

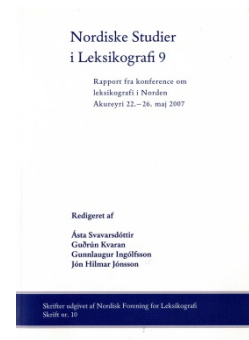
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Verber med fælles syntaktiske konstruktioner og semantisk slægtskab

Forfatter: Anna Braasch

Kilde: Nordiska Studier i Lexikografi 9, 2008, s. 43-60
Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007

URL: <http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive>



© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

ANNA BRAASCH

Verber med fælles syntaktiske konstruktioner og semantisk slægtskab

This paper deals with the issue of observable relationships between the syntactic behaviour and semantics of selected verb groups. The aim is to investigate the feasibility of exploiting syntactic similarity for effective uncovering of semantic kinship. The investigation is inspired by the demand for large-scale lexical resources combining morphological, syntactic and semantic descriptions on the one hand, and on the other by the fact that semantic networks, like WordNet usually do not at all, or only to a very limited degree provide syntactic information. Both aspects show that there is a need for effective means that combine syntactic and semantic description methods and resources containing the two information types.

The pilot project presented here discusses verbs sharing selected syntactic descriptions; they are extracted from the Danish Lexicon for Language Technology Applications (STO) and grouped together according to semantic labels, which in the first run are intuitively selected. The idea is to derive underlying semantic information systematically from the syntactic surface structures described. Although this approach starts from the 'wrong side' compared to other, well-known analyzes of the semantics-syntax relationship, it might be exploited in prediction of verb senses.

1. Indledning

Denne artikel præsenterer en undersøgelse af i hvilket omfang man kan observere nogle fælles semantiske træk ved ord der har identiske eller stærkt lignende syntaktiske konstruktionsmønstre. Formålet er at undersøge om man kan berige ords beskrivelser i en leksikalsk resurse med betydningsrelaterede oplysninger ved at udnytte de eksisterende syntaktiske oplysninger. Metoden sigter på at finde frem til grupper af ord med fælles betydningsindhold for derefter at kunne beskrive dem gruppevis vha. leksikalsk-semantiske skabeloner.

Udgangspunktet for denne tilgang er en stor dansk leksikalsk resurse for sprogteknologiske anvendelser, den SprogTeknologiske Ordbase (herefter STO), der indeholder udførlige, formaliserede morfologiske oplysninger om mere end 81.000 lemmaer og syntaktiske beskrivelser af mere end 45.000 ord (jf. Braasch & Olsen, 2004). En mindre del af lemmaerne har også leksikalsk-semantisk beskrivelse, i alt ca. 10.000 betydninger – heraf ca. 1.000 verbbetydninger.

Semantikdelen er oprindeligt udarbejdet i SIMPLE-projektet (jf. Pedersen & Paggio, 2006) og er baseret på SIMPLE-ontologien, der er et flerdimensionelt semantisk beskrivelsessystem med 135 ontologiske overbegreber. Det er disse

overbegreber som de enkelte ords betydninger er relateret til. Grunden til at væsentlig færre opslagsord er forsynet med semantiske beskrivelser end det er tilfældet mht. morfologi og syntaks, er, at arbejdet med semantikken var særdeles tidskrævende og dermed omkostningstungt, derfor var der i første omgang fokus på en central delmængde af ordforrådet.

Ordbasen er i anvendelse i flere projekter, og brugerne har fortalt os at de meget gerne vil have semantiske oplysninger sammen med den omfattende syntaks. Da vi så præsenterede dem for den meget komplekse SIMPLE-baserede semantik, syntes de at selv om den er rigtig interessant fra en forskningsmæssig synsvinkel, virker den meget omfattende i forhold til mulige sprogteknologiske anvendelser. Der blev udtrykt behov for noget "mere enkelt", fx grundlæggende, formaliserede semantiske beskrivelser af verber/verbgrupper, der skulle være anvendelig i praktiske applikationer. I sidste ende stiller disse anvendelser nemlig nogle særlige krav til hvilke betydningsrelaterede oplysninger der skal beskrives – og hvordan.

Det blev den direkte anledning til at se på hvilke muligheder der er for at udbygge STO med oplysninger om ordenes betydning(er). I denne forbindelse overvejes også hvilken type semantik der bør vælges, hvilken ordklasse der var mest relevant at starte med, og hvordan man kan etablere forbindelse mellem STOs syntaktiske beskrivelser og den valgte type semantiske beskrivelse af ord.

Formålet med den her præsenterede undersøgelse er derfor at afprøve hvilke muligheder der er for at tilføje STOs ordforråd formaliserede betydningsbeskrivelser på en praktisk håndterlig og økonomisk måde. Antagelsen var at der med udgangspunkt i de enkelte ords syntaktiske beskrivelser, kan afgrænses grupper af ord der udviser (næsten) identiske egenskaber mht. deres konstruktionsmønstre, og at disse ordgrupper vil have en række fælles betydningskomponenter. Indsigterne mht. arten og graden af sådanne fællesskaber skal siden danne grundlag for arbejdet hen imod en leksikalsk resurse af en betydelig størrelse med integrerede morfologiske, syntaktiske og semantiske beskrivelseslag.

2. Baggrund og inspiration

Emnet er ikke nyt i leksikografiske sammenhænge, men forskningen tager som regel udgangspunkt i ords betydningsrelationer (fx hyponymi, meronymi, troponymi) og undersøger derefter ordenes syntaktiske konstruktionspotentiale med henblik på at beskrive eventuelle sammenhænge. Der findes mange store undersøgelser af relationerne mellem ords semantiske og syntaktiske egenskaber, ikke blot ud fra den leksikologiske/leksikografiske (Levin, 1993), men også andre, fx den datalingvistiske (Fillmore et al., 2003) eller psykolingvistiske synsvinkel (Fellbaum, 1998). Den generelle, fælles erkendelse er at der er en naturlig sammenhæng mellem ords meningsstruktur og dennes syntaktiske udtryk. Fellbaum skriver eksempelvis at "verbs whose LCSs (viz. lexical conceptual structures) are

identical in terms of specific meaning components tend to share syntactic behaviour” (Fellbaum, 1998:95).

