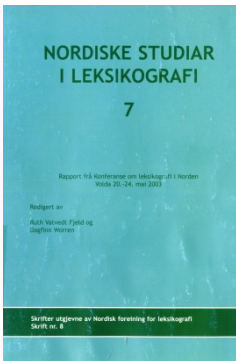


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Vi tager nogle kollokationer i øjesyn (En analyse af verbalkollokationers lingvistiske egenskaber)	
Forfatter:	Anna Braasch	
Kilde:	Nordiske Studier i Leksikografi 7, 2005, s. 67-77 Rapport fra Konferanse om leksikografi i Norden, Volda 20.-24. maj 2003	
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.

Vi tager nogle kollokationer i øjesyn

(En analyse af verbalkollokationers lingvistiske egenskaber)

In practical lexicography, both for human use and natural language processing applications, one of the most discussed topics is the appropriate linguistic description of complex lexical units. The subject and approach described here are closely related to the Danish STO project producing a lexicon for language technology applications. The paper focuses on a subtype of complex lexical units, viz. verbal collocations (operational term, VK) and describes a method of systematic analysis of the linguistic features of the collocations selected. First, the subject of the article is presented and the term 'verbal collocation' is defined. Second, the method of empirical investigations of corpus examples is discussed. Third, a number of selected VK's are analysed wrt. their formal, structural and lexical variation. Both monolingual and contrastive aspects are discussed. Finally, we illustrate the method of classification of the observations; these classifications serve as a basis for formalised descriptions of collocations.

1 Baggrund

Der er bred enighed om at sproget har en mængde leksikalske enheder som består af flere ord (med en neutral term også kaldet 'flerordsenheder', andre betegnelser er fx frasem, faste eller fraseologiske udtryk mv.) En flerordsenhed kan beskrives som «en sproglig størrelse, som består af mere end ét ord, og hvis betydning og/eller brug ikke kan beskrives udtømmende alene ved generelle syntaktiske regler og en leksikalsk beskrivelse af enkeltordene» jf. Boje & Braasch (1992, p. 484.)

Såvel mennesker som sprogteknologiske programmer tager udgangspunkt i det regelmæssige i sproget, de kombinerer viden om ords egenskaber og generelle grammatiske regler.

Kollokationer udgør et velkendt problem både i fremmedsprogstilegnelse og i sprogteknologi, fordi de kan afvige fra 'det regelmæssige' på forskellig vis, hvilket ofte forårsager fejl i deres reception og (re)produktion. Fejl opstår på følgende vis.

Et menneske overfører fx nogle generelt gældende strukturer fra sit modersmål til fremmedsproget (interferens) eller danner en konstruktion ud fra et kendt morfosyntaktisk regelsæt og en regulær valensstruktur (fejlanalogi). En 'maskine' (et program) anvender de generelle regler ved analysen af kildesprogtekst som på overfladen ligner en regelmæssig, frit dannet struktur (og 'opdager' derved ikke at der er tale om en kollokation). I oversættelse til målsproget fører dette til fejl, da de generelle regler ikke gælder fuldt ud, men kun med begrænsninger.

For at eliminere de nævnte fejlkilder for maskiner er det nødvendigt at undersøge hvilke forskelle der kan registreres mellem et lemmas regelbundne strukturer (med frie og valensbundne led) og i kollokationelle strukturer. Hvis disse beskrivelser er utilstrækkelige, så vil et oversættelsessystem producere ukorrekte – og i værste fald uforståelige – ytringer.

2 Fraseologi som leksikografisk emne

Håndtering af fraseologi i ordbøger er et emne som optager mange lingvister og leksikografer; således er der også tradition for at der på alle NFL-konferencer er en række indlæg som beskæftiger sig med forskellige aspekter og typer af fraseologi, eksempelvis Boje & Braasch (1992), Clausén (1999), (2001), (2003), Bönneberg (2003), Jónsson (2003) og Nilsson (2003). Det samme gælder andre, store internationale leksikografikongresser, også konferencerapporterne fra EURALEX indeholder en række relevante bidrag om emnet. Dansk fraseologi fra en praktisk leksikografisk synsvinkel er behandlet fx i Nimb (1996) og Trap-Jensen (1996) i forbindelse med udarbejdelsen af Den Danske Ordbog.

