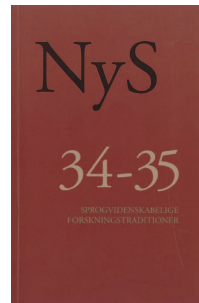


NyS

Titel: Formel Sproganalyse med Lexical-Functional Grammar
Forfatter: Bjarne Ørsnes
Kilde: *NyS – Nydanske Sprogstudier 34-35. Sprogvidenskabelige forskningstraditioner*, 2006, s. 53-91
Udgivet af: Multivers Academic
URL: www.nys.dk



© NyS og artiklens forfatter

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 NyS-numre (NyS 1-36) 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.

Formel Sproganalyse med Lexical-Functional Grammar

BJARNE ØRSNES

DATAMATISK SYNTAKSANALYSE MELLEM TEORI OG PRAKSIS

Der kan være mange gode grunde til at man gerne vil have computere til at analysere naturligt sprog. Man kan være interesseret i at udvikle egentlige applikationer som støtter både erfarne og mindre erfarne brugere i deres arbejde med sprog (oversættelse, grammatiktjek mv.), man kan bruge computere som værktøj til kvantitative og kvalitative undersøgelser af sprogligt materiale, og man kan bruge computere til at teste hypoteser ved at lave egentlige simulationer af menneskelig omgang med sprogligt materiale efter forudgående formalisering. Betegnelsen "datalingvistik" synes at blive brugt (en smule i flæng...) om stort set alle disse bestræbelser, og derfor er det undertiden meget vanskeligt at lave generaliserende udsagn om dette fagområde, ikke mindst fordi vi vel har en situation hvor "data" (i betydningen computere) indgår i større eller mindre omfang i alle videnskabsgrene. Dertil kommer at formaliserede sprogbeskrivelser typisk har en eller anden tilknytning til computere, hvorfor datalingvistik og formel lingvistik i Danmark har en tendens til at blive synonyme. Samtidig bliver formel lingvistik ofte synonymt med en chomskiansk version af syntaks, og ofte en chomskiansk opfattelse af ældre dato. Der er med andre ord en vis terminologisk forvirring, og derfor skal man nok være meget forsigtig med at udtale sig om datalingvistik som et sammenhængende fagområde.¹

Mødet mellem computere og (naturligt) sprog er et mangefacetteret og omfattende område, og jeg vil derfor i denne artikel kun koncentrere mig om et lille delområde, nemlig formaliserede beskrivelser af (dansk) syntaks i teori og praksis. Jeg vil koncentrere mig om en bestemt teori,

Lexical-Functional Grammar (Bresnan 2001; Dalrymple 2001). Lexical-Functional Grammar (LFG) er en teori om repræsentation og organisation af sproglig viden og dermed om hvordan sproget fungerer som system. Hertil anvender LFG et så veldefineret beskrivelsesapparat at beskrivelsen kan anvendes og bliver anvendt i datamatiske applikationer. Og ikke sjældent sker det at den datamatiske anvendelse sender lingvisten tilbage til skrivebordet igen.

LFG har sine rødder i den chomskianske lingvistik, men fandt – ligesom sin “bror” Head-Driven Phrase Structure Grammar (HPSG) – sin udformning i tæt samarbejde med industrielle virksomheder med interesse for at kunne processere naturligt sprog automatisk. På den måde har det implementeringsmæssige aspekt altid været centralt for LFG, om end teorien inden for den teoretiske lingvistik ofte anvendes på et rent deskriptivt, teoretisk plan.

Det vil nok være at gå en hel del for vidt at påstå at LFG er en central teori i dansk lingvistik. Alligevel er der god grund til at man som lingvist kan have glæde af at stifte bekendtskab med LFG. LFG ligger på mange måder i forlængelse af europæisk sprogvidenskabelig tradition (måske i særdeleshed valensgrammatikken) ved at operere med forskellige sproglige niveauer som hænger tæt sammen uden dog at stå i noget 1:1-forhold. Samtidig har det altid været væsentligt for LFG at finde frem til en arkitektur for den sproglige viden som netop giver plads til typologisk variation. Alt i alt er LFG et eksempel på at det man hertilands ofte forstår ved formel syntaks, er meget andet end den specifikke chomskianske teori i dens forskellige videreudviklinger.