Der er forskellige måder at beskrive disse relationer på, i traditionelle almene ordbøger sker det grundlæggende ved en betydningsbaseret inddeling af ord-bogsartikler, ved hver betydning af et ord anføres også det relevante konstruktionsmønster og et eksempel. Anvendelsen af databaseteknologien har imidlertid banet vejen for store systematiske undersøgelser af ordforrådet ud fra forskellige synsvinkler, og enkelte af disse projekter er kommet frem til grundlæggende nye måder for sprogbeskrivelse.

De projekter der inspirerede min undersøgelse, er forskellige mht. deres formål, teoretiske grundlag, praktiske tilgang og beskrivelsesmetode. Disse inspirationskilder er vigtige netop fordi de forholder sig forskelligt til sammenhængen mellem semantik og syntaks (eller omvendt!), nedenfor gives et par eksempler på hvordan de behandler forholdet mellem semantik og syntaks.

Levins semantiske kategorisering og syntaktiske gruppering af engelske verber (*English Verb Classes and Alternations*, 1993) “...examines a large number of semantically based verb classes and shows how syntactic patterns systematically accompany the semantic classification” (Fellbaum op.cit.: 95). Dette værk har ikke blot dannet skole med hensyn til registreringen af systematiske relationer mellem semantik og syntaks, men også udfordret andre forskere til at gå videre med arbejdet og udbygge, forfine systemet (fx Kipper et al. 2006). Som Fellbaum (op.cit:11) konkluderede “Levin has amassed impressive evidence showing that the semantic makeup of verbs is intimately linked with their syntactic behavior”.

De semantiske net, Princeton WordNet, (WN, jf. Fellbaum, 1998) og senere EuroWordNet (EWN, jf. Vossen, 1998) har fået ganske stor udbredelse i Europa idet der er lavet WordNets for en lang række europæiske sprog foruden engelsk (startende med udviklingen af Princeton WordNet), spansk, italiensk, tysk, ungarsk, tjekkisk osv., men også uden for Europa og USA udarbejdes der WN’s. De fleste semantiske net af WN-typen har kun meget få – hvis overhovedet nogen – syntaktiske oplysninger med. Dermed mangler de en vigtig parameter.

Det forholder sig noget anderledes med FrameNet (Fillmore et al., 2003), et semantisk net for engelsk, hvis grundlæggende ide omfatter en afbildning af ords indholdsside (semantiske argumentstruktur) og udtryksside (deres syntaktiske komplement) i de såkaldte “semantic frames”. Det er en konceptuel struktur der beskriver fx kommunikation og specificerer hvilke semantiske roller der er relevante i forhold til kommunikationen som handling, og hvordan de realiseres i tekster.

Det er derfor en oplagt ide at berige STO’s syntaksmateriale med semantiske beskrivelser og inddrage de erkendelser der er kommet ud af dette pilotprojekt mht. de relationer der kan betegnes som “syntaktisk lighed – semantisk nærhed”.

Sidst, men ikke mindst bør det nævnes det danske DanNet-projekt (se artik-

len af Bolette S. Pedersen, Sanni Nimb og Lars Trap-Jensen herom i dette bind). Projektet er i gang med at opbygge et semantisk net for dansk efter WordNet-modellen, under inddragelse af Den Danske Ordbogs materiale og SIMPLE/STO's formelle beskrivelser. På Center for Sprogteknologi, Københavns Universitet er der i 2007 nedsat en arbejdsgruppe til at undersøge mulighederne for at sammenkoble DanNets semantiske repræsentationer med STOs syntaktiske beskrivelser. En sådan leksikalsk resurse vil kunne imødekomme sprogteknologiens behov for en ordbase der dækker både syntaks og semantik (samt morfologi). Resultaterne af denne pilotundersøgelse vil så vidt muligt blive inddraget i arbejdsgruppens overvejelser og anbefalinger der vil ligge klar sidst på efteråret i 2007.

3. Arbejdsgrundlaget

Denne undersøgelse fokuserer på verber af følgende grunde. For det første er verbers syntaktiske konstruktionsmønstre ganske komplekse: De er en kombination af mange lingvistiske egenskaber, og dette forhold afspejles godt af STOs finmaskede syntaktiske beskrivelser. Derfor er denne ordklasse særlig interessant. For det andet er der et særligt forskningsmæssigt perspektiv i at arbejde med en formaliseret beskrivelse af verbers semantik set i sammenhæng med deres syntaks, da det er et aktuelt emne i forskellige projekter der vedrører andre sprog end dansk (jf. eksempelvis Kipper et al.)

3.1 Den syntaktiske beskrivelse i STO

VALENSTYPE (aritet)	Antal KONSTRUKTIONSMØNSTRE (ledtype og -funktion, kontrol, styret præposition...)	Antal SYNTAKTISKE ENHEDER (verballæsninger)	Antal SYNT. ENHED pr. MØNSTER (gennemsnit)
Dv0 (zerovalent)	4	47	~11,8
Dv1 (monovalent)	95	1.164	~12,2
Dv2 (divalent)	479	5.722	~12,5
Dv3 (trivalent)	304	1.573	~5,17
Dv4 (tetravalent)	6	52	~8,66
Total	888	8.558	~9,65

Tabel 1. Oversigt over ca. 5.770 verbers syntaktiske beskrivelse og fordeling mht. deres konstruktionspotentiale

Arbejdsgrundlaget er de formaliserede beskrivelser af verbers syntaktiske opførelse i STO (se fx Braasch, 2006). STO er den største danske leksikalske database udarbejdet for sprogteknologiske anvendelser – med detaljerede syntaktiske beskrivelser for en relevant mængde af det danske ordforråd. De syntaktiske

beskrivelser er valensbaserede (tager udgangspunkt i ords syntaktiske komplementstruktur) med en del lingvistisk forskning som grundlag. Den syntaktiske komplementstruktur er sproglig realisering af ordets semantiske argumentstruktur, derfor er der en nær sammenhæng mellem et ords betydning (leksikalske semantik) og dets prototypiske syntaktiske omgivelser (syntaktiske konstruktionspotentiale/valensmønstre).