3 Artiklens emne og indfaldsvinkel

Denne artikel fokuserer på en operationelt afgrænset, særlig hyppig type af flerordsenheder, nemlig verbalkollokationer (vores term). Der præsenteres en analysemetode der er udarbejdet med henblik på kollokationers lingvistiske beskrivelse i en leksikalsk datasamling for sprogteknologiske anvendelser, et såkaldt leksikon. Behandlingen af emnet i denne artikel er tæt forbundet med arbejdet på den danske SprogTeknologiske Ordbase, STO, idet målet er at udarbejde principper for flerordsenheders leksikografiske beskrivelse der er velegnede til implementering i ordbasen. Den monolingvale SprogTeknologiske Ordbase findes kun på elektronisk medium, primært beregnet til anvendelser i sprogteknologiske produkter, fx i systemer til maskinoversættelse eller informationssøgning, programmer til støtte for sprogundervisning, tekstresumering osv. Ordbasen er desuden også velegnet som leksikalsk ressource såvel i almen og datamatisk lingvistik som i leksikologi på grund af materialets omfang, finmaskethed og systematik. STO er korpusbaseret, både med hensyn til udvælgelse af ordforrådet og dets beskrivelse. Teksterne stammer fra perioden 1990 – 2003. Ordforrådet omfatter moderne almensprog og domænesprog med lav fagsproglighedsgrad fra forskellige emneområder (edb, miljø, helse, forvaltning, økonomi samt handel & erhverv).

STO-projektet er afsluttet d. 29. februar 2004; for yderligere oplysninger om projektet se <http://cst.dk/sto/sto.html>. Der er online adgang (med udførlig vejledning) til at søge i ordbasen via brugergrænsefladen (<http://cst.dk/sto>), og der er også nogle søgetips der er værd at afprøve. Gennem denne grænseflade kan der også søges i relevante danske elektroniske ordbøger og et par korpusser.

Den færdige, første version af ordbasen indeholder ikke fraseologi (herunder kollokationer), da fokus ligger på enkeltords morfologi og syntaks – og delvis semantik. Med hensyn til kollokationer drejer det sig primært om at udvikle og afprøve ovennævnte analysemetoder, samt at udføre kodning på testbasis for et udvalg af sådanne ord der hyppigt indgår i kollokationer.

4 Skematisk klassifikation af ordforbindelser

En løbende tekst består af sætninger der er forbindelser af ord, og disse ordforbindelser kan karakteriseres ud fra deres struktur, variation og grad af leksikalisering. En meget forenklet måde til at anskue forholdet mellem 'frie' og 'faste' ordforbindelser skitseres i Figur 1. De frie ordforbindelser karakteriseres ved fri variation mht. form og struktur i overensstemmel-

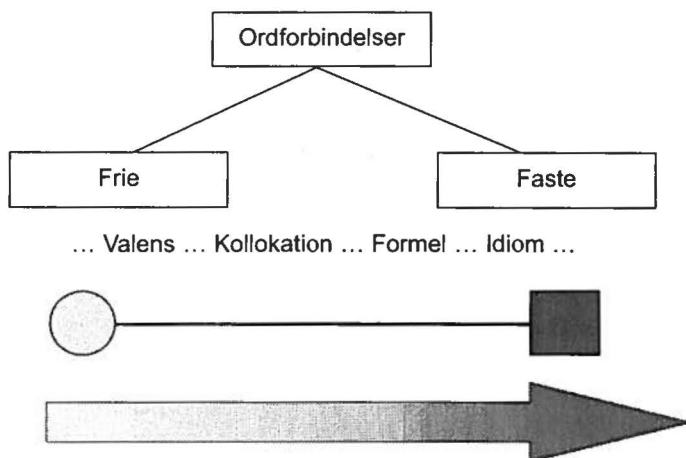


Fig. 1. Kollokationens placering mellem frie og faste ordforbindelse

se med sprogets generelle regler og kompositionel betydning. Faste ordforbindelser er leksikaliserede udtryk med ingen eller meget ringe grad af form- og strukturvariation og non-kompositionel betydning. Det betyder at jo mere frihed til variation, jo mindre leksikaliseret er ordforbindelsen – og omvendt: mindre variationsfrihed er et tegn på højere grad af leksikalisering, men kohæsiionsprocessen har mange mellemstadier uden skarpe grænser imellem.

