

*Anna Braasch\**

## **Kollokationer som fraseologisk kategori set fra forskellige synsvinkler**

### **Abstract**

The delimitation of collocations as a category on its own can be done along different lines. The purpose of this paper is to discuss the definition and use of the term 'collocation' from various points of view. Collocations differ from free word combinations on one hand and from idioms on the other, having two characteristic features: a certain degree of meaning compositionality and a certain degree of formal, structural and lexical fixedness. These features appear in several combinations which form the basis of different classifications. First, criteria and classifications are discussed; second, the use of the term in lexicographic theory and corpus linguistics is analyzed. The discussion is illustrated by a few examples of lexicographic representation in Danish dictionaries. Finally, an outline of various descriptive approaches with their focus on particular features is presented.

### **1. Terminologi og forvirring omkring begreber**

En af de væsentligste problemstillinger i diskussionen om sprogets fraseologiske materiale er den om ikke vilkårlige, så dog uensartede brug af fraseologiens centrale termer i forskellige beskrivelser af samme fænomener. Det første DANFRAS-Symposium i november 2004 havde derfor også som et af sine mål at diskutere "fraseologispecifikke termer og begreber på dansk", og i DANFRAS-Nyhedsbrev nr. 2 skrev Ken Farø at "Det manglende kendskab til eller *forvirring omkring begreber* (min fremhævelse) er nemlig en stor hindring for udviklingen af et fælles diskussionsgrundlag, der rækker ud over det rent impressionistiske."

---

\* *Anna Braasch*  
*Center for Sprogteknologi*  
*Københavns Universitet*  
*Njalsgade 80*  
*DK-2300 København S*

Der er flere forskellige grunde til at man har svært ved at nå en bred enighed om fraseologiens begrebsdefinitioner og betegnelser og dermed skabe en homogen terminologi. En af de væsentligste grunde er at fraseologiforskning og fraseografi har mange indfaldsvinkler og disse er afgørende for hvordan begreberne defineres og termerne bruges.

Problemstillingen med hensyn til terminologisk klarhed har to sider: dels bruges en og samme betegnelse om forskellige sproglige størrelser eller fænomener, dels bruges mere end en betegnelse (mere eller mindre synonymt) om samme sproglige størrelse eller det samme fænomen. Et par velkendte eksempler er rækker af synonyme termer (1) og flertydige termer (2).

- (1) frase – frasem – fraseoleksem – fraseologisk enhed, flerordsenhed – flerordsudtryk – polyleksikalsk enhed idiom – idiomatisk udtryk – fast udtryk – fast vending
- (2) kollokation: en kategori af etablerede ordforbindelser med særlige strukturelle og leksikalske egenskaber; fra en korpuslingvistisk synsvinkel: to eller flere ord der hyppigt forekommer i hinandens nærhed

Dette er et velkendt problem, ikke kun i dansk – og ikke kun i den fraseologiske terminologi. Generelt er det vigtigt at præcisere i hvilken betydning man anvender en term når der kan være tale om usikkerhed eller forvirring omkring dens brug. Behovet for præcision er naturligvis særlig stort i videnskabelig kontekst.

## 2. Udgangspunkt for diskussion af termen kollokation

Etablerede ordforbindelser eller flerordsenheder kan identificeres, klassificeres og beskrives ud fra flere teoretiske og praktiske synsvinkler – eksempelvis med henblik på lingvistiske, leksikografiske, sprogdidaktiske eller oversættelsesmæssige formål. Fraseologiforskningen byder på mange forskellige metoder og fremstillinger, og brugen af terminologi er divergent i de forskellige fremstillinger.

Det er netop den tidligere nævnte forvirring, eller måske rettere sagt mangfoldighed, der kendetegner billedet med hensyn til hvad man i almindelighed forstår ved en **kollokation**. Termen kollokation dækker over en undertype af fraseologiske enheder, men kriterierne for hvordan man definerer denne undertype, dvs. afgrænsningen i forhold til idiomer og andre etablerede flerordsenheder på den ene side og frie ordkombinationer på den anden side, bestemmes af den synsvinkel man anlægger.

I det følgende forsøger jeg at belyse hvad begrebet kollokation dækker i en række forskellige lingvistiske sammenhænge. Udgangspunktet er en overordnet opdeling af etablerede flerordsenheder eller faste ordforbindelser i fire hovedkategorier: **citragtige fraser** (her er sentensen valgt som prototype, men også ordsprog, bevingede ord osv. hører til her), **kommunikationsformler** (skabeloner, rutineytringer osv.), **idiotmer** og **kollokationer**. Afgrænsningen af de enkelte hoved- og underkategorier og deres særskilte typer kan foretages på forskellig vis på grundlag af semantiske, leksikalske, formelle/strukturelle, funktionelle osv. kriterier. Da disse kriterier ofte overlapper hinanden eller optræder i kombination med hinanden, er det meget vanskeligt at tale om skarpe, entydige grænser. Enhver fraseologisk typologi afspejler den særlige kombination af klassifikationskriterier som den er baseret på og af samme grund er typologierne svære at sammenligne. Det er ikke hensigten her at opstille en typologi eller et klassifikationssystem, men blot at skitsere problematikken i denne forbindelse.

Termen kollokation stammer oprindeligt fra den britiske kontekstorienterede sprogforskning (Firth, 1957) og er blevet et centralt begreb i engelsk korpuslingvistik (fx hos Sinclair siden 1966). Bahns (1996: 6-25) giver et kortfattet og godt indblik i kollokationsbegrebets dannelse og udvikling i den engelske sprogforskning. I dag defineres, fortolkes og anvendes termen på mange måder, idet man kan fokusere på forskellige kriterier alt efter indfaldsvinkel og formål.