Kendskab til LFG er samtidig en god investering idet LFG har vist sig at være en utroligt stabil teori. Der er ikke sket radikale ændringer i teoriens skelet siden den blev lanceret tilbage i 1982 – men den er blevet udvidet og forfinet på mange punkter. Og alligevel har LFG vist sig at kunne integreres i nyere lingvistiske retninger som ellers i mange henseender bryder med mere grundlæggende antagelser. Et eksempel er anvendelsen af LFG som teoretisk ramme inden for optimalitetsteori, som bryder med den opfattelse at grammatiske velformethedsbetingelser er absolutte sådan at forstå at de nødvendigvis skal være opfyldt. Et andet

eksempel er anvendelsen af LFG inden for statistisk baseret, data-orienteret parsing, som bryder med opfattelsen af sproglig viden som en regelbaseret grammatik og i stedet mener at sprogevnens baser sig på repræsentationer af tidligere input som så kan kombineres på nye måder og bruges til at analysere og generere nye sætninger.

Dertil kommer at LFG er en internationalt etableret teori. Ved at beskrive dansk inden for LFG opnår man at beskrivelserne bliver tilgængelige for et internationalt forum af lingvister. Den store mængde af grundige beskrivelser af andre sprog giver desuden en solid basis for at beskrive dansk og afdække dets ligheder og særheder i relation til andre sprog. Teorien giver en fælles referenceramme. Adgangen til eksisterende ressourcer har også været udgangspunktet for bestræbelserne på at udvikle en datamatisk LFG-grammatik for dansk da der allerede fandtes et internationalt projekt om parallel grammatikudvikling inden for LFG (Wedekind & Ørsnes 2004; Ørsnes 2004).

Vinklen i denne artikel er dog ikke kun at beskrive LFG, men også at beskrive udfordringerne og gevinsterne ved at arbejde med en formaliseret, implementeringsvenlig teori. Det vil således være en rød tråd gennem artiklen at teori og anvendelse mødes, og hvordan de mødes. Mødet mellem teori og data og mødet mellem computer og naturligt sprog fører på ingen måde altid til afhuggede hæle. Tværtimod fører det i høj grad til en øget forståelse af naturligt sprog.

LFG OG FAMILIEN AF "CONSTRAINT"-BASEREDE SYNTAKSTEORIER

Som antyd det ovenfor tilhører LFG en hel familie af syntaksteorier som ud over LFG selv tæller blandt andre Generalized Phrase-Structure Grammar (GPSG), Head-Driven Phrase-Structure Grammar (HPSG), som på mange måder er en videreudvikling af GPSG) og Tree-Adjoining Grammar (TAG). Disse teorier er vokset ud af den chomskyianske generative lingvistik, men har på forskellige områder distanceret sig fra Chomskys specifikke teorier, i særdeleshed Government and Binding med dens antagelse om at sproglige størrelser kan flytte sig mellem forskellige underliggende niveauer på vej til en overfladestruktur. Den væ-

