A FrameNet for Danish

Eckhard Bick
University of Southern Denmark
eckhard.bick@mail.dk

Abstract

This paper presents work on a comprehensive FrameNet for Danish (cf. www.framenet.dk), with over 12.000 frames, and an almost complete coverage of Danish verb lemmas. We discuss design principles and frame roles as well as the distinctional use of valency, syntactic function and semantic noun classes. By converting frame distinctors into Constraint Grammar rules, we were able to build a robust frame tagger for running Danish text, using DanGram parses as input. The combined context-informed coverage of the parser-frametagger was 94.3%, with an overall F-score for frame senses of 85.12.

1 The FrameNet concept

Classification of the lexicon is central to many aspects of linguistic research, and modern computational linguistics in particular has a need for robust classification systems to support on the one hand automatic analysis, on the other hand applicational tasks such as information extraction and question answering. As the pivot of the sentence, verbs play a special, integrative role in lexical ontologies. While noun ontologies are relatively easy to build around ISA/hypernym-relations, verbs are somewhat harder to classify because structural aspects are meshed with semantics, with complex combinatorial restrictions residing in both a verb's meaning and its syntactic nature. While one of the largest ontological resources, WordNet (Fellbaum 1998), does cover verbs, but provides little structural-relational information, a number of other classification projects link verb classes to certain verbo-nominal combination patterns, providing information on the form, function and semantics of complements. For English, Levin's original verb classification (Levin 1993) has been expanded in the VerbNet project (Kipper et al. 2006) to include non-np complements and employs 23 (25) thematic roles and 94 semantic predicates. In the FrameNet project (Baker et al. 1998, Johnson & Fillmore 2000, Ruppenhofer et al. 2010), semantic frames like *Commerce* are drawn up with roles like *Buyer, Seller, Goods* and *Money,* which are then associated with verbs (or nouns and adjectives) from corpus examples. Since the same verb may appear in more than one frame, verb sense lists are created implicitly, with no guarantee for full coverage. Conversely, the PropBank Project (Palmer et al. 2005) departs from syntactically annotated corpus data to assign both roles and argument structure to each verb consecutively. Both FrameNet and PropBank provide morphosyntactic restrictions, while FrameNet also adds ontological information on slot fillers.

For Danish, the target language of our own work, some semantic verb classification has been undertaken as part of the Danish DanNet project (Pedersen et al. 2008), covering ca. 3000 verbs with 6000 senses falling into 80 top classes, e.g. BoundedEvent + Physical + Location. However, while some incorporated adverbial material and reflexivity are provided as verb sense discriminators, no frame roles or systematic selection restrictions are listed. Earlier work comprises the STO database, with almost 6000 verbal entries of which 4/5 offer syntactic, and 1/5 semantic information (Braasch & Olsen 2004), and the Odense Valency Dictionary (Schösler & Kirchmeier-Andersen 1997), that classified verbal argument semantics through the semantics of pronoun complements, covering ca. 4000 verbs.

The project described here, launched in 2006, also regards valency as a useful stepping stone towards the semantic classification of verb structures, assuming that almost all subsenses of a given verb can be distinguished, and a full thematic role frame assigned, if the form, function and (noun) semantics of complements are known. Thus, using the DanGram parser's valency dictionary (Bick 2001) as a point of departure, we manually assigned verb classes and thematic role frames to each valency "sense" of a given verb, using corpus data and dictionaries to check sense coverage, and adding sense-based subdistinctions for the broader valency senses where necessary. Syntactic functions and forms of complements were already implicit in the valency tags and could therefore be assigned semi-automatically. At the same time, our methodology of building semantic frames from "syntactic

Bolette Sandford Pedersen, Gunta Nešpore and Inguna Skadiņa (Eds.)
NODALIDA 2011 Conference Proceedings, pp. 34–41

frames" considerably facilitated locating and checking corpus examples, since all syntactic complementation patterns were already available - and searchable - in corpora annotated with the DanGram parser (Bick 2001), allowing focused inspection of semantic variation.