## 2.1. Klassifikation af ordforbindelser

Den grundlæggende antagelse er at alle sproglige ytringer består af ordforbindelser, og disse kan være mere eller mindre etablerede og faste med hensyn til deres betydning, form og komponentstruktur (disse egenskaber i kombination kan siges at udgøre graden af leksikalisering).

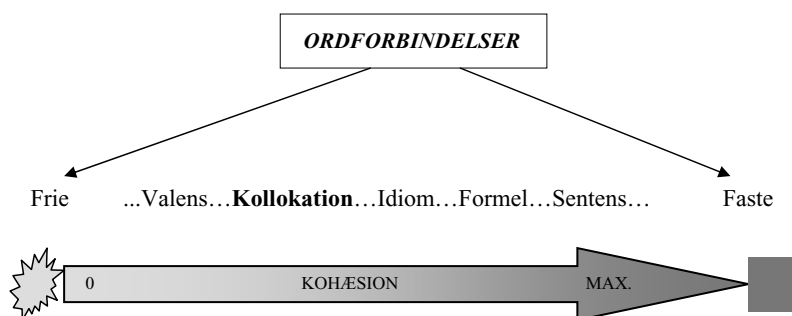
En ordforbindelses betydning som helhed karakteriseres ved graden af dens **kompositionalitet**, og den kan karakteriseres som

- **forudsigelig** (jf. termen i Brink, 2004), som er i overensstemmelse med sprogets generelle produktionsregler og dermed er fuldt kompositional

- **gennemsigtig** (delvis kompositionel)
- **uigennemsigtig** (ikke-kompositionel)

Ordforbindelser med gennemsigtig egentlig betydning kan bruges i overført eller metaforisk betydning.

Ordforbindelsers **variabilitet** sammenfatter forskellige typer og grader af ændring i form og struktur uden at det medfører en (uforudsigelig) betydningsændring, såsom morfosyntaktisk tilpasning (fleksibilitet), modifikation ved indskud, negation (kontinuitet), mv. eller udskiftning af en komponent (leksikalsk stabilitet). Nedenstående skitse (Figur 1) viser hvordan disse to fundamentale træk, kompositionalitet og variabilitet, danner grundlag for indplaceringen af kollokationer på et kontinuum mellem frie og faste ordforbindelser. Kombinationen af disse to træk benævnes her som **kohæsion** – i mangel af en bedre term.



Figur 1. Ordforbindelser. Klassifikationskriterium: kohæsion mellem komponenterne.

Fuld kompositionalitet betyder oftest også at ordforbindelsens komponenter frit kan kombineres med andre ord, varieres (eksempelvis erstattes med et synonym eller et andet, semantisk kompatibelt ord) osv. Der er ingen særlig affinitet eller gensidig afhængighed mellem komponenterne, hverken grammatisk, leksikalsk eller semantisk. De frie ordforbindelser består af enkeltord med fuld egenbetydning som er kombineret med hinanden i overensstemmelse med sprogets generelle regler. Disse ordforbindelser udviser ingen indre kohæsion.

Ovenstående figur illustrerer at affiniteten (tilbøjeligheden til at selektere hinanden) mellem ordforbindelsernes komponenter vokser fra yderpunktet **frie** i venstre mod yderpunktet **faste** til højre, og samtidig mindskes både deres grammatiske og leksikalske variabilitet, dvs. der er tale om en øget indre kohæsion hen imod de helt faste ordforbindelser med maksimal indre kohæsion. Dette er omtalt hos Sinclair (1991: 110 ff.) som 'the idiom principle', dvs. måden som et sprogs ord organiserer sig på i faste flerordsenheder.

## 2.2. Afgrænsning af kollokationer som kategori

Der er mange kombinationer og grader af både kompositionalitet og variabilitet mellem de to yderpunkter, derfor er det begrundet at tale om et kontinuum. Selv om man opstiller en ganske detaljeret klassifikation, vil man finde glidende overgange og grænsetilfælde mellem de definerede undertyper af ordforbindelser langs med dette kontinuum. Et sådant tilfælde er overgangen mellem semantiske selektionsrestriktioner for valensbundne led og kollokationer, idet mange kollokationer ligner syntaktiske konstruktioner med stærke, specifikke eller også kaldet **strengt selektionsrestriktioner** mellem de leksikalske komponenter, eksempelvis konstruktioner med det transitive verbum *begå*. Verbet subkategoriserer for et valensbundet objekt som skal være en nominalfrase, fx et substantiv. For dette objekt gælder selektionsrestriktionen 'negativ eller kriminel handling'. I dette tilfælde er det endda muligt at opstille en endelig liste over de substantiver som verbet kan stå sammen med eller mere præcist være støtteverbum for, nemlig {*fejltagelse... forbrydelse, indbrud, mord, tyveri, hærværk, vold, voldtægt...* }. Ordforbindelser af denne type kan, alt efter klassifikationens formål, betragtes som valenskonstruktion dvs. som et verbum med det pågældende valensbundne led eller som kollokationer med det pågældende substantiv som kerne.

Ovenstående illustrerer også begrundelsen for at kollokationer er placeret mellem valensbundne konstruktioner og idiomer på den glidende kohæsionsskala. En anden måde at anskueliggøre den beskrevne klassificering af ordforbindelser på ses nedenfor. Tabel 1. viser nogle eksempler der er forsøgt klassificeret i overensstemmelse med kompositionalitets- og variabilitetskriteriet.