Tabel 1 giver en oversigt over verbernes syntaktiske beskrivelse i STO mht. antallet af konstruktionstyper og verbernes syntaktiske læsninger, fordelt på de fem valenstyper (ariteter). Det fremgår af oversigten at de 5.770 verballemmer har i alt 8.558 syntaktiske beskrivelser, hvilket afspejler at et verbum naturligvis godt kan have mere end et konstruktionsmønster. Der er defineret i alt 888 forskellige konstruktionsmønstre for verber ud fra STOs valensbaserede formelle beskrivelsessystem. Dette ganske omfattende materiale lægger op til den valgte metode, nemlig at nærme sig verbernes semantiske indhold fra deres syntaktiske beskrivelse ved at undersøge verber med identiske eller stærkt lignende syntaktiske egenskaber ud fra tre grundspørgsmål:

- Har de betydningsmæssige ligheder (“semantiske fællesnævner”), hvis ja, hvilke?
- Kan man ud fra de fundne ligheder opstille nogle grundlæggende betydningsgrupper?
- Kan man tildele de enkelte grupper nogle generaliserende semantiske etiketter?

3.2 Empirisk materiale

De syntaktiske oplysninger om de enkelte ord er baseret på ordets korpusforekomster og afspejler de almindelige syntaktiske forbindelser som ordet indgår i, dvs. deres autentiske brug. Naturligvis belyser ordets kontekst i korpusforekomsterne også dets semantiske egenskaber, såsom argumentstruktur, selektionsrestriktioner og betydning. Ved den korpusbaserede udarbejdelse af de syntaktiske beskrivelser er der også taget et vist hensyn til de semantiske læsninger af ordet.

3.3 Valensbaseret tilgang til syntaks

Den lingvistiske tilgang i STO bygger på valensteorien, dvs. ords *syntaktiske konstruktionspotentiale* er beskrevet med fokus på kategoriel og funktionel valens. Ords *semantiske argumentstruktur* bestemmer dets syntaktiske valens, idet komplementerne er argumenternes realisering i ytringer. I den valensbaserede tilgang beskrives syntaksen med udgangspunkt i ordets komplementstruktur og semantikken ud fra dets argumentstruktur (Grimshaw 1990, Levin 1993). Argumentstrukturen realiseres som ordets konstruktionspotentiale og er dermed bindeled mellem dets betydning og syntaks. Eksempelvis beskriver de fleste zerovalente verber vejrfænomener (*regne, tordne, sne...*); forholdet mellem syntaks og seman-

tik er tydeligt om end ikke entydigt, jf. undersøgelsen der refereres i Braasch (2006). Det forholder sig anderledes i tilfælde af verber med mere kompleks syntaks, fx konstruktioner med flere styrede led. Her er der naturligvis mange flere kombinationsmuligheder mellem syntaks og semantik. Det gælder fx for trivalente verber (fx *beskyldte/rose/straffe* eller *undersøge/behandle/helbrede...* nogen for noget); disse tager subjekt, objekt og præpositionalobjekt. Det sidstnævnte kan have forskellig syntaktisk form, dvs. et nominalsyntaxme, infinitiv eller en (in)fininit at-sætning. De ca. 1.570 trivalente konstruktioner i STO er beskrevet med 304 forskellige konstruktionsmønstre, disse kan samles i større grupper på grundlag af de enkelte komplementers syntaktiske egenskaber.

Den syntaktiske beskrivelse omfatter antallet og arten af styrede (valensbundne) led, og for hvert styret led angives følgende: *position* i den kanoniske rækkefølge, syntaktisk *kategori*, fx NP, PP (styret præposition og styrelse), at-sætning osv., syntaktisk *funktion*, fx subjekt, direkte, indirekte eller præpositionsobjekt, objektprædikat, infinitivkomplement, osv., og hvorvidt leddet er obligatorisk i konstruktionen eller ej. Andre registrerede egenskaber er refleksivitet, partikelverbum, subjekt/objektkontrol, mv. Hver unik kombination af disse egenskaber udgør et konstruktionsmønster. Egenskaberne formaliseres i STO i såkaldte *attribut/værdi-par* (fx kategori = NP), og deres unikke kombination kaldes *syntaks-mønster*. En sådan formaliseret beskrivelse er ofte en generalisering over flere af verbets konstruktioner, hvor det har den samme grundbetydning.

I eksemplerne nedenfor bruges følgende notation:

[]: omslutter et obligatorisk komplement

(): omslutter et optionelt komplement

{ } : omslutter mulige realisationer af et obligatorisk eller optionelt komplement.

Et verbum kan naturligvis optræde med mere end et mønster, fx i tilfælde af polysemi som i

(1a) *anholde* [+ ngn] + (*for* ngt) i betydningen 'arrest'

(1b) *anholde* [*om* ngt] i betydningen 'apply'

hvor (1a) er en trivalent konstruktion (subjektet er altid til stede, det er derfor underforstået), mens (1b) er divalent.

Et væsentligt spørgsmål i denne sammenhæng er, hvad der egentlig regnes som styret (valensbundet) led. I gængse fremstillinger (fx i tyske valensgrammatikker) anskues valensbegrebet binært, dvs. et led i en given konstruktion er enten valensbundet eller ej, hvorimod Somers (1987:27 & 266) beskriver det ud fra sin datalingvistiske synsvinkel som et kontinuum med seks forskellige grader af "bundethed". I Somers' kontinuum udgør de såkaldte "middles" en kategori der ligger mellem valensbundne led (også kaldet komplement) og de frie led (også kaldet adjunkter). Denne kategori udviser nogle særlige egenskaber, eksempelvis kan det ofte være svært at tildele et sådant led en af de traditionelle semantiske

roller (se fx Trask, 1993: 249), og det udtrykkes ofte med et præpositionssyntagma. Semantisk set er “middles” bestemt af det valensbærende ord (her af verberne *bo* og *gå*), hvorimod deres syntaktiske form eller kategori kan variere, ligesom de kan være obligatoriske eller optionelle, som det fremgår af eksemplerne (2a-2b) nedenfor.