2 The Danish FrameNet

After 4 years, our framenet (inspection demo at www.framenet.dk) has a very good coverage for the DanGram lexicon, and while further senses and patterns are being added and existing ones revised, the overall number of lexemes is now fairly stable, at 6825, with an average of 1.77 frames and 1.46 senses per lexeme. At the time of writing, this corresponds to about 11.000 valency patterns and 12.075 different verb frames, roughly twice the volume of DanNet. We use 494 different verb categories[1] (cp. Appendix 1) that are grouped using the original Levine senses and VerbNet numbering system, albeit with a modified naming system[2] and expanded subclassification system. Thus, though syntactic alternations such as diathesis or word order are not considered frame-distinctors, we do deviate from WordNet and VerbNet by making a class distinction for polarity antonyms like *increase - decrease, like - dislike,* and for the self/other distinction (*move_self, move_other).* We also try to avoid large underspecified classes (e.g. change_of_state), while at the same time keeping the classification scheme as flat as possible, in order to facilitate the use of our categories as corpus annotation tags or Constraint Grammar disambiguation tags. We have therefore introduced classes like *heat - cool, activate - deactivate* or *open - close,* reducing the larger *change_of_state* to a kind of wastebin rest category.

3 Frame role distinctors: valency, syntactic function and semantic classes

The distinctional backbone of our frame inventory are syntactic valency frames like <vt> (monotransitive), <vdt> (ditransitive), <på^vtp-ind> (prepositional ditransitive with the preposition "på" and a verb-incorporated 'ind'-adverb). Each of these valency frames is assigned at least one (or more[3]) verb senses, each with its own semantic frame. Depending, for instance, on the number of obligatory arguments, several valency or semantic frames may share the same verb sense, but two different verb senses will almost always differ in at least one syntactic or semantic aspect of their argument frame - guaranteeing that all senses can in principle be disambiguated exploiting a parser's argument tags and dependency links.

For each of our 12.000 verb sense frames, we provide a list of arguments with the following information:

1. Thematic role (Table 1)
2. Syntactic function (Table 2)
3. Morphosyntactic form (Table 4)
4. for np's, a list of typical semantic prototypes to fill the slot (Table 3)
5. An English language gloss / skeleton sentence

For about half the frames (46%), a best-guess link to a DanNet verb sense is also provided, based on semi-automatic matches on adverb incorporation and hypernym classification.

Our FrameNet uses 38 thematic roles (or case/semantic roles, Fillmore 1968), leaving out adverbial roles that never occur as valency-bound elements in a frame, but only in free adverbials (such as §COND for conditional subclauses). The 38 roles are far from evenly distributed in running text. Table 1 provides some live corpus data, showing that the top 5 roles account for 2/3 of all role taggings in running text.

	Thematic Role	**in corpus**
§TH	Theme	31.75%
§AG	Agent	12.25%
§ATR	Attribute	12.25%
§PAT	Patient	5.12%
§COG	Cognizer	4.69%
§SP	Speaker	4.15%
§RES	Result	3.78%
§LOC	Location	2.95%
§DES	Destination	2.86%

[1] A smaller set of 200 frame senses was also established, with a hypernym-mapping from the more fine-grained set, in part to allow some generalisation when used in e.g. syntactic disambiguation rules, in part to facilitate robust cross-language comparison - and possibly transfer - of frame types.

[2] We wanted the class names to on the one hand be real verbs, on the other to reflect hypernym meanings wherever possible. Therefore, we avoided both example-based names (common in VerbNet) and - mostly - abstrac concept names (commen in FrameNet) that are not verbs themselves.

[3] In 780 cases multiple verb senses share the same valency frame - in other words, in 6.5% of cases, verb senses cannot be disambiguated on syntactic function and form alone, but need help from semantic (noun) classes.