Kategori	Eksempel	Kompositionalitet	Variabilitet	Kohæsion
Enkeltord i komb.	<b>en dyr bog</b>	fuld	fuld	ingen
Valenskonstruktion	<i>tage</i> + objekt-np	fuld	fuld	ingen
<i>Uægte sammensat verbum m. valens/ Kollokation?</i>	<i>tage på</i> (i vægt) (+ np {angivelse;})	forudsigelig	leksikalsk begrænset NP: grad/mængde	nogen
Kollokation	<i>sige farvel</i>	forudsigelig	leksikalsk helt fast; verbet bøjes	betydelig
	<i>tage afsked</i>	gennemsigtig	leksikalsk helt fast, verbet bøjes	ganske betydelig
	<i>tage sin afsked</i>	gennemsigtig	leksikalsk helt fast, verbum+poss.pron. bøjes kongruent	ganske betydelig
<b>Kollokation /Idiom?</b>	<i>tage bestik af</i> + np {situation}	? gennemsigtig forudsat viden om søfart	leksikalsk begrænset NP: (symptom på) situation	ganske betydelig
	<b>tage affære</b>	?? forudsat viden om etymologi	leksikalsk helt fast; verbet bøjes	ganske betydelig
Idiom	<i>tage det sure med det søde</i>	metaforisk	leksikalsk helt fast; verbet bøjes	ganske betydelig
Formel, skabelon mv.	<i>at tænke sig!</i> ; <i>alt andet lige</i>	gennemsigtig	ingen	maksimal
Citatagtig frase/ Sentens mv.	<i>tag det som en mand, frue!</i>	metaforisk	ingen	maksimal

Tabel 1: Eksempler på ordforbindelsernes hovedkategorier

Tabellen illustrerer den svære grænsedragning mellem kategorierne, eksempelvis kan en ordforbindelses kompositionalitet vurderes subjektivt ud fra adgang til en vis mængde specialviden, fordi en kollokations eller et idioms betydning vil blive vurderet forskelligt mht. kompositionalitet afhængigt af om man (gen)kender komponenternes faktiske, oprindelige betydning eller det bagvedliggende billede. Dette er et af de svage punkter i den skitserede klassifikation som blandt andet fører til inkonsistent brug af termer.

### 2.3. Kollokationer som sprogligt problem

Det er almindelig kendt at kollokationer udgør et særligt problem i fremmedsprogsindlæring, tekstproduktion og leksikografi. Det samme forhold gælder også i sprogteknologi hvor computerprogrammer anvendes

til behandling af naturlige sprog. De vigtigste grunde til dette problem er:

- For det første er kombinationen af basis og dens kollokat(er) arbitrær og ikke forudsigelig, men er sprogbrugsbestemt (fx *tage* el. *træffe en beslutning* vs. kun *træffe en afgørelse*).
- For det andet afviger kollokationers indre variabilitet fra sprogets generelle grammatiske og leksikalske produktionsregler – de kan udvise en begrænset variabilitet på flere plan (med hensyn til deres morfologiske, syntaktiske og leksikalsk-semantiske fleksibilitet).
- For det tredje udviser kollokationens interne morfosyntaktiske struktur ofte idiosynkrasi (fx med hensyn til bestemthed jf. *tabe hovedet/ sit hjerte/ ansigtØ*).

Disse tre grundlæggende egenskaber eventuelt i kombination med en ikke fuldt ud forudsigelig eller gennemsigtig (ikke fuldt kompositionel) betydning gør at kollokationer skal læres på samme måde som andre etablerede faste udtryk; følgelig skal de registreres i ordbøger (for mennesker) og ordbaser (for maskiner).

### 3. Metaleksikografiske beskrivelser og definitioner

Kollokationer omtales generelt som en særskilt kategori af ordforbindelser, men deres leksikografiske beskrivelse i metaleksikografisk teori og leksikografisk praksis udviser nogle forskelle.

I det følgende illustreres mangfoldigheden i definitionen og brugen af termen ”kollokation”; eksemplerne er udvalgt fra velkendte leksikografiske håndbøger og indledende afsnit (brugervejledning mv.) fra relevante ordbøger. Kilderne er udvalgt hovedsageligt på grund af deres pragmatiske tilgang til definition og leksikografisk beskrivelse, og kilder med dansk eller nordisk relevans er foretrukket, da det drejer sig om en diskussion af dansk fraseologi. Der er dog givet plads til enkelte andre kilder med betragtninger som skønnes relevante i denne sammenhæng.

Sammenstillingen, der på ingen måde er udtømmende, viser to væsentlige ting: hvilke basale træk der er enighed om blandt flere af de citerede kilder og hvilke træk der er uenighed om eller forskelle mellem kilderne.

### 3.1. Nordisk Leksikografisk Ordbog

Nordisk Leksikografisk Ordbog (NLO, 1997) er et godt udgangspunkt for diskussionen, fordi formålet med at udgive NLO netop var at "...medvirke til udviklingen af nordisk leksikografisk teori, som var mangelfull" (NLO 1997:7), og dens centrale opgave var at "...få samlet nordisk leksikografisk terminologi og presentere den på en ensartet måde og som egen kategori." (ibid.)

NLO giver en meget kortfattet definition og opsummering af de centrale lingvistiske træk der karakteriserer kollokationer. Artiklen afsluttes med en liste over udvalgt litteratur om emnet (i dette tilfælde engelsk, tysk dansk og svensk). Listen er en henvisning til de benyttede kilder dvs. et fingerpeg om hvad de aktuelle definitioner på kollokationer er baseret på samt vejen til mere information om emnet. Definitionen lyder i sin korthed (op.cit., p. 154):

**"kollokasjon**

ordforbindelse der betydningen har en klar sammenheng med betydningen av de enkelte delene (...)"