Ifølge vores definition består en kollokation af to (ord)grupper af indholdsord, det kan være verbum, substantiv eller adjektiv og et (antal) valensbundne led. Den ene af komponenterne bærer (hoved)betydningen og er den syntaktisk og semantisk faste del ('basis'), mens den anden komponent er semantisk 'svag' ('kollokat').

I Figur 1 er kollokationerne placeret efter valensbundne forbindelser på vej mod faste ordforbindelser eller udtryk. En kollokation er transparent mht. udtryk, men idiomatisk mht. struktur og udgør en slags glidende overgang mellem obligatorisk leksikalisk valens og idiomatiske, leksikaliserede strukturer, som det eksempelvis afspejles i ordforbindelsers klassifikation i Somers (1987).

Det er svært at definere en præcis grænse mellem de ordforbindelser der er valensbundne og dem vi behandler som kollokationer, fordi overgangen fra et ords stærke selektionsrestriktion med tiden manifesterer sig som kollokation. At det på ingen måde er trivielt at drage grænsen mellem valensbundne strukturer i 'fig.' betydning og 'idiomer' kan man konstatere, hvis man ser på hvor forskelligt det samme udtryk er behandlet i forskellige ordbøger, både mht. hvor i makro-/mikrostrukturen det anbringes, og hvordan det beskrives.

5 Verbalkollokationer

Termen 'verbalkollokation' (VK) refererer til en type kollokationer der udviser en kombination af følgende lingvistiske egenskaber.

Grammatiske og strukturelle egenskaber:

- I kollokationens kanoniske form er det første ord et verbum (kollokat)
- Kollokationen har en struktur der minder om verbers valensbundne strukturer
- Kollokationen som helhed har en verbumlignende funktion i sætningen (prædikat) og kan ofte erstattes med et synonymt verbum
- Kollokationens morfosyntaktiske variabilitet er begrænset i forhold til den der gælder for en parallel fri struktur

Leksikalsk-semantiske egenskaber:

- Verbets betydning er 'fig.' (figurativ), 'overf.' (overført) eller er svækket
- Udtrykket er ofte kun delvis semantisk transparent, graden af transparens varierer meget
- Udtrykkets betydningshelhed er mere eller anderledes end summen af dets komponenters betydning
- Udtrykkets leksikalske variabilitet er meget begrænset.

Valget af termen 'verbalkollokation' kan også begrundes operationelt med at en maskine læser teksten fra venstre mod højre, og et analyseprogram (eksempelvis en grovparser) udfører opmærkning af hvert tekstord efter opslag i det tilknyttede leksikon. I den kanoniske form og i de fleste forekomster af de her behandlede kollokationstyper står verbet i dansk først, dvs. til venstre for basen. Derfor er det hensigtsmæssigt at der, ved opslag af verbet i ordbasen, er adgang til oplysninger om dets potentielle funktion i kollokationer. I andre sprog, fx tysk er forholdet anderledes, idet ordstillingen i den kanoniske form afviger fra den bøjede form, jf. nedenstående eksempel:

*tabe hovedet: han taber / tabte/ har tabt hovedet vs.
den Kopf verlieren: er verliert/verlor den Kopf /hat den Kopf verloren*

I den efterfølgende beskrivelse er der ikke taget særlig hensyn til semantiske egenskaber som graden af transparens, betydningkohæsion eller andre forhold der traditionelt danner grundlag for forskellige typologier og klassifikation af kollokationer.