§ACT	Action	2.19%
§REC	Recipient	1.75%
§BEN	Beneficiary	1.65%
§EV	Event	1.56%
§EXP	Experiencer	1.31%

Table 1: Thematic Roles

Other roles: §STI - Stimulus; §REFL - Reflexive; §DON - Donor; §PATH - Path; §ORI - Origin; §EXT - Extension, §VAL - Value; §EXT-TMP - Duration; §MES - Message, §TP - Topic; §SOA - State of Affairs; §CAU - Cause; §ROLE - Role; §INS - Instrument, §MNR - Manner; §FIN - Purpose; §COMP - Comparison; §HOL - Whole, §PART - Part; §POSS Possessor, §ASS - Asset; §CONT - Content; §COM - Co-role; §INC - Incorporated

Even in a case-poor language like Danish, we found some clear likelihood relations between thematic roles and syntactic functions (table 2). Thus, agents (§AG, §COG, §SP) are typical subject roles, while patients (§PAT), actions (§ACT) and results (§RES) are typical direct object roles, and recipients (§REC) and beneficiaries (§BEN) call for dative object function.

	Function	**most likely role**
@SUBJ	Subject	TH > AG > COG > SP > PAT > EV > REC > EXP
@S-SUBJ	Situative subject	TH-NIL
@F-SUBJ	Formal subject	TH-NIL
@ACC	Accusative object	TH > PAT > ACT > RES > REFL > BEN > EV
@DAT	Dative object	REC > BEN > EXP
@PIV	Prepositional obj.	TH > DES > LOC > TP > AG > RES
@...-REFL	Reflexive	REFL
@SC	Subject complem.	ATR > RES
@OC	Object complem.	RES > ATR
@SA	SC Adverbial	LOC > DES
@OA	OC Adverbial	DES > MNR > LOC
@AUX<	Argument of aux.	-

Table 2: Syntactic Functions

The prototypical frame consists of a full verb and its nominal, adverbial or subclause complements. Like most other languages, however, Danish employs also verb incorporations that are not, in the semantical sense, complements. The simplest kind are adverb incorporates, which we mark in the valency frame, but not in the argument list:

kaste op (vomit) - <vi-op>

slå fra (deactivate) - <vt-fra>

komme ind på (discuss) - <på^vt-ind>

More complicated are support verb constructions, where the semantic weight and - to a certain degree - valency reside in a nominal element, typically a noun that syntactically fills a (direct or prepositional) object slot, but semantically orchestrates the other complements. While adverb incorporates are marked as such already at the syntactic level (@MV<)[4], noun incorporates receive an ordinary syntactic tag (@ACC), but are marked with an empty §INC (incorporate) role tag at the semantic level:

holde kæft (shut up) - <vt-kæft>

have brug for (need) - <for^vtp-brug>

One could argue that the real frame arguments (like the noun expressing what is needed in *have brug for*) should be dependency-linked to the §INC noun *brug* and the frame class marked on the latter, but for consistency and processing reasons we decided to center all dependency relations on the support verb in these cases, and also mark the frame name on the verbal element of support constructions.

Pp incorporates are in principle handled in the same way, with a syntactic @PIV tag on the preposition and an §INC role tag on its argument:

træde i kraft (take effect) - <vi-i=kraft>

However, some of these incorporates, especially those containing dative case, which is otherwise extinct in Danish, can be said to be so "frozen" that a preprocessing stage can turn them into one token, assigning an adverb tag to the pp, and allowing the role-free adverb incorporation solution:

have i sinde (intend) - <vt-i=sinde>

være på færde (be going on)- <vi-på=færde>

Independently of the one- or two-token treatment, incorporated pp's are treated alike in our FrameNet dictionary, as '-prp=noun' parts in the valency frame, without a separate argument line, the annotational difference being triggered solely by preprocessing conventions.

4 Frame annotation

One would assume that using argument information from our verb frame lexicon on the one hand and a functional dependency parser on the other,

[4] We are here taking into account the (syntactic) annotation performed by DanGram, the parser used as input for our frame annotation system.

it should in theory be possible to annotate running text with verb senses and frame elements, simply by checking verb-argument dependencies for function and semantic class. To prove this assumption, we implemented our annotation module in the Constraint Grammar formalism, choosing this particular approach in part because that made it easier to exploit the DanGram-parser's existing CG annotation tags, but also to allow for later fine-tuning and contextual exceptions.