- (2a) Han *bor* her/på gården/ikke ret langt herfra
– det semantiske argument angiver *stedet* (lokativisk), det syntaktiske komplement kan udtrykkes med et præpositionssyntagma eller et adverbial og er obligatorisk.
- (2b) Han *gik* ud på pladsen/ned til bækken/hen til landsbyen
– det semantiske argument angiver retningen (direktionelt), det syntaktiske komplement kan udtrykkes med et præpositionssyntagma eller et adverbial og er optionelt.

Den syntaktiske beskrivelse af verber i STO inkluderer som regel også “middles”, og det er i mange tilfælde en rigtig god hjælp i sprogteknologiske anvendelser – og i denne undersøgelse selv om kodningen desværre ikke er helt konsekvent gennemført.

4. Undersøgelsen

Hvert verbum (samt substantiv og adjektiv) har således mindst et konstruktionsmønster, og hvert konstruktionsmønster er repræsenteret ved et antal verber. Der er mønstre som beskriver mange verbers konstruktionspotentiale som i tilfælde af almindelige transitive verber (*klippe*, *spise*), og der er andre der kun gælder for ganske få, fx de zerovalente vejrligsverber som *sne* og *hagle*. Kombinationen af lemma og et af dets syntaksmønstre udgør en såkaldt *syntaktisk enhed* i STO (disse kaldes også syntaktiske læsninger). Det er sådanne syntaktiske enheder af verber med fælles syntaksmønster der udtrækkes fra databasen, og materialet undersøges mht. relationen mellem deres syntaktiske og semantiske egenskaber. Verbet “tale” har eksempelvis otte syntaktiske enheder, så forskellige som bl.a. i (3a-c):

- (3a) Vi taler (*med dem*) (*om* {alt muligt/*at* komme på besøg})
med subjekt og to optionelle præpositionsobjekter
(STO-kode: Dv3P0Pnw0-med-om)
- (3b) Han taler [{*stille* og *dæmpet*}]
med subjekt og obligatorisk adverbiumssyntagma
(STO-kode: Dv2ADVP)
- (3c) S-SF-politikere talte [*mod* {KV-regeringen}]
med subjekt og et obligatorisk præpositionsobjekt
(STO-kode: Dv2Pnti-mod)

Syntaksmønstrer i (3a) beskriver 20 verber i STO, hvoraf ti er entydigt talehandlingsverber, fx *konferere*, *passiare*, *småsnakke*, *pladre*, og de resterende ti beskriver handlinger der inkluderer en form for kommunikation, ofte ad mundtlig vej, fx *forhandle*, *vadde*, *skændes*, *enes*, *toppes*, *strides*. Her er det altså ganske tydeligt at disse verber også semantisk er stærkt beslægtede, selv om ikke alle er (nær)synonymer. Det er denne type umiddelbare iagttagelser der er efterprøvet i undersøgelsen. I denne sammenhæng er det interessant at bemærke at Fellbaum (2002:8) diskuterer et lignende forhold mellem to beslægtede verbalgrupper. I hendes beskrivelse vil den første gruppe udgøre forskellige realisationer af verbet *tale* (det overordnede verbum) mht. måden ("manner") handlingen udføres på, den defineres derfor som *genuine troponymy*. I den anden gruppe har verberne derimod en klar semantisk komponent af kommunikationens funktion ("function"), og dermed står disse verber i en anden relation til *tale*, det kaldes *function relation*.

Der er desuden mange tilfælde hvor flere verber deler mere end ét konstruktionsmønster, men dette forhold bliver ikke behandlet her. Levins klassifikation af engelske verber tager sit udgangspunkt netop i dette og giver en meget omfattende redegørelse over alternationerne for næsten 4.200 verber.

4.1 Valg af materiale til undersøgelse

Udvælgelsesprocessen omfatter tre trin. Det indledende trin er at udtrække alle syntaktiske konstruktionstyper for verber for at skaffe et overblik over det samlede potentielle materiale. Den talmæssige fordeling af syntaktiske enheder er som vist i tabel 1 (ovenfor): der er fem valenstyper defineret ud fra ariteten, og for hver valenstype er der registreret et antal syntaktiske konstruktionsmønstre der hver især beskriver en unik kombination af konstruktionsegenskaber.

Næste trin var at udvælge en delmængde af en valenstypes konstruktionsmønstre som den egentlige genstand for undersøgelsen. Der er udvalgt en delmængde (otte) af de (i alt 304) trivalente konstruktionsmønstre der beskriver verber med følgende styrede led: subjekt, direkte objekt og præpositionsobjekt (POBJ). Verber med denne konstruktionstype kaldes "ditransitive verbs" (Grimshaw, 1990:28) eller "double object constructions" (Levin, 1993:274). Subjektet og objektet er i de valgte konstruktioner et nominalsyntaxme (NP). Det er præpositionsobjektet der er særlig interessant ud fra forskellige aspekter. Valget af præpositionen er i mange tilfælde (men ikke altid?) bestemt af verbets semantik (i det præsenterede materiale er præpositionen *til* eller *for*). Styrelsen for præpositionen kan have forskellige syntaktiske egenskaber mht. kategori: i nogle konstruktioner kan det kun være nominalsyntaxme, i andre alternerer nominalsyntaxmet med en at-infinitiv (sætning); og sidstnævnte kan optræde med eller uden objektskontrol. Den sidste egenskab vedrører leddets tilstedeværelse i konstruktionen: det kan være optionelt eller obligatorisk. Oplysningen om hvorvidt leddet er obligatorisk eller ej har dog vist sig ikke at være nogen særlig indikator med hensyn til semantiske forskelle i forhold til de her undersøgte verber. I andre tilfælde kan det forholde

sig ganske anderledes, fx "manden taler sandt" er forskellig fra "manden taler". Det kan ofte dreje sig om en stærk valensbinding med strenge selektionsrestriktioner, og konstruktionen nærmer sig til kollokationsstatus. (Et sådant stærkt styret led betegnes hos Somers (1987:27) som "integral complement", denne type konstruktioner falder uden for undersøgelsen.)