5.1 Korpusbaserede undersøgelser

Kernen i det praktiske arbejde i klassifikationsfasen og analyseprocessen er baseret på empiriske undersøgelser. Det første skridt er at afgrænse verbalkollokationer fra valensstrukturer og undersøge deres interne frasestruktur. Det andet skridt er at sammenligne disse kollokationers morfosyntaktiske, strukturelle og leksikalske egenskaber med parallelle, frie ordforbindelser (syntagmer).

Følgende korpusser er anvendt i undersøgelsen:

Berlingske Aviskorpus 1990-92 (på ca. 20 millioner ord), Korpus2000 (DSL, på ca. 28 millioner løbende ord) og DK Korpus 87-90 (på ca. 4 millioner ord).

I det efterfølgende behandles en række verbalkollokationer (VK), der optræder hyppigt i korpus, primært avistekster. Der er lagt vægt på følgende egenskaber:

- VK består af to hovedkonstituer, nemlig
 - (a) det verbum der i kanonisk form står på første plads i læseretningen, det har ofte svækket egenbetydning ('kollokat') og er ofte arbitrært valgt mellem synonymer
 - (b) et nominalsyntaxme og/eller præpositionssyntaxme (+ evt. valensstyrrelse) denne er den betydningsbærende konstituent ('basis')
- VK udviser begrænset variabilitet mht. form, struktur og leksik i forhold til parallelle frie ordforbindelser.

5.2 Analyse af intern frasestruktur – Strukturelle undertyper

Et bredt udvalg af et ords korpusforekomster undersøges (i form af konkordans) og de prototypiske kollokationer klassificeres. Nedenfor gives nogle eksempler på de udvalgte verbalkollokationens interne struktur med verberne *tage*, *tabe* og *fælde* som kollokat.

Forklaring til parenteserne i eksemplerne nedenfor:

[] omslutter et optionelt led, { } et obligatorisk led.

- V + NP (NP er base, forskellig mht. bestemthed i eksemplerne)
 - tage affære*
 - tage sin afsked*
 - tabe hovedet*
 - tabe sit hjerte*
 - fælde en tåre*
- V + NP [+ PP] (NP er base, PP er valensbunden, optionel)
 - tage bestik [af ngn/ngt]*
- V + {NP} + PP (NP er valensbunden, obligatorisk, PP er base)
 - tage {ngn} på ordet*
- V + PP1 + PP2 (PP1 er base, PP2 er valensbunden, obligatorisk)
 - tage til takke med {ngt}*
- V + PP1 [+PP2] (PP1 er base, PP2 er valensbunden, optionel)
 - tage til genmæle [over for ngn]*

Verbet (kollokatet) er den konstituent der sikrer kollokationens tilpasning til sætningskonteksten ved hjælp af bøjning, desuden tilføjer det ofte noget modificerende til udtrykkets betydningshelhed (fx aspektuel nuance), men samtidig beholder verbet også nogle elementer af sin egen kernebetydning. Verbet er den – stort set – frit variable del af en verbalkollokation.

Kollokationer er særlig interessante i maskinel sprogbehandling fordi de på overfladen ligner frie ordforbindelser der er dannet ved fri kombination af sprogets grammatiske og leksikalske regler, men de udviser begrænsninger i forhold til hvilke kombinationer der er mulige eksempelvis med hensyn til formvariation, og de oversættes ikke kompositionelt.

5.3 Systematisk analyse af korpusforekomster

For hver VK undersøges hvilke begrænsninger der gælder med hensyn til dens syntagmatiske, paradigmatiske og leksikalske variabilitet sammenlignet med strukturelt set parallelle frie ordforbindelser. Følgende egenskaber er analyseret:

- Nominalsyntagmets leksikalske modifikation ved attributivt adjektiv og/eller negation (inds kud)
- Kollokationens modifikation ved et adverbium eller negation (inds kud)
- Substitution af kollokatet eller (en del af) basen med et synonym uden at betydningsenheden ændres
- Nominalsyntagmets (basens) variation mht. numerus og bestemthed
- Verbets variation i tempus og diatese

Anden form for modifikation eller udvidelse af VK-strukturen, hvis relevant er fx syntaktisk transformation som topikalisering, nominalisering.