Derefter følger en kommenterende og forklarende tekst på knap en spaltens længde, der nævner nogle generelle, centrale træk og kriterier mht. til definition, klassifikation og tilgang:

- en kollokation består af to dele
- den ene del af forbindelsen udgør en (overordnet) **basis** (dvs. er den semantiske kerne), mens den anden har en karakteriserende funktion som **kollokator** (dvs. specificerer, præciserer basis)
- basis er den konstante del af kollokationen
- kollokationers leksikografiske angivelse og beskrivelse er afhængig af ordbogstypen (reception/produktion)

Desuden gives der yderligere kommentarer til kollokationsbegrebet som vedrører

- kollokatorens eventuelle udskiftelighed uden at det medfører ændring i kollokationens betydning (dikotomi mht. leksikalsk variabilitet: åben / lukket kollokation)
- kollokatorens type dvs. indholdsord eller funktionsord (dikotomi: leksikalsk / grammatisk kollokation)



- supplerende tekstlingvistisk definition (nævnte kriterier: typiskhed og frekvens af samforekomst samt syntaktisk nærhed mellem kollokerende ord)

For at kunne få et rimeligt udbytte af artiklen *kollokasjon* i NLO er det nødvendigt at slå de termer op der henvises til som beslægtede termer, først og fremmest udgangspunktet i definitionen, nemlig ordforbindelser (op.cit., p. 206). Denne artikel lister og forklarer de tre grundlæggende kriterier som er almindeligt anvendte i definition og klassifikation af ordforbindelser og dermed også kollokationer som er en undertype af ordforbindelser.

Disse kriterier er:

- Semantisk transparens ( i hvilken grad svarer ordforbindelsens betydning til kombinationen af dens komponenters egenbetydning); her stipuleres en dikotomi mellem kollokationer (a) og idiomer (b)
  - a. "...der betydningen av ordforbindelsen som helhed er en kombinasjon av betydningen av de enkelte ord"
  - b. "...der betydningen ikke svarer til en semantisk kombinasjon av de enkelte ledd."
- Konventionalisering som omfatter en række tilstande mellem "åpne kollokasjoner til helt faste udtrykk" da kriteriet vedrører
 

"... i hvilken grad det [udtrykket] er en fast og uforanderlig leksikalsk enhet, eller om en viss variasjon er tillatt i enkelte led."
- Syntagmatisk status (dvs. sætning eller syntagme).

De første to kriterier er de centrale mht. placeringen af ordforbindelser i en ordbogs makro- og mikrostruktur, idet disse kriterier udtrykker ordforbindelsernes karakteristiske træk: kompositionalitet og variabilitet. Disse træk kan manifestere sig i mange forskellige former, og deres vurdering og behandling i praktisk leksikografi afspejler ikke altid lige gennemskueligt eller konsekvent dette forhold.

### 3.2. Handbok i lexicografi

I diskussionen af principper og metoder for ordbogsarbejde nævner eksempelvis Svensén (1987: 92ff) at ords **kombinatoriske egenskaber** har to grundlæggende aspekter, det grammatiske og det 'semantisk-leksikalske', og begge skal behandles i ordbøger, men han kalder kun

de semantisk-leksikalsk betingede ordforbindelser for kollokationer. Svensén påpeger endvidere at

(...) kollokationer och idiom (...) kan i princip betraktas som två olika kategorier. Men i praktiskt ordboksarbete är det ofta svårt att skilja dem åt, varför de i många ordböcker redovisas tillsammans.” (ibid.)

Dette er en praktisk observation som også gælder for moderne ordbøger, for eksempel Den Danske Ordbog (DDO), hvor idiommer dog markeres med ’*ofø.*’ for overført betydning. I modsætning til dette kan nævnes at Den Danske Netordbog kun medtager kollokationer, idiommer skal slås op i Den Danske Idiomordbog (se videre i afsnit 5.2.). Set fra brugerens synspunkt er den valgte kategorisering (kollokation vs. idiom) og dermed indplaceringen af ordforbindelsen i mikrostrukturen og/eller dens præsentation ikke altid umiddelbart indlysende, nogle gange endda kontraintuitiv.

### 3.3. Manual i fagleksikografi

Den fagleksikografiske tilgang i manualen (Bergenholtz & Tarp 1994: 117 ff.) tager sit udgangspunkt i den almensproglige definition og afgrænsning af ordforbindelser med kollokationer i fokus, fordi denne kategori er stærkt repræsenteret i fagsprog sammen med flerordstermer. Der stilles desuden særlige krav til den leksikografiske behandling af kollokationer i fagordbøger som er præciseret ud fra to faktorer: ordbogens brugertype og funktioner. Faglig og fagsproglig kompetence nævnes som de centrale kriterier for brugertypen. Ordbøgers funktioner defineres som reception, oversættelse og produktion, og disse funktioner kombineres med brugerens forhold til kildeproget (L1) og målproget (L2), dvs. dennes modersmåls- eller fremmedsprogskompetence. Bergenholtz & Tarp (op.cit. p. 119) henviser til at grænsedragningen mellem ”banale (= frie) ordforbindelser” og ”semantisk specifikke kombinationer” (= kollokationer) ikke er absolut, men til en vis grad afhænger af leksikografens opfattelse. Det er dette forhold der er kernen i problematikken mht. leksikografisk praksis som illustreres med nogle eksempler fra ordbøger i afsnit 5.

### 3.4. The BBI Combinatory Dictionary of English

I indledningen giver Benson et al. (BBI; 1986: ix) udtryk for samme principielle udgangspunkt som Svensén (1987) med hensyn til de to

typer kombinatoriske træk, og disse træk danner grundlaget for to adskilte kollokationskategorier, her kaldet grammatisk og leksikalsk betingede kollokationer. I modsætning til Svensén henregnes i BBI altså begge til hovedkategorien kollokation, hvor (traditionelt, ligesom hos Svensén) kun de leksikalsk betingede ordforbindelser ellers kaldes kollokationer.