Derefter fulgte udtræk af alle verber (216 stk.) med de valgte konstruktioner sorteret i otte lister, svarende til de otte konstruktionsmønstre. Tabel 2 viser en oversigt over fordelingen af verbers syntaktiske enheder på de enkelte konstruktionsmønstre. Disse er nært beslægtede selv om præpositionsobjektets egenskaber varierer noget: det kan være obligatorisk eller optionelt i konstruktionen, den styrede præposition kan være *til* eller *for*, og styrelsen kan være et NP, eller en at-infinitiv eller at-sætning med objektskontrol. Den talmæssige fordeling viser at der er en overvægt af syntaktiske enheder med optionelt præpositionsobjekt med NP uanset præpositionen.

KONSTRUKTIONSMØNSTER (STO-kode)	EKSEMPEL	ANTAL
med POBJ: til + NP, optionelt (<i>Dv3NP0-til</i>)	genopstille ngn (til posten)	56
med POBJ: til + NP, obligatorisk (<i>Dv3NP-til</i>)	udråbe ngn [til vinder]	31
med POBJ: til + NP/at-sætn. med objektskontrol, optionelt (<i>Dv3NPnio0-til</i>)	besnakke ngn (til {et lån/at låne...})	20
med POBJ: til + NP/at-sætn. med objektskontrol, obligatorisk (<i>Dv3NPnio-til</i>)	beordre ngn [til{march/at marchere}]	30
med POBJ: for + NP, optionelt (<i>Dv3NP0-for</i>)	snynde ngn (for penge)	35
med POBJ: for + NP, obligatorisk (<i>Dv3NP-for</i>)	tiltale ngn [for drab]	21
med POBJ: for + NP/at-sætn. med objektskontrol, optionelt (<i>Dv3NPnio0-for</i>)	kritisere ngn (for {sjuskethed/ for at være sjusket})	16
med POBJ: for + NP/at-sætn. med objektskontrol, obligatorisk (<i>Dv3NPnio-for</i>)	anse ngn [for {en god læge/at være en god læge}]	7
TOTAL		216

Tabel 2. Oversigt over de undersøgte trivalente syntaktiske enheder fordelt på konstruktionsmønstre

4.2 Materialets behandling

Herefter fulgte den egentlige behandling af de enkelte lister, den blev udført i tre hovedtrin; disse skitseres nedenfor.

4.2.1 Tildeling af semantisk etikette og gruppering af verber

Det første trin var at gennemse alle lister og tildele en foreløbig *semantisk etikette* til hvert enkelt verbum. Etiketterne er valgt intuitivt, dvs. ikke ud fra et på forhånd defineret system, men ud fra verbets hovedbetydning i den konkrete eksempelsætning med den konstruktion, som syntaksmønstret beskriver. De fungerer som en slags semantiske fællesnævner med udgangspunkt i de åbenbare betydningsmæssige ligheder mellem listens verber. Denne metode er i god overensstemmelse med andres erfaringer, som Fellbaum skriver (1998:70): “Dividing the verb lexicon into semantic domains initially on a purely intuitive basis (min fremhævelse, A.B.) might lead one to discover relations that organize verbs and concepts.”

De tildelte etiketter er overskrifter på engelsk (fx SELECT, ESTIMATE, CHANGE_of <FEATURE>), og der er to grunde til dette valg. For det første er det for at undgå at skulle vælge et dansk overbegreb til verbet allerede fra starten, hvilket kunne fastlåse analysen til et bestemt begrebshierarki, da et sådant hierarki er mere velegnet til substantiver end til beskrivelse af verber. For det andet er formålet at forberede hovedtrin 3, sammenligning af de semantiske grupperinger med eksisterende beskrivelser for engelske verber (Levins og EuroWordNets klassifikationer), som kunne vælges senere som inspirationskilde og arbejdsgrundlag for den semantiske beskrivelse af danske verber. Verberne blev derefter sorteret i foreløbige semantiske grupper (lister) på grundlag af deres fælles semantiske etikette.

SEMANTISK GRUPPE	VERBUM (eksempler)	POBJ (præp + NP)	POBJ (præp + at-inf.)	ANTAL (tentativt)
FORCE	tvinge ngn beordre ngn	til denne handel til march	til at rejse til at lukke alle døre	12
URGE/ INQUIRE	overtale ngn formane ngn	til en kop kaffe til større tolerance	til at melde sig til at holde aftalen	9
ASK/ADVISE/ INSPIRE	råde ngn motivere ngn lokke ngn	til modstand til træning til det tomme hus	til at gøre modstand til at lære noget nyt til at holde vinterferie	8

ELECT	udtage ngn nominere ngn/ngt	til kampen til en Oscar	til at spille mod Polen	14
RELATE	henregne ngt forbinge ngn	til en ... kategori til konflikten	Ø	10
QUANTIFY/ ESTIMATE	estimere ngt tidsfæste ngt	til 3 måneder til foråret 2000	til at ligge over 3 m	7
QUALIFY	uddanne ngn	til et job	til at kunne bestride jobbet	10
PROSECUTE	straffe ngn	for indbruddet	for at have begået indbrud	7
ACCUSE/EXCUSE	beskyldte ngn frikende ngn	for illoyalitet for ansvaret	for at have krænket loven for at have optrådt forkert	5
COMMUNI- CATE_ JUDGE- MENT	klandre ngn rose ngn	for beslutningen for sangen	for at tage fejl for ikke at sylte sagen	8
.....
CHANGE_of POSS. 1	støvsuge ngt	for sand og snavs	Ø	5
CHANGE_of POSS. 2	narre ngn	for et stort beløb	Ø	7
CHANGE_of POSS. 3	donere ngt	til kampagnen	Ø	4
.....
CHANGE_of LOC.	udskibe ngt	til Amerika	Ø	6
CHANGE_of STATE	forarbejde ngt	til gelé	Ø	9
.....