As a first step, we wrote a converter program (framenet2cgrules.pl) that turned each frame into a verb sense mapping rule - a relatively simple task, since argument checking amounts to simple LINKed dependency contexts in the CG formalism:

```
SUBSTITUTE (V) (<fn:consist> <r:SUBJ:HOL> <r:PIV:PART/MAT> V) TARGET ("bestå" <mv> V)
(1 (*) LINK *-1 VFIN LINK c @SUBJ LINK 0 <cc>) (c @PIV LINK 0 ("af") LINK c @P< LINK 0 <cc> OR <mat>) ;
```

In the example rule, apart from the <fn:consist> framenet class (implicitly: sense), argument relation tags (<r:....>) are added indicating a HOL role (whole) for the subject and a PART/MAT role (part/material) for the prepositional “af”-object, IF the former is a concrete object (<cc>) and the latter a physical object (<cc> = concrete countable) or a material (<mat>). In the definition section of the grammar, such semantic noun sets are expanded to individual semantic prototype classes (table 3):

```
LIST <cc> = <cc.*>r <cloH.*>r <con> <fruit> <furn> <tool.*>r <V.*>r ; (subtypes, clothing, containers, fruits, furniture, tools, vehicles)
LIST <mat> = <mat> <mat-cloth> <cm-chem> <cm.*>r ; (materials, chemicals, mass nouns)
```

	Semantic (prototype) noun class
<H>	human (<Hprof>, <Hfam>, <Hideo> ....)
<cc>	concrete object
<act>	action
<L>	location (<Lh, Ltop, Lwater, Labs ...)
<fact>	fact
<event>	event
<A>	animal
<sem-r>	“read”-semantical
<sem-s>	“speak”-semantical
<food>	food
<cm-liq>	liquid
<mon>	money
<sit>	situation
<sem-c>	semantic concept
<cm>	substance (concrete mass noun)
<Lsurf>	surface
<V>	vehicle (<Vground>,<Vair> ...)
<conv>	convention
<HH>	group
<an>	anatomical (body part)
.....	(about 200 classes)

Table 3: Semantic prototypes

Apart from semantic classes, the frame mapping rules in step one may exploit word class or phrase type (table 4). With noun phrases being the default, special context conditions will be added for finite or non-finite clausal arguments, adverbs or pronouns.

	Form type	
np	noun phrase or noun phrase in @PIV	
refl	reflexive pronoun	
fcl	finite subclause	
icl	non-finite subclause	
advl	adverb, adverb phrase or adverbial pp	
pl	plural np	
pron	impersonal pronoun (usually 'det')	
adj	adjective	
num	numeral	
pp	prepositional phrase, not in @PIV	
lex	incorporated lexical item	

Table 4: Morphosyntactic Form

For the second step, assigning thematic roles to arguments, we needed to either perform mappings on multiple (argument) contexts, or to target arguments and unify their function with the head verb's new <r:....> tag in order to retrieve (and map) the correct thematic role from the latter. To the best of our knowledge, no current CG compiler allowed either method, so we had to make changes in the compiler code of the open source CG3 variant we were using, for the first time allowing unification between tag-internal string variables and ordinary tag and map sets.

```
MAP KEEPORDER (VSTR:§$1) TARGET @SUBJ (*p V LINK -1 (*) LINK *1 (<r:.*>r) LINK 0 PAS LINK 0 (<r:ACC:\(.*\)>r)) ;
```

The rule above is a simple example, retrieving a thematic role variable from the verb's accusative argument tag (<r:ACC.:..>) and mapping it as a VSTR expression onto the subject in case the verb is in the passive voice. Complete rules will also contain negative contexts (omitted here), for instance ruling out the presence of objects for intransitive valency frames.

While helping to distinguish between verb senses with the same syntactic argument frame, using semantic noun classes as context restrictions raises the issue of circularity in terms of corpus example extraction, and also reduces overall robustness of frame tagging, not least in the presence of metaphor. Therefore, all frame mapping rules are run twice - first with semantic noun class restrictions in place, then - if necessary - without. This way "skeletal-syntactic" (semantics-free) argument structures can still be used as a backup for frame assignment, allowing corpus-based extension of semantic noun class restrictions.