”In English, as in other languages, there are many fixed, identifiable, non-idiomatic phrases and constructions. Such groups of words are called **recurrent combinations**, **fixed combinations**, or **collocations**. Collocations fall into two major groups: **grammatical collocations** and **lexical collocations**.” (ibid.)

Dette danner grundlaget for The BBI Combinatory Dictionary of English (BBI: 1986). Ordbogens indledning er en oversigt over de hyppigste grammatiske (G1-G8, 8 i alt) og leksikalske (L1-L7, 7 i alt) kollokationstyper som er medtaget i ordbogen. For verbers grammatiske kollokationer (G8) oprettes et **mønster** (”*pattern*”); et mønster er en unik kombination af prototypiske træk som en række verber har tilfælles. (I lingvistisk terminologi kaldes dette også **kategoriel** og **funktionel valens** eller **syntaktisk konstruktionsmønster**). Hvert mønster er formaliseret i en kode som anvendes i ordbogsdelen.

I beskrivelsen af de leksikalske kollokationer, som er den type ordforbindelse der almindeligvis betegnes som egentlige kollokationer, påpeger Benson (op.cit., p. xxiv) to forhold der er særligt vigtige i praktisk leksikografi.

I tilfælde af kollokationer der består af et (transitivt) verbum og et substantivsyntagme (eller et præpositionssyntagme), henvises der for det første til at valget af verbet er arbitrært og ikke forudsigeligt og derfor skal medtages i ordbogen. Et dansk eksempel herpå er de to semantisk set parallelle kollokationer

*tage* eller *træffe* en beslutning vs. (*\*tage*) *træffe* en afgørelse

For det andet er det, hvis der er tale om et flertydigt substantiv som kerne, vigtigt at anføre kollokationerne der knytter sig til de enkelte betydninger under opslagsordets pågældende betydning. Eksempelvis *affære*, med to groft skitserede betydninger, 1: ’sag, anliggende (ofte med negativ konnotation)’, kollokation: *tage affære*; 2: ’kærlighedsforhold uden for ægteskabet’ (med brugseksempel *have/indlede... en affære*).

#### 4. Kollokationsbegrebet i korpuslingvistik

I nyere tekstorienteret lingvistisk forskning – korpuslingvistik – har termen oprindeligt en bredere definition, nemlig to eller flere ords samtidige forekomst inden for en sætning. Denne definition indskrænkes dog i praksis af to kriterier: hyppighed (frekvens) og den afstand der er mellem de ord der er genstand for undersøgelsen. Sinclair (1991: 170) definerer kollokation på følgende vis:

”Collocation in its purest sense (...) recognizes only the lexical co-occurrence of words. This kind of patterning is often associated with grammatical choices as well, leading to the wealth of idioms and fixed phrases that are found in everyday English.”

Forholdet mellem kollokation og idiom er et af de centrale emner i korpuslingvistikken og beskrives som følger:

”An idiom is a group of two or more words which are chosen together in order to produce a specific meaning in speech or writing. (...) The individual words which constitute idioms are not reliably meaningful in themselves, because the whole idiom is required to produce the meaning. Idioms overlap with collocations, because they both involve the selection of two or more words. In principle, we call co-occurrences idioms, when we interpret the co-occurrence as giving a single unit of meaning. If we interpret the occurrence as the selection of two related words, each of which keep some meaning of its own, we call it collocation.” (op.cit. p.172)

Det er vigtigt at bemærke at i en korpuslingvistisk tilgang består en kollokation ikke af en a priori defineret basis og dens kollokat, men alt efter hvilket af ordene der er søgeord (*”node”*) eller genstand for undersøgelse og beskrivelse, tales der om det pågældende ords kollokater (*”collocates”*). Tabel 2 viser et par eksempler fra Korpus2000; de benyttede funktioner er *”vis nabo-ord”*, *”hyppighed”* og *”sortering”* på hhv. foregående og efterfølgende ord:

Søgeord	Kollokater (eksempler)	Bemærkning til betydning
begå (1)	aftalebrud, bedrageri, fejl, fejltagelse, hæleri, indbrud, lovbrud, mord, selvmord, svindel, tyveri, vold, voldtægt ...	objekt: negativ eller ulovlig handling
begå (2)	genistreg, straffespark, video, cd, libretto, forestilling	objekt: særlig positiv (kunstnerisk) præstation
fejl	begå, lave	gøre/lave uoverlagt handling
hærværk	begå, lave, forøve	udføre overlagt handling
indbrud	begå, lave	”
mord	begå	”
selvmord	begå	”
vold	begå, forøve, øve+, bruge	”
genistreg	begå (8), lave (1), gøre (1)	præstere (vurderende ytring)
affære (1)	tage+	(pinlig) sag, anliggende
affære (2)	have, indlede	kærlighedsaffære

Tabel 2. Eksempel på søgeord og deres hyppige kollokater (nabo-ord) i Korpus2000

Note: Tegnet '+' markerer sådanne ordkombinationer som registreres som ”**Vendinger**” i DDO (i form af sublemmaer) på grundlag af signifikant hyppig samforekomst af komponenterne i Korpus2000.

Korpuslingvistiske metoder, som beskrevet eksempelvis i Sinclair (1991) er særdeles nyttige i fraseologiforskning, idet de giver et pålideligt empirisk grundlag for leksikologiske undersøgelser. På den anden side er det indlysende at frekvente kombinationer af nabo-ord ikke i sig selv er **vendinger**. Et godt eksempel herpå er at ordet *affære* som typiske nabo-ord til venstre har *Clintons* og til højre *Monica Lewinsky*, hvilket afspejler faktiske men ikke sproglige relationer mellem ordene.