Tabel 3. Eksempler på tildelte semantiske etiketter og verber med angivelse af tentativ fordeling (udsnit)

Tabel 3 viser et udsnit af de mulige semantiske grupperinger, og det illustrerer at samme verbum kan indplaceres under forskellige etiketter – alt efter hvilket semantisk element/træk eller konnotation der fokuseres på, eksempelvis verberne *råde* eller *lokke*. Den kontekst verbet optræder i bidrager naturligvis betydeligt til at præcisere betydningen, fx hvorvidt der er tale om tvang eller falske forudsætninger i forhold til handlingen. (Dette er et generelt problem som bliver taget op i afsnit 5.1.)

4.2.2 Behandling af verbalgrupperne

Det andet trin var behandlingen af de enkelte semantiske grupper. Det viste sig nemlig at tildelingen af semantisk etikette ikke var helt konsistent, hvilket afspejlede fx at de enkelte verbers konnotationer i første omgang blev vægtet forskelligt, men når verberne var sorteret i grupper, blev denne inkonsistens mere tydelig. Der blev udført justeringer mht. tildelingen, og enkelte etiketter viste sig at være for snævre eller for brede med hensyn til deres betydningsomfang, så de blev erstattet med mere dækkende overskrifter. De enkelte gruppers præpositionsobjekter blev derefter gennemset, og det viste sig at de havde en større eller mindre grad af semantisk lighed. Deres prototypiske semantiske type er registreret i et første forsøg, til dette er foreløbig brugt nogle generelle overbegreber fx *genstand*, *person*, *beløb*, *ydelse*. Desuden er de semantiske roller noteret i form af etablerede betegnelser som fx PATIENT, BENEFICIARY, GOAL, MEASURE. (jf. eksempelvis hos Grimshaw, op.cit og Trask, op. cit). CAUSE hører egentlig ikke til de centrale roller, den betragtes som en perifer rolle, denne kan tildeles bl.a. komplementer af typen "middles" (jf. diskussionen i Somers 1987:74 ff.). PATIENT-rolle er her delt op i to undertyper der refererer til forholdet mellem det første objekt i konstruktionen (*nogen* eller *noget*) og præpositionsobjektet (*for* + *nogen* eller *noget*) som i eksemplerne i (4a) og (4b):

- (4a) Postbuddet tømmer *bilen* (OBJ) *for pakker* (POBJ: PATIENT, LOCATED)
- (4b) Selskabstømmeren lænsede *virksomheden* (OBJ) *for aktiebeholdningen* (POBJ: PATIENT, POSSESSED).

GRUPPE m. SEMANTISK ETIKETTE	EKSEMPEL PÅ VERBER	EKSEMPEL PÅ POBJ: for + NP	POBJ's SEM. TYPE	POBJ's SEM. ROLLE
CH_of Location (REMOVE)	tømme, rydde tappe, lænse_1	{pakker, passagerer}, {saft, vand}	genstand, person	PATIENT (LOCATED)
CH_of Possession (CLEAR/ REMOVE)	rippe, lænse_2	{spillere, aktie- beholdningen}	genstand (værdi)	PATIENT (POSSESSED)
CH_of Possession (CHEAT/ FOOL)	afpresse narre, blokke snyde	{penge} {et stort beløb} {en gave}	værdi (beløb) værdi (ydelse)	PATIENT (POSSESSED)
CH_of Possession (OFFER)	skænke, ofre	{gæsterne} {familien, sagen}	person	BENEFICIARY
CH_of Possession (SELL)	sælge videresælge	{sin bror} (=fri dativ) {500 kroner}	person beløb	BENEFICIARY MEASURE

CH_of Posses- sion (DEBET/ CHARGE)	debitere, bon'e	{reparationerne} {400 kroner}	værdi (ydelse, vare) værdi (beløb)	CAUSE MEASURE
--	--------------------	--------------------------------------	--	------------------

Tabel 4. Trivalente verber med POBJ (præposition 'for' + NP) og deres semantiske egenskaber (STO-kode: *Dv3NP-for*) (udsnit)

Tabel 4 viser et lille udsnit af resultaterne for de undersøgte verber med den overordnede semantiske etikette *CHANGE_of_ <FEATURE>* efter at dette trin var afsluttet.

4.2.3 Sammenligning af grupperingen med beskrivelse i to etablerede systemer

I det tredje trin udførtes sammenligninger af materialet med Levins verbalklasser og med EuroWordNets metode for tildeling af de såkaldte *situationskomponenter* og indordning i dets topontologi (EWN, 2002:61) på et overordnet plan. Dette betyder at undersøgelsen ikke går i dybden på dette punkt, fordi formålet er at få et første overblik over om de her forslåede syntaksbaserede semantiske grupperinger er kompatible med beskrivelsessystemer der har en anden tilgang.

Sammenligningen med Levins verbalklasser er udført på den måde at de enkelte semantiske grupper er identificeret i forhold til Levins klassifikation, og de relevante kategorier derfra er noteret for den pågældende gruppe. Det er indlysende at der er betydelige forskelle mellem verbernes syntaks på dansk og engelsk, og derfor er klassifikationen af syntaktiske alternationer også forskellig for de to sprog. Desuden medtager Levin som regel ikke præpositionsobjekter, især ikke med infinitivsætning. Derfor kan det være svært at indplacere enkelte verber med sådanne valensled. Dette problem skyldes formentlig forskellen mellem den binære (hos Levin) og den graduerede anskuelse af valensbundethed (i STO/hos Somers, op.cit.). På den anden side er der et ganske stort sammenfald med hensyn til den semantiske gruppering af verberne, og det er dette forhold undersøgelsen fokuserer på.