Analysens hovedformål er at registrere hvilke begrænsninger der gælder mht. variabilitet i forhold den regelmæssige kombinatorik i sproget. Formålet med dette er at udforme verbal-kollokationers beskrivelse på en sådan måde at fejl, der vil kunne opstå ved brug af de generelle regler i analyse og produktion (såkaldt overgenerering), undgås.

Efter systematiseringen af relevante korpusbelæg sammenfattes analyseresultaterne, og de danner grundlag for den formaliserede beskrivelse af kollokationens individuelle egenskaber. Beskrivelsesmetoden er i overensstemmelse med STOs generelle mønsterbaserede repræsentation af lingvistiske egenskaber. Nedenstående eksempler illustrerer analysen af nogle udvalgte egenskaber. (I eksemplerne markeres forkerte komponenter eller former med *.)

5.3.1 Nominalsyntagmets paradigmatiske variation

En kollokation, jf. eksemplerne i (1a) ligner på overfladen en fri ordforbindelse som i (1b) der er dannet ved kombination af viden om ords egenskaber og generelle grammatiske regler, men der er individuelle begrænsninger i de anførte eksempler i (1a) mht. basens dvs. objektets mulighed for formvariation.

- (1a) *Maria tager en pause/*pausen/*sin pause fra arbejdet*
*Maria tager affære /*affæren*
*Maria taber hovedet/*sit hoved*
Maria taber ansigt vs. redder/bevarer ansigtet
*Maria tager initiativ/initiativet/*sit initiativ til en udflugt*

- (1b) *Maria tager en bog/bogen/sin bog/et par bøger fra bordet*

Der er to iagttagelser med hensyn til objektets formvariation: for det første er den meget begrænset, i de fleste tilfælde er formen invariabel, for det andet er der tilsyneladende ingen logisk begrundelse for bestemthedsforholdet. Det understøttes af at en lignende arbitrær distribution kan ses i sammenligningen mellem forskellige sprog (1c).

- (1c). DA: *tabe ansigt* (umarkeret) *redde/bevare ansigtet* (bestemt)
 SV: *förlora ansiktet* (bestemt) *rädda ansiktet* (bestemt)
 EN: *lose (one's) face* (poss./ umarkeret) *save face* (umarkeret)
 TY: *sein Gesicht verlieren* (poss.) *das Gesicht wahren* (bestemt)

Dette udgør altså også et kontrastivt problem ved oversættelse da man ved kollokationer ikke kan overføre bestemthedsforhold fra et sprog til et andet efter de samme regler som der gælder for frie syntagmer.

5.3.2 Leksikalsk modifikation og variation af nominalsyntaxmet

I frie ordforbindelser kan nominalsyntaxmet modificeres ved attributivt brugte adjektiver, og deres betydning kan varieres ved brug af synonymmer på passende steder. Dette gælder kun i ringe, om end noget varierende omfang ved verbalkollokationerne; i nogle tilfælde kan der bruges nogle ganske få, semantisk beslægtede adjektiver. Eksemplerne i (2a) og (2b) viser at adjektivisk kvantificering ikke bruges, hvorimod hele verbalkollokationen kan som regel negeres eller modificeres ved brug af et adverbium.

- (2a) *X tager* <ofte, aldrig, vist ikke> *{(en) stor/lille} del i nyt
 (2b) *X tager* *<ingen, mange, nogle gevaldige> kegler

Den monolingvalt set uforudsigelige, arbitrære distribution af kollokatet mellem en række synonymmer udgør et grundlæggende problem, fx i to tæt beslægtede kollokationer. Eksemplerne (2c) og (2d) viser at affiniteten mellem VKs komponenter er så stærk at en komponent ikke kan udskiftes med et synonym, hvorimod der i (2e) er frit valg mellem to verber. Hvorvidt en af komponenterne kan erstattes med sit synonym, dvs. VKs leksikalske variabilitet, afhænger af graden af VK's indre kohæsion og graden af leksikalisering.