Sprogteknologiske ressourcer og værktøjer, såsom store elektroniske korpusser og konkordansprogrammer, programmer til automatisk analyse og opmærkning af tekster (parsere og taggere) og søgemaskiner på internettet kombineret med redigeringsværktøjer, er vigtige komponenter i moderne ordbogsarbejde. Konkordanser og frekvensundersøgelser inddrages i høj grad til at finde og klassificere ordforbindelser. I denne sammenhæng bør nævnes Det Danske Sprog- og Lit-

teraturselskabs efterhånden velkendte og meget benyttede ressourcer, Korpus2000 (jf. Asmussen 2001 og Asmussen 2004) og Korpus90 som er tilgængelige via internettet.

## 5. Kollokationer i leksikografisk praksis

Nedenfor vises et par eksempler på hvordan aktuelle danske ordbøger behandler fraseologisk materiale, med fokus på kollokationer.

### 5.1. Den Danske Ordbog (DDO 2003)

Den Danske Ordbogs (DDO) udgivelse er påbegyndt i 2003 og indtil videre (marts 2005) er udkommet 4 af ordbogens i alt 6 bind. Brugervejledningen i bind 1 (pp. 45-47) redegør for hvilke typer **faste udtryk** der er medtaget i ordbogen, hvor de er placeret i mikrostrukturen, samt hvordan de er beskrevet.

Idiomer defineres som faste udtryk der ikke har en kompositionel betydning, men er en helhed der har en **overført betydning** (markeret med *ofø*). Foruden idiomer og talemåder medtager ordbogen ”forskellige typer **faste udtryk uden overført betydning**”, listen omfatter flerordstermer, partikelverber, udtryk der indeholder fossilerede ordformer, refleksive verber og en kategori der beskrives som ”udtryk hvori indgår et tryksvagt verbum, hvis betydning ikke altid er forudsigelig, fx gå i arbejde og komme på benene”.

Det interessante er at selve termen ”kollokation” ikke er nævnt, dvs. at man ikke opererer med en kategori som er afgrænset i forhold til frie eller valensbundne ordforbindelser på den ene side og idiomer på den anden side. Men den anførte liste giver konkrete eksempler på de typer af ordforbindelser (”faste udtryk”) der ikke er markeret som idiomer. Det nævnes i forbindelse med to typer hvor et tryksvagt verbum indgår, at disse udtryk ikke altid er fuldt kompositionelle.

I modsætning til de ovennævnte typer behandles en række almindelige (frekvente) ordforbindelser af typen *begå en genistreg*, *træffe afgørelse*, *træffe/tage beslutning* ikke som faste udtryk, men sådanne, der ”(...) blot kan nævnes blandt **sprogbrugseksempler**” (op.cit. p.45). En anden måde at anføre sådanne ordforbindelser som er både hyppige og ganske faste på er at markere dem som **konstruktioner**, fx *gøre anskrig*.

Disse typer bestående af et tryksvagt verbum og et substantiv opfylder ellers et væsentligt kriterium med hensyn til fasthed, nemlig leksikalsk stabilitet, idet verbet ikke kan erstattes frit med et synonym, fx *\*udstøde* eller *\*råbe anskrig*. Disse typer bliver i de tidligere omtalte metaleksikografiske kilder betragtet som kollokationer, nemlig hos Svensén: "semantisk-lexikalsk kollokation" (1987: 95), i NLO: 'begrænset kollokation' (1997: 155) og hos Benson et al.: 'Lexical collocation 1/CA: (creation and/or activity) (1986: xxiv). I en frekvensbaseret opgørelse (fx på grundlag af Korpus2000) viser disse sig at være signifikant hyppige samforekomster og vil dermed ifølge Sinclairs tidligere citerede definition også falde ind under kollokationer.

Dermed kan det konkluderes at DDO er mere restriktiv end de ovennævnte kilder med hensyn til hvilke typer ordforbindelser der kategoriseres som faste uden at være idiomer (dvs. kollokationer). Nimb (1996) redegør for de leksikografiske overvejelser der ligger til grund for klassifikation og præsentation af de ovenfor nævnte verbum+substantiv-ordforbindelser og påpeger de principielle fordele samt mulige ulemper af de valgte løsninger for brugeren.

## 5.2. Den Danske Netordbog og Den Danske Idiomordbog

Disse to ordbøger, der er udarbejdet på Center for Leksikografi, Handelshøjskolen i Århus, er tilgængelige i elektronisk form via internettet (<http://www.dendanskenetordbog.dk> og <http://netdob.asb.dk>). De er komplementære mht. repræsentation af kollokationer og idiomer ifølge de to brugervejledninger på følgende vis:

Den Danske Idiomordbog (DIO 2003) indeholder udelukkende "**særlige ordforbindelser** (kaldet idiomer)", hvorimod Den Danske Netordbog (DNO 2002-2004) indeholder opslagsord og deres "**normale forbindelser** (kaldet kollokationer)".

Hvis man forgæves søger et udtryk i DIO, så kan man hurtigt slå det op i DNO og omvendt, idet der er direkte link mellem de to ordbøger. Denne tekniske løsning reducerer brugerens problemer noget mht. at finde oplysninger man har brug for i dette tilfælde, dvs. når fraseologien opdeles i kollokationer og idiomer og beskrives strengt adskilt. Denne form for leksikografisk løsning, hvor man fysisk helt adskiller de to centrale fraseologiske kategorier, ville være ganske uhensigtsmæssig i

trykte ordbøger. Det stiller også store krav til brugeren, især med henblik på fremmedsproglig tekstproduktion.