In a vertical, one-word-per-line CG notation, the frame-tagger adds <fn:sense> and <v:valency> tags on verbs, and §ROLE tags on arguments. So far, free adverbial adjuncts are not role-tagged. The example demonstrates a frame sense distinction for the Danish verb *nedsætte.* Dependency arcs are shown as #n->m ID-links.

Nu "nu" <atemp> ADV @ADVL> #5->6
nedsætter "nedsætte" <mv> **<v:vt> <fn:establish>** PR AKT @FS-STA #6->0
regeringen "regering" <HH> N UTR S DEF NOM @<SUBJ **§AG** #7->6
en "en" ART UTR S IDF @>N #8->9
kommission "kommission" <HH> N UTR S IDF NOM @<ACC **§RES** #9->6
der skal undersøge, hvordan ...

(Literally: *Now establishes government-the a commission that shall investigate how ...)*

I Odenses Vollsmose er det først og fremmest miljøets manglende anseelse,
der "der" <clb> <rel> INDP nG nN NOM @SUBJ> **§AG** #12->13
nedsætter "nedsætte" <cjt-head> <mv> **<v:vt> <fn:decrease>** V PR AKT @FS-<SUBJ #13->5
forventningerne "forventning" <f-psych> N UTR P DEF NOM @<ACC **§PAT** #14->13
og "og" <co-fin> KC @CO #15->13
øger "øge" <nosubj> <cjt> <mv> **<v:vt> <fn:increase>** PR AKT @FS-<SUBJ #16->13
problemerne "problem" <ac> N NEU P DEF NOM @<ACC **§PAT** #17->16

(Literally: *In Odense's Vollsmose is it first of all the environment's lacking standing,that decreases expectations-the and increases problems-the.)*

N=noun, V=verb, ADV=adverb, INDP=independent pronoun, ART=article, KC=coordinating conjunction, @SUBJ=subject, @ACC=accusative object, @ADVL=adverbial, @CO=coordinator, @>N prenominal, @FS=finite clause, @STA=statement, §AG=agent, §PAT=patient, §RES=result

5 Evaluation

To evaluate the coverage and precision of our frame tagger, we annotated a 2.4 million word chunk of newspaper text from the Danish daily *Information,* building on a DanGram dependency annotation (Bick 2005) as input, and using only the rules automatically created by our FrameNet conversion program, with no manual rule changes, rule ordering or additions.

Out of 289.720 main verbs, 98.8% were assigned a frame verb sense, albeit 19.2% of assignments were default senses for the verb in question because of the lack of surface arguments to match for sense-disambiguation. 15.0% of frames were subject-less infinitive and participle constructions, but of these, two thirds (10.9%) did have other, non-subject arguments to support frame assignment. The corpus contained 4051 verb lexeme types, and the frame tagger assigned 9195 different frame types, and 5929 verb sense types. Type-wise, this amounts to 2.26 frames, and 1.46 senses per verb type (similar to the distribution in the frame lexicon itself), but token-wise ambiguity is about double that figure, as we will discuss later in this chapter.

	frame slots	expressed surface arguments with frame roles	percentage of filled slots
SUBJ	176831	90981	51.45%
ACC	92610	71336	77.03%
DAT	806	433	53.72%
PIV	22718	22542	99.23%
SC	15120	15120	100.00%
OC	432	432	100.00%
SA	6024	6024	100.00%
OA	191	191	100.00%
ADVL	92	92	100.00%

Table 5: Surface expression of arguments

Table 5 contains a break-down of surface expression percentages for individual argument types. Apart from subjects in non-finite clauses, dative objects are the least obligatory category. Predicative arguments, of verbs like *være (be), blive (become),* are 100% expressed, and prepositional arguments (PIV) have almost as high an expression rate simply because most verbs have alternative valency frames of lower order (intransitive or monotransitive accusative) that the tagger would have chosen in the absence of a PIV argument. In other words, PIV arguments are strong sense markers, and there absence will sooner lead to false-positive senses of lower valency-order than to PIV-senses without surface PIV.