EuroWordNets system opererer med andre parametre, beskrivelsen omfatter tre væsentlige oplysningsniveauer, for verber er det "EntityType" (som altid er *2nd Order Entity*), "SituationType" som kan være *Dynamic* (herunder *BoundedEvent*: verber med tidsmæssig afgrænsning og et forventet resultat; jf. eksemplerne i tabel 5) eller *Static*, samt en – eller oftest en kombination af – flere semantiske egenskaber der kaldes "SituationComponents", fx *Communication*, *Location*, *Manner*, *Purpose*, *Quantity* og *Social*. Tildelingen af disse beskrivelseskomponenter er vanskelig, men også meget vigtig i sammenligningen. Vanskeligheden består i at disse semantiske etiketter efter min begrænsede erfaring dels ikke kan tildeles konsistent, dels at der i EWN i princippet kan tildeles et stort antal sådanne etiketter til samme semantiske verballæsning (fx *blow_2*: Bounded Event+Agentive+Physical+Location+Purpose+Manner), og de kan optræde i

en hvilken som helst kombination, idet deres antal og rækkefølge ikke er fastlagt på forhånd (EWN, 2002:65 ff). Dette resulterer i overlapninger og krydsninger i klassifikationen.

Tabel 5 giver et lille indblik i at det godt kan lade sig gøre at indplacere de semantiske grupper i EWN på et ikke for specifikt niveau, samtidig med at de enkelte verbalgrupper også har deres pendant i Levins' klassifikationssystem.

SEMANTISK GRUPPE	LEVIN'S VERB CLASS	EWN-PLACERING BoundedEvent+Agentive+ <SituationComponent>
CH-of Location (REMOVE)	10: Verbs of Removing	+Physical+Location+Possession
CH_of Possession 1 (CLEAR/REMOVE)	10.1+3 Remove, Clear	+Physical+Location+Possession
CH_of Possession 2 (CHEAT/FOOL)	10.5+6 Steal, Cheat Poss. Deprivation	+Possession+Manner
CH_of Possession 3 (OFFER)	13.3 Future having (Ch_ Poss)	+Possession+Social
CH_of Possession 4 (SELL)	13.1 Give (Ch_Poss)	+Purpose+Possession
CH_of Possession 5 (DEBET/CHARGE)	54.5 Bill (Measure)	+Purpose+Possession+Social

Tabel 5. Eksempel på resultater af sammenligningen: Semantiske grupperinger, Levin's klassifikation og EWN's indplacering af udvalgte trivalente verber (*udsnit*)

Resultatet af denne sammenligning skulle give et fingerpeg om hvorvidt den her afprøvede metode kan bidrage til en systematisk og effektiv kodning af semantiske egenskaber i forhold til eksisterende syntaktiske beskrivelse af verber. En anden mulighed der er under overvejelse i skrivende stund, er en semiautomatisk sammenføring af to resurser for dansk: den færdige STO databases indhold og DanNets semantiske beskrivelser. Erfaringerne fra denne undersøgelse vil kunne bidrage dertil fx med en metode for selektionen af syntaktiske enheder. I denne sammenhæng vil en sammenligning mellem de her beskrevne semantiske grupper og DanNets kodninger af verber være en yderligere forudsætning.

5. Relevante iagttagelser vedrørende udnyttelse af STO-materialet

5.1 Generelle forhold

Det er indlysende at de divalente verber udgør den allerstørste gruppe (med 479 forskellige konstruktionsmønstre for 5.722 syntaktiske enheder), og at denne gruppe er meget kompleks mht. semantiske egenskaber. Det er også evident at forholdet mellem syntaks og semantik er mere entydig i tilfælde af zero- og te-

travalente verber (jf. Braasch 2006) fordi disse valensgrupper er mere homogene, dvs. de omfatter kun få konstruktionsmønstre (jf. tabel 1). En anden iagttagelse er at samme syntaktiske komplement (objekttype) kan være syntaktisk manifestation af mange forskellige semantiske argumenter, dvs. der er ikke nogen 1:1 relation på alle planer mellem argument- og komplementstruktur. Dette sætter en naturlig begrænsning for i hvor høj grad den beskrevne metode kan generaliseres.

Det graduerede syn på valensbundethed ved kodningen af "middles" sætter fokus på det problematiske i at tildele præpositionsobjekter semantiske standardroller, de tenderer til at have perifere og mindre entydige roller.

Det viste sig, som tidligere nævnt, i mange tilfælde at være svært at tildele en dækkende semantisk etikette til verber i en intuitivt set semantisk homogen verbalgruppe, fx verberne *indbyde*, *invitere*, *tilskynde*, *motivere*, *formane*, *lokke*, *råde* og lignende (jf. afsnit 4.2.1). Valget stod mellem etiketter som fx URGE, ASK, INQUIRE, ADVISE, INSPIRE, alt efter hvilken betydningskomponent der fokuseres på i kombinationen af denotation/konnotation, fx graden af intensitet i handlingen, eller formålet med handlingen. Et skift i dette fortolkningsfokus kan også medføre en ændring af selve grupperingen. Det er meget mere kompliceret at udnytte hierarkiske og semantiske relationer (i form af troponymi) til at organisere verber i semantiske domæner end for substantiver (ved fx hyponymi/hyperonymi); de semantiske relationer mellem verbers betydninger er mere komplekse.

5.2 Specielle forhold

Den systematiske sammenstilling af verber der er kodet med ens eller stærkt lignende konstruktionsmønstre gav mulighed for et konsistent stykke af kodningerne. Gennemgangen af materialet afslørede enkelte problemer/mangler i forhold til verbernes syntaktiske beskrivelse i STO som kan sammenfattes i følgende punkter:

- (a) Manglende data mht. opslagsord, fx *skose* [nogen] (*for* {noget/at}) eller relevant syntaktisk enhed inden for den undersøgte valensgruppe, fx for verberne *elske* og *hade* [nogen] (*for* {noget/at}).
- (b) Inkonsekvens i kodning mht. hvilke sætningsled der er registreret som valensbundet (fx *anholde* er beskrevet som divalent med eksemplet "Politiet anholdt ham, da færgen...", hvorimod verbet i de fleste korpusforekomster er trivalent med optionelt præpositionsobjekt, der fremgår af konteksten: *anholde* nogen *for* noget/at + infinitiv)
- (c) Inkonsekvens mht. hvilket præpositionsobjekt der er markeret som optionelt, fx *presse* [nogen] [*for* noget] vs. *afpresse* [nogen] (*for* noget)).