Vejledningen til DNO oplyser følgende om kollokationer:

”Eksempler på typiske ordforbindelser med opslagsordet som kerne. Angivelser af kollokationer er ment som hjælp til at vælge den ordforbindelse, som man har brug for, men ikke lige kan komme i tanke om eller ikke kender.”

Om idiomer står der:

”En fast ordforbindelse med opslagsordet som kerne, der kun kan bruges i særlige sammenhænge og som har en særlig betydning. I modsætning til kollokationer er idiomer næsten fastfrosne i brugen. Et idiom identificeres ved hjælp af følgende regel: Hvis summen af betydningen af de enkelte ord i en ordforbindelse ikke svarer til betydningen af hele ordforbindelsen, er der tale om en idiom (...). Når det derimod drejer sig om udtryk, der i sig selv ikke har en klar betydning, taler vi ikke om et idiom, fx gå fra snøvsen, hvor snøvsen i sig selv ikke har en klar betydning eller kan forstås som ’forstanden’ eller ’besindelsen’.”

I vejledningen til DIO står nogenlunde det samme, blot med den yderligere forklaring at

”Der foreligger en ikke helt klar adskillelse mellem det brede begreb ’fast forbindelse’ og idiom.”

I fortsættelsen argumenteres endvidere for at udtrykket *gå fra snøvsen* ikke er et idiom og at

”betydningen af hele udtrykket kan forstås som en sum af betydningen af de enkelte ord...”.

Konklusionen er at der synes at være principiel enighed mellem DDO, DNO og DIO om at et idioms betydning ikke kan forstås kompositionelt, hvorimod den praktiske grænsedragning mellem kompositionel og ikke-kompositionel betydning synes at være mere problematisk. Den leksikografiske behandling af de enkelte kategorier er følgelig også noget divergerende. Tabel 3 sammenligner nogle eksempler herpå.



Ordkombination	Korpus2000	DDO	DNO	DIO
tage <b>affære</b>	nabo-ord; vending	sublemma, fast udtryk	særligt sublemma (4) + dets kollokation!	intet
gøre <b>afbigt</b>	'for sjælden'	konstruktion + for ngt	særligt sublemma (2) + dets kollokation!	intet
afgå ved <b>døden</b>	'for sjælden'	sublemma, fast udtryk	idiom	idiom
gå på <b>aftægt</b>	'for sjælden'	brugseksempel	brugseksempel	intet
gøre <b>anskrig</b>	'for sjælden'	underbetydning ofø	(sub)lemma	intet
tabe <b>ansigt</b>	nabo-ord; vending	sublemma ofø	idiom	idiom

Tabel 3. Eksempel på leksikografisk behandling af ordforbindelser

## 6. Kollokationer set fra forskellige synsvinkler med fokus på specifikke træk

I de tidligere afsnit blev det vist at de lingvistiske træk der kendetegner kollokationer kan defineres og vægtes forskelligt i forhold til hinanden, afhængigt af tilgang og formål. Nedenfor gives et par eksempler på nogle ofte benyttede synsvinkler og deres fokus på specifikke træk, uden nærmere detaljer om de enkelte træks indhold og konkrete brug. Med synsvinkel menes her en bestemt måde at betragte emnet på, og formålet med denne oversigt er at vise bredden af mulighederne, derfor er billedet her også noget simplificeret og listen langt fra udtømmende.

Empirisk-statistisk baseret beskrivelse

- *kvantitativ synsvinkel* (baseret på målelighed af hvorvidt to eller flere ord optræder signifikant hyppigt sammen eller i nærheden af hinanden i konkrete tekster)
- *distributionel synsvinkel* (ords fordeling og typiske kontekst)

Lingvistisk beskrivelse

- *strukturel-semantisk synsvinkel*  
en frase der består af to ord eller ordgrupper med forskellig semantisk vægt, således at den ene del bærer hovedbetydningen og derfor kaldes den (semantiske) basis; hvorimod anden del har en karak-

teriserende eller modificerende funktion og kaldes kollokat (el. kollokator)

- *formrelateret synsvinkel* (kollokationens grundform og graden af dens morfosyntaktiske variabilitet)
- *leksikalsk-semantisk synsvinkel*  
kollokationens samlede betydning i forhold til betydningen af dens komponenter, kollokationens leksikalske stabilitet
- *typologisk synsvinkel*  
afgrænsning fra hhv. frie ordkombinationer og faste ordforbindelser,  
klassifikation mht. kollokators udskiftelighed, dvs. stabilitet,  
klassifikation mht. kollokators type og funktion
- *funktionel synsvinkel* (leksikalsk flerordsenhed)
- *produktionsrelateret synsvinkel*  
monolingvalt: en enhed der ikke kan produceres fuldt kompositionelt (arbitrært valg af kollokat, begrænset morfosyntaktisk variabilitet mv.),  
kontrastivt: en enhed der ikke kan oversættes fuldt kompositionelt (forskelle mellem sprog mht. valg af kollokat, idiosynkratisk morfologi mv.)
- *receptionsrelateret synsvinkel*  
graden af betydningens kompositionalitet (mindre problem i almensprog, større i fagsprog)
- *stilistisk synsvinkel*  
sammenhæng mellem kollokationstyper og deres brug mht. stilniveau.

Leksikografisk beskrivelse omfatter både empirisk og lingvistisk beskrivelse og fokuserer desuden på

- *selektionssynsvinkel*: hvilke kollokationstyper har leksikografisk relevans set fra ordbogstypen (og skal derfor medtages i ordbogen)
- *repræsentationssynsvinkel*: hensigtsmæssig placering i makrostruktur (evt. som opslagsord), i mikrostruktur, lingvistisk beskrivelse (grundform, brugseksempel, eksplicit markering af variabilitet/restriktioner mv.)