On a random 5000-word chunk of the frame-annotated data, a complete error count was performed for all verbs. All in all, there were 566 main verb tags, 4 of which (0.7%) had been wrongly verb-tagged by the parser, in one case due to a spelling error. For 3 verbs (0.5%), the parser offered a wrong (same-form) lemma. Our frame tagger assigned 561 frames, missing out on 3 regular verbs, and (wrongly) tagging 2 of the false-positive verbs. Only 1 verb was not covered by the frame lexicon, suggesting a very good raw coverage (99.82%). In 15.7% of cases, the frame tagger assigned a default frame, usually a low-order valency frame without incorporates[5]. Of 562 possible frames, 478 were correctly tagged, yielding the following correctness figures:

	Recall	Precision	F-score
total	85.05%	85.20%	85,12
ignoring parse errors	85.51%	86.91%	86,20

Table 6: Recall and precision

These figures are an encouraging result, despite the "weak" (inspection-based) evaluation method. No other frame-/role-tagger could be found for Danish, but Shi & Mihalcea (2004), also using FrameNet-derived rules, report an F-score of 74.5% for English, while Gildea & Jurafsky (2002), using statistical methods, report F-scores of 80.4% and 82.1% for frame roles and abstract thematic roles, respectively. For copula and support verb constructions, not included in the earlier evaluations, Johansson & Nugues (2006) report tagging accuracies for English of 71-73%, respectively, but a comparison is hard to make, since we only looked at support constructions that our FrameNet does know, with no idea about the theoretical lexical "coverage ceiling".

A break-down of error types revealed that 39% of all false positive errors (but only 5.7% of all frames) were cases where the human "gold sense" was not on the list of possible senses in the framenet database. 11 false positives (13.3%) were caused by errors from the parsing stage (wrong lemma, auxiliary or syntactic tag). Ignoring these errors, i.e. assuming correct parsing input, would influence precision, in particular, and raise the overall F-score by 1 percentage point. As one might expect, default mappings accounted for a higher percentage (24.7%) among error verbs than in the chunk as a whole (15.7%), and contributed to almost a third of the "frame-not-in-lexicon" cases.

Frequent verbs have a high sense ambiguity, and verbs with a high sense ambiguity were more error-prone than one-sense verbs, as can be seen from table 7. Thus, the verbs occurring in our evaluation chunk had 4.21 potential senses per verb (6.77 for the ambiguous ones), and the verbs accounting for frame tagging errors had a theoretical 10.08 senses each.

	count	theoretical sense count	senses / verb	sense count in chunk (as tagged)
framenet lexicon	6825	9933	1.46	-
verb types in chunk	243	1022	4.21	275
sense ambiguous	135	914	6.77	167
frame error verbs	40	403	10.08	51

Table 7: Sense ambiguity per verb

6 Conclusion and future work

We have reported work on a comprehensive framenet for Danish, with over 12.000 frames, and a lexeme coverage of almost 100%. After conversion of our framenet into CG rules, the combined parser-frametagger coverage was 94.3% (i.e. only 5.7% match-less default mappings), with an overall F-score for frame senses of 85.12.

Still, given the fact that almost 40% of frame tagging errors were due to missing frame senses, the current framenet should be checked against larger amounts of corpus data to identify senses not captured by our valency-based approach. In particular, noun-incorporations (e.g. *finde sted - take place*) may require further research, since the original DanGram valency lexicon only treated adverb incorporations, and all other incorporations were added in a piecemeal fashion.

On the frametagger side, our CG conversion approach should allow improvements by manually ordering or modifying frame-derived map-

[5] The default frame is not currently based on statistics, but decided upon when converting the framenet lexicon into a Constraing Grammar, as the first intransitive or monotransitive valency frame by order of appearance in the lexicon. Ultimately, therefore, the default choice is under the control of the lexicographer, who can change the frame order in the lexicon.

ping rules, adding more complex context conditions where necessary. Finally, to confirm our intuition as to the effectiveness of the CG conversion approach, it should be compared to a scoring method where frame conditions are matched and counted individually against the parse tree. With either method, the Danish FrameNet could be used to annotate large corpora for manual revision, ultimately allowing hybridization with a statistical frame tagger.