Disse iagttagelser er blevet noteret, og rettelser kan udføres ved lejlighed, hvilket bliver en sidegevinst for STO. Der skal naturligvis også tages hånd om disse problemer inden eller under en (semi)automatisk udnyttelse af STO ved at rette inkonsekvenser og evt. tilføje af manglende data.

6. Sammenfatning og perspektivering

De hidtidige erfaringer viser at der, i hvert fald for en del af verbernes vedkommende, kan foretages en sortering og gruppering ud fra verbers valensstruktur, og denne gruppering kan danne grundlag for en systematiseret semantisk beskrivelse. Ifølge Fellbaum (1998:95) "An exploration of the syntactic properties of co-troponyms occasionally provide the basis for distinguishing semantic subgroups of troponyms", pilotundersøgelsens resultater falder godt i tråd med dette.

De her præsenterede semantiske grupper kan revideres/tilpasses til at kunne bruges i forhold til forskellige beskrivelsessystemer, fx et semantisk net for danske verber. Til semantisk kodning kan der opstilles – eller vælges fra eksisterende semantiske klassifikationer – leksikalsk-semantiske kodningsskemaer, profiler eller skabeloner passende for de enkelte grupper og undergrupper. En sådan skabelon skal indeholde (generaliserede, formaliserede) oplysninger om grundlæggende betydningsindhold (fx semantisk type i den valgte klassifikation) og prototypiske semantiske argumenter specificeret med hensyn til deres betydningsmæssige egenskaber (formuleret som selektionsrestriktioner, igen i forhold til en på forhånd fastlagt typologi). Der er forskellige muligheder for valget af semantisk beskrivelsestype og formalisme; det væsentlige er fra sprogteknologiens synsvinkel at den leksikalske ressource ikke på forhånd må være dedikeret til en bestemt applikation, men må kunne anvendes bredt i forskellige sammenhænge, fx semantisk opmærkning af tekster, i systemer til avanceret/indholdsbaseeret informationssøgning, tekstresumering osv. Endvidere kan STO-materialet udnyttes i lingvistisk forskning vedrørende prædiktation om forholdet mellem syntaks og semantik, eksempelvis ved at undersøge verber der har flere syntaktisk identiske konstruktioner for at få overblik over alternationstyper i dansk og sammenligne dem med Levins klassifikation.

Referencer

- Braasch, Anna & Olsen, Sussi (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.
- Braasch, Anna (2006): Exploitation of Syntactic Patterns for Sense Group Identification. In: *Proceedings, XII. EURALEX International Congress*, pp.133-140. Edizioni dell'Orso, Alessandria.
- EWN= EuroWordNet: General Documentation, LE2-4003, LE-8328. Ed.: Piek Vossen.

- Elektronisk dokument (www.vossen.info/docs/2002/EWNGeneral.pdf).
- Fellbaum, Christiane (1998): A Semantic Network of English Verbs. In: *WordNet. An Electronic Lexical Database* (ed. Fellbaum). The MIT Press. Cambridge, Massachusetts, London.
- Fellbaum, Christiane (2002): On the Semantics of Troponymy. In: *The Semantics of Relationships: An Interdisciplinary Perspective*. R. Green, C. A. Bean, S. H. Myaeng (eds.) Kluwer. Dordrecht, Holland.
- Fillmore, Charles J., Christopher R. Johnson, Miriam R. L. Petruck (2003): Background to FrameNet. In: *International Journal of Lexicography*, Vol. 16 No.3. OUP, Oxford.
- Grimshaw, Jane (1990): *Argument Structure*. The MIT Press. Cambridge, Massachusetts, London.
- Kipper, Karin, A. Korhonen, N. Ryant, M. Palmer (2006): Extending VerbNet with Novel Verb Classes. Elektronisk publikation (www.cl.cam.ac.uk/~alk23/lrec06.pdf).
- Levin, Beth (1993): *English Verb Classes and Alternations*. A preliminary Investigation. The University of Chicago Press. Chicago and London.
- Pedersen B., Paggio P. (2004): The Danish SIMPLE Lexicon and its Application in Content-based Querying. In: *Nordic Journal of Linguistics Vol.27:1*, s. 97-127. United Kingdom.
- Pedersen, B.S., S. Nimb, N., J. Asmussen, N. Sørensen, L. Trap-Jensen, H. Lorentzen (2006). DanNet - A WordNet for Danish. In: *Proceedings from Third International Conference on Global Wordnets*, pp. 329-331. Jeju, South Korea.
- Somers, Harold (1987): *Valency and Case in Computational Linguistics*. Edinburgh University Press. Edinburgh.
- Trask, R.L. (1993): *A Dictionary of Grammatical Terms in Linguistics*. London and New York.
- Vossen, Piek (1998): Introduction to EuroWordNet. In: *EuroWordNet, A Multilingual Database with Lexical Semantic Networks*. (ed. Vossen). Kluwer Academic Publishers. Dordrecht/ Boston/London.

Hjemmesider på internetet (pr. 12. december 2008)

- DanNet: <http://wordnet.dk>
 EuroWordNet: <http://www.illc.uva.nl/EuroWordNet/>
 FrameNet: <http://framenet.icsi.berkeley.edu/>
 SIMPLE: <http://www.ub.es/gilcub/SIMPLE/simple.html>
 STO: http://cst.ku.dk/sto_ordbase
 Wordnet, Princeton: <http://wordnet.princeton.edu/>

Anna Braasch
 seniorforsker, f. 1946
 Center for Sprogteknologi
 Københavns Universitet
 Njalsgade 80
 DK-2300 København S
 e-mail: braasch@hum.ku.dk

