   CONC/concession

Verbo-nominal inflection (rare): 'råbe' (shout) + -n = 'råben' (shouting)

Hans §SP evige råben efter mere øl §FIN
(his constant shouting for more ale)

Verbo-nominal derivation (common), creating true nouns from verbal roots, through suffixation with -else / -[n]ing:

- *Firmaets §AG nylige udfasning /V:udfase/ af bonusordninger §PAT*
  (the company's recent curb on bonus schemes)
- *fornyelse /V:forny/ af offentlige bygninger §PAT*
  (repair of public buildings)

Building a lexicon of frame-carrying nouns:

- Find nouns where stripping -else/-ing leads to a recognizable verb root
- retrieve the corresponding verb frames
- exclude nouns with semantic class tags incompatible with actions, activities, events or processes
  - *følelse* [emotion, not "to feel"]
  - *forretning* [shop, not "to do business"]
- consider compounding with incorporation of arguments:
  - subjects: *kvindesvømning* [women swimming]
  - objects: *atomspaltning* [atom cleaving]
  - adverbials: *dialysebehandling* [dialysis treatment]
- for compounds, the act/event-condition is applied to the second part, and in case of a conflicting tag for the compound as a whole, the compound is treated as unsafe for automatic frame extraction

Latin loan word derivation: **VERB** -ere --> **NOUN** -ion/ation
- traditional:
  - *adoptere -> adoption* [adoption]
  - *approksimere -> approksimation* [approximation]
- naturalized with -else/-ing, prefixes and argument incorporation:
  - *afnaziﬁcere/-ing* ["**de**nazify"]
  - *detailregulere/-ing* ["regulate **in detail**"]
- -ion/-ation derivation increased coverage, but irregular stemming (e.g. phonetically motivated c/k shift in *kvaliﬁcere - kvaliﬁkation* [qualify - qualification]) makes the method less automatic and more dependent on derivational lexicon entries.

3. Lexicon scheme for nominal frames

From scratch frame entries for predicating nouns without a deverbal morphology:

- frame name + list of possible semantic role arguments

- optional slot filler conditions for each argument
  - primary syntactic conditions (left/genitive position, self and bound preposition lexeme)
  - secondary categorial conditions (semantic class of slot filler)
  - syntactic form conditions, e.g. *icl* (non finite clause), *fcl* (finite clause), with or without a preposition condition

1. *hjælp* (help for/with)
   FN:help / til §**BEN**'all / til §**FIN**'act / fra §**AG**
2. *krav* (demand for/to)
   FN:demand / på §**TH** / om §**ACT** / til §**REC**'H / til §**TP**'all
3. *betaling* (payment to/for)
   FN:pay / af §**REC**'H / af §**CAU**'act / for §**CAU** / til§**REC** / med§**INS**
4. *hensyn* (consideration for)
   FN:adjust / til §BEN
5. *ret* (a right to)
   FN:allow / til §**ASS** / icl §**ACT**

Special case where the noun does not denote the predicating core, but rather one of the arguments, usually the subject. Even without a predicator, such nouns still evoke their frame and will take arguments representing other roles in the frame:

vært (host for/to somebody/an event)
   FN:socializeO / self §**AG** / for,gen §**BEN**'H / for,gen §**EV**'occ

Automatic annotation:

- new module for the DanGram Frametagger
- identifies nominal frames by matching argument slots
- with +HUM genitive dependents, use §AG as a fall-back

4. Free role mapping

As for verbs, some PP dependents of nouns are not valency-bound, but simply free satellites (adjuncts) with a low selection preference for a specific noun frame, e.g. §LOC (location), §LOC-TMP (time).

Consider the following examples from our corpus, all of which contain an §EXT (extension) role complement mediated by the preposition *"på"*, ranging from strongest-bound (1) to weakest-bound (4).

1. *nedskæringer på 750 millioner* [cuts amounting to 750 million] frame: **decrease** (recoverable verb template *skære ned på - cut down*)
2. *håndteringsbeløb på 50 kroner* [a handling fee of 50 crowns] compound, frame: **cost** (second part noun [beløb/amount] has a frame in the lexicon)
3. *fedtindhold på 0,5%* [a fat content of 0.5%], frame: **contain** (frame from second part [indhold/contents], but only loose connection to a degree role, and only triggered by 'fat')
4. *ikke så interessant efter 11 bind på 2 timer* [not so interesting after 11 volumes in 2 hours], implied frame: **read** (elliptic frame, implied only by the reading object 'bind' [volume])

Problem: (1) and (2) can rely on lexicon information, once a frame-carrier is identified, but (3) and (4) need independent role-mapping without identifying a frame first.