References

Bick, Eckhard. 2001. En Constraint Grammar Parser for Dansk. in Peter Widell & Mette Kunøe (eds.): *8. Møde om Udforskningen af Dansk Sprog,* 12.-13. oktober 2000, pp. 40-50, Århus University

Bick, Eckhard. 2005. Turning Constraint Grammar Data into Running Dependency Treebanks. In: Civit, Montserrat & Kübler, Sandra & Martí, Ma. Antònia (red.), *Proceedings of TLT 2005*, Barcelona, December 9th - 10th, 2005, pp.19-27

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

Braasch, Anna & Sussi Olsen. 2004. STO: A Danish Lexicon Resource - Ready for Applications. In: Fourth International Conference on Language Resources and Evaluation, Proceedings, Vol. IV. Lisbon, pp. 1079-1082.

Fellbaum, Christiane (ed.). 1998. *WordNet: An Eletronic Lexical Database*. Language, Speech and Communications. MIT Press: Cambridge, Massachusetts.

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

Gildea, D. and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles, Computational Linguistics, 28(3) 245-288.

Johansson, Richard & Pierre Nugues. 2006. Automatic Annotation for All Semantic Layers in FrameNet. *Proceedings of EACL 2006*. Trento, Italy.

Johnson, Christopher R. & Charles J. Fillmore. 2000. The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics* (ANLP-NAACL 2000), April 29-May 4, 2000, Seattle WA, pp. 56-62.

Kipper, Karin & Anna Korhonen, Neville Ryant, and Martha Palmer. 2006. Extensive Classifications of English verbs. *Proceedings of the 12th EURALEX International Congress.* Turin, Italy. September, 2006.

Levin, Beth. 1993. *English Verb Classes and Alternation, A Preliminary Investigation*. The University of Chicago Press.

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.

Pedersen, B.S., S. Nimb & L. Trap-Jensen. 2008. DanNet: udvikling og anvendelse af det danske wordnet. In: *Nordiske Studier i leksikografi Vol. 9, Skrifter published by Nordisk Forening for Leksikografi, pp. 353-370*

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/index.php?option=com_wrapper&Itemid=126)

Schøsler, Lene & Sabine Kirchmeier-Andersen (eds.). 1997. The Pronominal Approach Applied to Danish. Studies in Valency II. Rask Supplement Vol. 5. Odense University Press.

Shi, Lei & Rada Mihalcea. 2004. Open Text Semantic Parsing Using FrameNet and WordNet. In HLT-NAACL 2004, Demonstration Papers. pp. 19-22

Appendix 1 - verb categories

Groups	Verbal Classes (494)
Aux and simple construction verbs (30)	(1) be_copula, be_place, consist, be_name, be_part, be_like, be_attribute, be_valid, abound, lack_itr, become, (2) become_be, become_part, get_part, (3) do, work, work_as, work_for, function, do_leisure, take_action, resist, train, (5) have, have_attr, have_part, lack, contain_have (6) must, (7) can
Puttning (22)	(9) put, put_deposit, put_spatial, funnel, raise, lower, flow, pour, spread, coil, uncoil, spray, heap, cover_ize, pollute, fill, uncover_ize, cover, uncover, adorn, confine, park
Removing (16)	(10) remove, exclude, come_off, banish, empty, wipe, clean suck, steal, rid, cheat, exonerate, peel, mine, unhire, resign, renounce
Taking and Bringing (9)	(11) transfer, send, moveO, take, bring, carry, transport, (12) pull, push
Giving and Getting (20)	(13) give, sell, accrue_to, contribute, salary, future_having, supply, equip, man, burden, buy, gain, obtain, employ, get, lose, cause_gain, exchange, trade, berry
Handling (20)	(14) lean, study, get_to_know, forget, check_if, read, (15) hold, grasp, keep, handle, (16) hide, (17) throw, pelt, discard, (18) hit, beat, hit_goal, hurt, spank, bump