- (2c) *X tager*/*træffer affære /*anliggende
 (2d) *X træffer*/*tager en afgørelse
 (2e) *X tager/træffer* en beslutning

Set fra en kontrastiv synsvinkel viser disse arbitrære valg sig at være sprogspecifikke, selv i en kollokation der findes parallelt på flere sprog (2f)

- (2f) finde sted (DA)
äga rum (SV)
take place (EN)
avere luogo (IT)
stattfinden (DE)

(Den tyske kollokation har en kanonisk form der svarer til et præfigeret verbum.)

Det er derfor relevant af både monolingvale og kontrastive grunde at angive graden og arten af modifikation og variation i den enkelte VKs leksikalske beskrivelse.

5.3.3 Syntaktisk og strukturel variation

Verbet er den fleksible konstituent i en VK, det tilpasses mht. tempus og delvis også mht. diatese til den givne kontekst. Det viser sig dog ofte (som i 3(a)) at en VK ikke kan passivendes, dvs. dets transitive verbum ikke transformeres fra aktiv- til passivkonstruktion uden tab af den idiomatiske betydningshelhed.

- (3a) *X tabte ansigt/hovedet/sit hjerte ved den omtalte begivenhed*
 **Ansigt/hovedet/ens hjerte blev tabt (af X) ved den omtalte begivenhed*

I andre tilfælde er passivkonstruktionen en syntaktisk variationsmulighed jf. (3b), som udnyttes i stilistisk øjemed fortrinsvis i upersonlige, formelle tekster. Der forekommer kun sjældent en direkte passiv vending af en aktiv konstruktion med det personlige subjekt udtrykt i et præpositionssyntagma med 'af' ('by-passiv').

- (3b) *Man tog affære ved denne lejlighed*
Der blev taget affære ved denne lejlighed.

Andre former for syntaktisk variation, der er undersøgt, er eksempelvis topikalisering af basen (3c), eller nominalisering af kollokationen (3d), koordinering (3e), brug af anaforer i forbindelse med kollokationer mv. Et par eksempler illustrerer nedenfor den generelle iagttagelse at verbalkollokationer udviser meget begrænset syntaktisk variabilitet i sammenligning med frie ordforbindelser.

- (3c) *tage kegler* (i betydningen 'vinde bifald')
 «Den rytmiske musik *tager* selvfølgelig *kegler* hos et klassisk publikum.»
 (korpuseksempel)
 **Kegler tager* den rytmiske musik selvfølgelig hos ... vs.
 Selvfølgelig *tager* den rytmiske musik *kegler* hos et klassisk publikum.
 (Her er en topikalisering af det modificerende adverbium mulig.)
- (3d) * Den rytmiske musik er en *kegletager* hos...
 * Der var ... tale om *kegletagning!*
- (3e) *tage affære* og *tage sin afsked tage* (transitivt verbum)
*Konen tog straks affære *og sin afsked. Konen tog affære *og bankbogen.*

Ud fra ovenstående eksempler på vore iagttagelser kan der konkluderes at den strukturelle og leksikalske affinitet samt den betydningsmæssige kohæsion mellem en verbalkollokations konstituenten viser sig tydeligt i deres ringe grad af variabilitet. Tabel 1 viser et udsnit af systematiseringen af analyseresultaterne mht. begrænsninger i verbalkollokationers muligheder for passivtransformation. (Tegnet 'Ø' markerer at der kan dannes såvel upersonlige som personlige passivkonstruktioner ('by-passiv') med kollokationen.)