Solution: Constraint Grammar role mapping rules
- use semantic noun class (e.g. measurability)
- use modifiers (here: numbers)
- use post-nominal trigger-preposition *'på'* in connection with unit nouns

```
MAP (§EXT) TARGET N-UNIT        (map extension role §EXT on units)
   (p ("på" PRP) LINK p N)        (if parent is 'på', and postnominal)
   ((-1 NUM) OR (-1 NUM-FRACT LINK -1 NUM OR ("en"))) ;
                                  (if preceded by a number or a fraction expression)
```

5. Results

Propbank project for Danish:

- Our annotation scheme & annotator are part of a larger project
- 87,000-token treebank with
  - morphosyntactic tags, semantic ontology for nouns
  - syntactic function tags and dependency links
- subsection of the larger, sentence-randomized Korpus2010 (Asmussen 2015), containing newspapers (15%), magazines (58%), blogs (8.5%), chat fora (2.5%), parliamentary speeches (10.5%) and various internet sources (6%)
- all annotation levels are manually revised
- complete verb frame annotation

Statistical breakdown of the noun frame annotation:

- 9.6% of all nouns (1,342/15,000) identified as frame carriers
- 4477 ad-nominal roles identified
  - 30% linked to, and identified through, a noun frame
  - 70% assigned by free mapping rules
  - role carriers were 50% nouns, 26% clauses (especially relatives) and 13% names
- Clear tendency for some roles to be frame-projected (ACT, RES, CAU, TH, PAT), while others were mostly identified by free mapping rules (ATR, ID,LOC, ORI, EXT)

Tag	role	%all	% frame arg
ATR	attribute	27.11	**4.9**
LOC	location	14.2	**7.2**
TH	theme	9.0	**67.9**
TP	topic	7.0	38.8
ID	idewntigy	6.4	**6.3**
PAT	patient	4.0	**65.5**
AG	agent	3.3	29.5
BEN	beneficiary	3.0	48.1
ORI	origin	2.9	**23.1**
FIN	purpose	2.7	**63.4**
HOL	whole	2.1	**77.4**
ACT	action	2.1	**90.3**
CAU	cause	2.0	**69.7**
EXT	extension	1.9	**26.5**
RES	result	1.5	**80.9**

- 255 different frames found (~50% of the Danish FrameNet inventory)
- Noun frames covered 741 different lexemes
- higher lexeme spread than for verbs (type/token ration 1:2 vs. 1:8), maybe because all verbs are frame carriers, while many frequent nouns are not
- frame statistics with top-3 lexeme realisation (high variation in lexeme spread, red = high, blue = low)

Frame	n	lexemes
be_part	57	del 44, halvdel 5, led 2
investigate	40	undersøgelse 30, forskning 3, analyse 3
run_obj	39	formand 12, leder 5, forvaltning 4
future_having	35	mulighed 32, udbud 2, anvisning 1
decide	31	bestemmelse 12, regel 8, afgørelse 6
discuss	30	debat 9, samtale 4, forhandling 4
relate	29	forhold 9, spørgsmål 6, relation 5
cause	28	årsag 7, grund 6, konsekvens 5
adjust	26	regulering 9, omstilling 7, hensyn 5
explain	25	forklaring 8, eksempel 7, redegørelse 6
create	24	udvikling 9, udmøntning 7, fremstilling 3
allow	24	ret 7, adgang 5, godkendelse 3
tell	23	oplysning 11, historie 4, meddelelse 2
assess	22	vurdering 16, beregning 3
pay	20	råd 4, udgift 3, ressource 3
help	19	grundlag 9, støtte 3, hjælp3

False negatives: Frames
Free role mapping as a shortcut to identification of new frames
- 20% of the ad-nominal roles assigned by free mapping were in fact arguments rather than satellites, warranting a frame tag on their head noun.
- Since certain roles are more likely to be arguments than others, these can be flagged for prioritized inspection, if linked to a frameless head.

False negatives: Roles
- Only 5% of non-transparent nouns (i.e. excluding Danish equivalents of "kind [of]", "lot [of]", "handful" etc.) did not receive a role tag, indicating a good coverage, especially since half of those were simple genitive modifiers of non-deverbal nouns, with a very low chance of being a role carrier.

6. Conclusion and perspectives

- We present a new method for extending Propbank annotation from verbal to nominal frames with a reasonable coverage using
  - verbo-nominal derivation
  - lexical argument-slotfiller information
  - ontology-based role mapping rules
- Our method allows informed prioritization of lexemes and categories for in-depth manual revision.

Danish PropBank:
ELRA-W0117, **ISLRN : 213-212-351-142-5**
Live parses: **http://visl.sdu.dk**
Frame information: **http://framenet.dk**

References

Asmussen, Jørg. 2015. *Corpus Resources & Documentation*. Det Danske Sprog- og Litteraturselskab, http://korpus.dsl.dk

Bick, Eckhard. 2011. A FrameNet for Danish. In: *Proceedings of NODALIDA 2011*, May 11-13, Riga, Latvia. NEALT Proceedings Series, Vol. 11, pp. 34-41. Tartu: Tartu University Library.

Baker, Collin F. Baker; J. Fillmore; J. Charles; John B. Lowe. 1998. The Berkeley FrameNet project. In: *Proceedings of the COLING-ACL*. Montreal, Canada

Fillmore, Charles J. 1968. The case for case. In Bach & Harms (Ed.): *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston. 1-88.

Palmer, Martha; Dan Gildea; Paul Kingsbury. 2005. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31:1., pp. 71-105, March, 2005.

Rehbein, Ines; Josef Ruppenhofer; Caroline Sporleder; Manfred Pinkal. 2012. Adding Nominal Spice to SALSA - Frame-Semantic Annotation of German Nouns and Verbs. *Proceedings of KONVENS 2012*, Vienna. pp. 89-97.

Ruppenhofer, Josef; Michael Ellsworth; Miriam R. L. Petruck; Christopher R. Johnson; Jan Scheffczyk. 2010. FrameNet II: Extended Theory and Practice. http://framenet.icsi.berkeley.edu/