Manipulating Entities (79)	(19) poke, (20) touch, touch_exp, (21) cut, crush, perforate, prune, (22) combine, add, absorb, connect, integrate, associate, contrast, link_soc, register, exempt, scramble, group, bond, fasten, cling, (23) separate, divide, split, unattach, differ, (24) colouring, lighting, (25) mark, write, note, label, transcribe, imitate, (26) make, grow, breed, cultivate, create_food, prepare_food, prepare, create, create_finish, create_semantic, shape, deflect, turn_into, convert, modify, perform, rehearse, adjust, process, (27) cause, interact, implement, (28) spawn, (29) appoint, predestine, characterize, portray, name, declare, declare_oc, proclaim, assume, predict, behave, role_as, role_oc, role_sc, now, think, regard_as, remember, classify, dicide, choose
Percepting and emoting (20)	(30) see, hear, sense, undergo, notice, watch, listen, percep, stimulus_subj, (31) affect_exp, emote_obj, like, dislike, obey, disobey, emote, suffer, marvel, attract, repel
Wanting (14)	(32) wish, prefer_to, prefer_oc, long, (33) judge, accuse, praise, speak_affect, analyze, (35) hunt, capture, search, investigate, rummage
Speaking and meeting (33)	(36) socialize, socializeO, play, encounter, fight, dispute, (37) explain, quote, dedicate, inquire, interrogate, teach, tell, identify, speak_mnr, speak_tool, talk, discuss, say, suggest, hint, answer, refuse, advertise, lie, speak_emot, advise, concede, elaborate, emphasize, promise, invoke, reveal
Body (32)	(38) sound_biocom, eat, drink, booze, chew, swallow, dine, thrive, feed, digest, (40) sound_body, excrete, breathe, show_emot, gesture, body_moveO, politing, bodystate, body_moveSA, die, pain, hurt_self, change_bodystate, (41) body_care. comb, dress, undress, serve, (42) kill, kill_method, subjugate
Emanating (8)	(43) light_emission, sound_emission, make_noise, smell_emission, substance_emission, reflect, burn, emit
Changing (27)	(44) destroy, collapse, (45) break, deform, heat, cool, alter, activate, deactivate, open, close, improve, worsen, tighten, loosen, change_process, decay, increase, decrease, oscillate, double, changeS, calibrate, repair, therapy, solve, damage
Moving and Placing (35)	(46) lodge, enter, invade, usurp, permeate, (47) exist, persist, endure, depend, moveS_fluidic, moveO_fluidic, moveO_lokal, moveS_local, orient, sound_move, swarm, collect, accumulate, gather, bulge, spatial_conf, shape_change, meander, border, cross, stretch, (48) appear, show, originate, result, disappear, occur, befall, occur_dynamic, (49) body_moveSC
Mdes of Movement (19)	(50) change_body_pos, body_pos, (51) move_dir, rise, fall, leave, roll, run, vehicle, steer, dance, chase, accompany, reach, (52) avoid, (53) linger, delay, rush
Measuring (7)	(54) measure_tr, measure_itr, cost, contain_quant, fit, assess, bill
Starting, stopping and ongoing (14)	(55) start, begin, start_movement, complete_process, continue, stop, end, hinder, halt, establish, unestablish, run_obj, sustain, (57) weather
Influencing (13)	(58) urge, beg, (59) force, (60) order, demand, summon, (61) try_to, test, (62) plan, (63) enforce, (64) allow, welcome, (65) facilitate
Social interaction (24)	(66) consume, economize, (67) forbid, (68) pay, (69) refrain, (70) rely, (71) conspire, (72) help, benefit, detriment, affect, punish, (73) cooperate, participate, vicariate, (74) succeed, fail, (75) neglect, (76) limit, (77) approve, reject, (78) indicate, confirm, (79) devote
Handling conflicts (19)	(80) liberate, (82) withdraw, (83) cope, (84) discover, (85) defend_phys, defend_cog, attack, (86) correlate, relate, compensate, match, (87) focus, comprehend, (88) mind, (89) agree, (90) exceed, vanquish, exaggerate, (91) matter
Rest - ressource allocation, complex operations (23)	(92) institutionalize, (93) adopt, (94) risk, (95) surrender, (96) accustom, (97) base, deduce, (98) confront, (99) ensure, insure, (100) own, belong_to, (101) patent, (102) promote, (102) require, (104) spend_time, (105) use, serve_as, serve_to, (106) void, (107) include, involve, (108) math

ISSN 1736-6305 Vol. 11
http://hdl.handle.net/10062/16955