6 Beskrivelsesmetode i STO

Analyseresultaterne sammenfattes som morfologiske, syntaktiske og leksikalske træk og derefter systematiseres disse med henblik på formel beskrivelse af kollokationernes egenskaber. Det næste skridt er at opstille en subklassifikation af de undersøgte kollokationer på grundlag af deres indre struktur og kombinationen af deres indre variationsmuligheder, Tabel 1 er en del af en sådan klassifikation. Det afsluttende skridt er at formalisere de begrænsninger der gælder for hver komponent (eller position i strukturen) i overensstemmelse med STOs beskrivelsessystem. Udgangspunktet er at de valgte typer af kollokationer skal kunne beskrives med det samme beskrivelsesapparat der anvendes for enkeltord og deres konstruktionspoten-

Subtype/ Struktur	Passivtransf. Begrænset	Kollokation: eksempel i kanonisk form
V+N(obj)	Kun upersonlig	<i>tage affære</i>
V+NP(obj +PP)	Kun upersonlig	<i>tage bestik af [ngt]</i>
	Ingen passiv	<i>tage æren for [ngt/ngn]</i>
V+NP(obj + <PP>)	Kun upersonlig	<i>tage stilling <til [ngt]></i> <i>tage hensyn <til [ngn/ngt]></i>
	Kun upersonlig	<i>tage ansvar/ansvaret <for [ngn/ngt]></i>
V+NP(obj)	Ingen passiv	<i>tage sin afsked</i> <i>tabe hovedet</i> <i>tabe sit hjerte</i>
V+[NP(obj)]+PP	Ø	<i>tage [ngt] i brug</i> <i>tage [ngt] i øjesyn</i>
V+PP	Kun upersonlig	<i>tage til genmæle</i>
	Ingen passiv	<i>tage til takke</i>

Tabel 1. Passivtransformation af kollokationen som helhed (udsnit)

tiale mv. Beskrivelsesapparatet består af en model og et formaliserebeskrivelsessprog. Modellen er tredelt: et lemmas beskrivelse er fordelt på tre lag: morfologi, syntaks og semantik. For hvert af disse lag er der udviklet et meget detaljeret beskrivelsessystem som også kan anvendes til kollokationers beskrivelse. Beskrivelsessystemet i STO er baseret på en systematisk liste over oplysningstyper (fx ords bøjning og styrelse) der er relevante ud fra data-lingvistiske og leksikografiske aspekter. Hver oplysningstype beskrives ved et sæt egenskaber eller træk (fx substantivbøjning: genus, numerus, kasus, bestemthed), og for hvert træk er der en udtømmende liste over dets mulige værdier (fx genus: fælleskøn, intetkøn, umarkeret). En oplysningstype beskrives følgelig vha. en kombination af træk+værdi-par, udtrykt i en struktureret kode. Hver unik kombination af koder udgør et mønster. Ords formrelaterede egenskaber, herunder bøjning, er beskrevet udtømmende i STO ved hjælp af morfologimønstre.

Beskrivelsen af ords syntaktiske egenskaber er baseret på valensteorien og er udtrykt i syntaksmønstre. I disse mønstre registreres antallet af led der knyttes til et verbum, substantiv eller adjektiv. Arten af det enkelte led beskrives mht. dets ledfunktion (fx subjekt, objekt), den syntaktiske kategori og struktur (fx nominal- eller præpositionssyntagme, infinitiv, led-sætning), den styrede præposition og andre relevante egenskaber.

Disse formrelaterede og kombinatoriske egenskaber er ligeså relevante når ord indgår i kollokationer, og mønstrene der er etableret for enkeltords beskrivelse, kan udnyttes på den måde at de begrænsninger der gælder for basens eller kollokatets variabilitet i forhold til en parallel, fri ordforbindelse, markeres i det anvendte mønster.

Der gives en mere detaljeret redegørelse for formaliseringen af relevante oplysninger i Braasch & Olsen (2000a) og Braasch & Olsen (2000b) som er tilgængelige på hjemmesiden <http://cst.dk/projekter/sto>.

7 Sammenfatning

Artiklen præsenterer en analyse af nogle egenskaber ved verbalkollokationer der er væsentlige i forbindelse med korrekt sproglig formulering. Maskinel sprogbehandling stiller nogle særlige krav til præcision og detaljeringsgrad i ords beskrivelse, men lignende behov har også et menneske i en oversættelsessituation.