## 7. Leksikografisk beskrivelse med og for computere

Kollokationsforskning er et særligt aktuelt emne, fordi computeren er i stand til at håndtere meget store tekstmængder, hvilket giver korpusanalysen – og dermed korpuslingvistikken – helt nye dimensioner med hensyn til omfanget og gyldigheden af empiriske undersøgelser. Samtidig påvirker brugen af computere i leksikografi (som redigeringsværktøj) ordbøgernes kvalitet mht. konsistenstjek og konsekvens i udvalg og præsentation af leksikalsk materiale.

Datalingvistik er en særlig gren af lingvistikken, og på samme måde er dataleksikografi en gren af leksikografien, nemlig leksikografi for computere, dvs. til datalingvistiske og sprogteknologiske formål. I dataleksikografi er der fokus på egenskaber der er relevante til disse formål og som kan formaliseres – og derved også håndteres – i sådanne applikationer.

Leksikografi for computere har meget tilfælles med almen leksikografi, med den grundlæggende forskel at i en leksikalsk database er der næsten ingen begrænsninger for hvor mange ord og oplysningstyper der kan medtages. De traditionelle makro/mikrostrukturelle problemer kan løses teknisk på mange fleksible måder og henvisningsstrukturen styres sådan at der opnås en meget høj grad af konsistens. Der er til gengæld andre problemtyper: ords beskrivelse skal opdeles i formaliserbare og kombinerbare småbidder og struktureres sådan at de i kombination med hinanden beskriver hver enkelt leksikalsk enheds egenskaber udtømmende.

I denne sammenhæng er klassifikation og etiketter for de enkelte kategorier i princippet mindre væsentlige, det afgørende er, at etablerede ordforbindelser beskrives systematisk med hensyn til grundform, intern struktur, syntagmatisk og paradigmatiske variation og betydning (Braasch & Olsen 2002, Braasch 2005). Disse træk er ligeledes relevante for ordbøger til tekstproduktion på fremmedsprog, men af mange praktiske grunde er det ikke muligt at opfylde dette behov i sin helhed i almen leksikografi.

## 8. Afsluttende bemærkninger

Formålet med artiklen var at bidrage til diskussionen om brugen af fra-seologiske termer og begreber på dansk og at bidrage til udviklingen af det fælles diskussionsgrundlag der blev efterlyst ved indkaldelsen til

DANFRAS-Symposiet i november 2004. Det blev vist at afgrænsning og definition af kollokationer som fraseologisk kategori kan foretages ud fra en række synsvinkler der opererer med forskellige kriterier. Dette ses som en oplagt forklaring på termens divergerende brug både i teori og praksis; i sidste ende styres deres lingvistiske beskrivelse og leksikografiske præsentation i ordbøger af formålet

### Referencer

- Asmussen, Jørg 2001: Korpus 2000. Et overblik over projektets baggrund, fremgangsmåder og perspektiver. In *NyS 30. Nydanske studier & almen kommunikationsteori*. København.
- Asmussen, Jørg 2004: Korpus 2000 – til hvilken nytte? Muligheder og grænser for empiriske sprogundersøgelser. In Duncker, Dorthe (ed.): *Studier i Nordisk 2002-2003*. København: Selskab for Nordisk Filologi.
- Bahns, Jens 1996: Kollokationen als lexikographisches Problem. Tübingen: Niemeyer. (Lexicographica: Series maior 74).
- BBI 1986 = *The BBI Combinatory Dictionary of English* compiled by Benson, Morton, Benson, Evelyn/Robert Ilson. Amsterdam/Philadelphia: John Benjamins.
- Benson 1986 = *Introduction*. In BBI 1986, ix – xxx.
- Bergenholtz, Henning/Tarp, Sven 1994: *Manual i fagleksikografi*. Århus: Systime.
- Braasch, Anna/Olsen, Sussi 2000: Formalised Representation of Collocations in a Danish Computational Lexicon. In *The Ninth EURALEX International Congress, Proceedings, Vol. II*. Stuttgart.
- Braasch, Anna 2005: *Vi tager nogle kollokationer i øjesyn*. In Vatvedt Fjeld, Ruth & Dagfinn Worren (eds.) *Nordiske Studier i Leksikografi 7*. Oslo.
- Brink, Lars 2004: *Fraseologiske definitioner*. Unpubl. handout ved DANFRAS-Symposiet. Kbh., 23.-24. november 2004.
- Firth, J. R. 1957: *Modes of Meaning*. In *Papers in Linguistics*. London: Oxford University Press, 190 – 215.
- Nimb, Sanni 1996: *Collocations of Nouns: How to Present Verb-noun Collocations in a Monolingual Dictionary*. In *Euralex '96 Proceedings I*, 265-271. Göteborg: Göteborg University, Dept. of Swedish.
- NLO 1997 = Bergenholtz, Henning/Cantell, Ilse/Vatvedt Fjeld, Ruth/Gundersen, Dag/Jónsson, Jón Hilmar/Svensén, Bo: *Nordisk Leksikografisk Ordbog*. Oslo: Universitetsforlaget
- Sinclair, John 1991: *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Svensén, Bo 1987: *Handbok i leksikografi*. Stockholm: Esselte Studium & TNC.

**Ordbøger og korpus:**

DDO 2003 = *Den Danske Ordbog I-IV*. Det Danske Sprog- og Litteraturselskab. København: Gyldendal

DNO = *Den Danske Netordbog*. Center for Leksikografi, ASB. <http://netordbogen.dk>

DIO = *Idiomordbogen*. Center for Leksikografi, ASB. <http://netdob.asb.dk/idiom>

Korpus2000. Det Danske Sprog- og Litteraturselskab. <http://korpus.dsl.dk/korpus2000>