Grænsen mellem traditionel og datamatisk leksikografi er i dag meget fleksibel, i begge bruges elektroniske korpusser og værktøjer. Datamatisk leksikografi har siden 1970erne arbejdet meget med udnyttelse af eksisterende leksikalske data for nye formål, herunder for sprogteknologiske ordbaser. Genbrug af materiale og metoder der udvikles i forbindelse med datamatisk leksikografi, herunder den præsenterede analysemetode kan også gå den anden vej, da netop de nye tekniske muligheder – brugen af computer til udarbejdelse, lagring og publicering af ordbøger for mennesker – giver mulighed for at integrere flere og mere detaljerede oplysninger.

Referencer:

- Boje, Frede Braasch A. 1992: Hvad får man skudt i skoene? Flerordsenheder i aktive ordbøger for mennesker og maskiner. I: R. Vatvedt Fjeld (red.): *Nordiska Studier i lexicografi 1, Rapport fra Konferanse om lexicografi i Norden*. Oslo.
- Braasch, Anna & S. Olsen 2000a: Formalised Representation of Collocations in a Danish Computational Lexicon. I: *The Ninth EURALEX International Congress, Proceedings, Vol. II*. Stuttgart.
- Braasch, Anna & S. Olsen 2000b: Towards a Strategy for a Representation of Collocations – Extending the Danish PAROLE-Lexicon. I: *Second International Conference on Language Resources and Evaluation, Proceedings, Vol. II*. Athens.
- Bönnemark, Margit 2003: Danska och svenska binominala uttryck – en jämförelse. I: Zakaris Svabo Hansen & A. Johansen (red.) *Nordiska Studier i lexicografi 6, Rapport fra Konferanse om lexicografi i Norden*. Tórshavn.
- Clausén, Ulla 1999: Om pragmatiska fraser i Svensk konstruktionsordbok. I: Peter Slotte, Westerberg, P. & Orava, E. (red.) *Nordiska Studier i lexicografi 4, Rapport från Konferansen om lexicografi i Norden*. Helsingfors.
- Clausén, Ulla 2001: Funktion och attityd i Svensk konstruktionsordbok. I: Martin Gellerstam, K. Jóhannesson, B. Raplh, L. Rogström (red.) *Nordiska Studier i lexicografi 5, Rapport från Konferens om lexicografi i Norden*. Göteborg.
- Clausén, Ulla 2003: Jämförelseskonstruktioner i Svensk konstruktionsordbok – struktur, variation och funktion. I: Zakaris Svabo Hansen & A. Johansen (red.) *Nordiska Studier i lexicografi 6, Rapport fra Konferanse om lexicografi i Norden*. Tórshavn.
- Jónsson, Jón Hilmar 2003: Fraseologien i forgrunnen – fraseologisk register som ledd i ordbokens tilgangsstruktur. I: Zakaris Svabo Hansen & A. Johansen (red.) *Nordiska Studier i lexicografi 6, Rapport fra Konferanse om lexicografi i Norden*. Tórshavn.
- Nilsson, Kåre 2003: Eksempler og uttrykk (idiomer) i tospråklige ordbøker. I: Zakaris Svabo Hansen & A. Johansen (red.) *Nordiska Studier i lexicografi 6, Rapport fra Konferanse om lexicografi i Norden*. Tórshavn.
- Nimb, Sanni 1996: Collocations of Nouns: How to Present Verb-noun Collocations in a Monolingual Dictionary. I: Gellerstam, Martin, J. Järborg, S-G. Malmgren, K. Norén, L. Rogström, C. Røjder Papehl (red.) *EURALEX '96 Proceedings*. Göteborg.

-
- Sinclair, John 1991: *Corpus, Concordance, Collocation*. Oxford University Press, Oxford.
- Somers, H. L. 1987: *Valency and Case in Computational Linguistics*. Edinburg University Press.
- Trap-Jensen, Lars 1996: *Word Realtions: Two kinds of Typicality and their Place in the Dictionary*. I: Gellerstam, Martin, J. Järborg, S-G. Malmgren, K. Norén, L. Rogström, C. Røjder Papmehl (red.) *EURALEX '96 Proceedings*. Göteborg.