sentligste forskel er (ud fra en syntaktisk betragtning) at disse nye teorier ikke skelner mellem dybde- og overfladestrukturer (som i de derivationselle teorier). Der er altså ikke tale om at en overfladestruktur deriveres fra en dybdestruktur, eller - omvendt - at man skal slutte sig tilbage fra en overfladestruktur til en underliggende struktur. Man kan på den måde sige at disse nye teorier tager overfladestrukturen meget alvorligt for så vidt som der stort set ikke postuleres tomme elementer og at sproglige størrelser ikke antages at flytte sig overhovedet. Der er kun ét syntaktisk niveau (som dog kan beskrives i flere parallelle dimensioner). Det betyder så samtidig at mange sproglige fænomener som tidligere blev beskrevet som forskelle mellem en dybde- og en overfladestruktur (fx passiv og raising), nødvendigvis må beskrives anderledes. Ét gennemgående træk ved disse teorier er at leksikon tillægges en meget stor betydning i beskrivelsen af også regelbundne sproglige fænomener (derfor omtales disse teorier ofte som leksikalistiske). De omtalte fænomener, passiv og raising, analyseres fx som valensalternationer, altså som leksikalske og ikke syntaktiske fænomener. Variation i konstituentrækkefølge som fx topikalisering analyseres ikke som en flytning, men som en værdideling mellem det topikaliserede led og dets valenssyntaktiske funktion: et led kan fx både være subjekt og topik på én gang. Denne form for værdideling udtrykkes ikke naturligt i en træstruktur, men derimod i en anden mere fleksibel datastruktur (mere herom senere). I det hele taget tillægger disse teorier ikke træstrukturer den samme dominerede rolle som fx Government & Binding og dens videreudviklinger. Hvor Government & Binding analyserer fx anaforisk binding ud fra strukturelle forhold i træstrukturen og definerer grammatiske funktioner ud fra positioner i træstrukturen, inddrager disse nye teorier andre typer af prominensrelationer som ikke er defineret ud fra træstrukturer. Det kan være relationer som at direkte objekter er mere prominente end indirekte objekter, hvilket igen kan afspejles i repræsentationen af de leksikalske størrelses argumentsstruktur. Også af den grund anvender disse teorier i højere grad andre og mere fleksible datastrukturer (såsom AVM-strukturen der beskrives senere), enten i stedet for træer (HPSG) eller som supplement til træer (LFG).

Opkomsten af disse nye formelt orienterede teorier er imidlertid ikke kun en reaktion på den specifikt chomskianske lingvistik, men også et resultat af udviklingen inden for datalogi. Teorier som HSPG og LFG opstod også ud fra ønsket om at kunne gøre den sproglige beskrivelse til genstand for datamatisk bearbejdning, dels som afprøvning af sproglige hypoteser, dels af egentligt anvendelsesorienterede hensyn. Udforskningen af formelle og naturlige sprogs matematiske egenskaber og også de datalogiske fremskridt inden for automatisk syntaktisk analyse af sprog har afgørende præget udviklingen af disse teorier.

Teorier som HPSG og LFG går under betegnelser som leksikalistiske teorier, unifikations-baserede eller "constraint"-baserede grammatikker. Den første betegnelse sigter som allerede nævnt til den rolle leksikon tillægges i sprogbeskrivelsen. Betegnelsen unifikations-baseret sigter til at teorierne gør udstrakt brug af unifikation som en grundlæggende operation til opbygning af information (se afsnittet om datamatiske grammatikker nedenfor). Unifikation er faktisk en væsentlig forudsætning for at kunne analysere en lang række fænomener som tidligere blev tilskrevet transformationer mellem en dybde- og en overfladestruktur. Endelig sigter betegnelsen "constraint"-baseret til måden at afgøre om en sproglig størrelse er velformet (eller grammatisk) på. Teorien opstiller en række betingelser som en sproglig størrelse skal opfylde for at være velformet. Grammatikalitet er således resultatet af at opfylde disse velformethedsbetingelser på én gang. Velformethedsbetingelserne kan være enten universelle eller sprogspecifikke, og det er selvsagt et væsentligt mål for teorien at afdække hvilke principper der er universelle, og hvilke der er sprogspecifikke.

Netop dette sidste punkt viser så også disse teories tilhørsforhold til den generative lingvistik. De forsøger at karakterisere den menneskelige sprogevne, og de gør det ved at opstille matematisk eksplicite modeller for forskellige aspekter af naturligt sprog. I de formelle modeller er det muligt at udtrykke og teste påstande om hvordan naturligt sprog fungerer. Og når disse påstande udtrykkes så præcist at det er muligt for en automat (fx i form af en computer) at beregne konsekvenserne af disse påstande, banes vejen for både en stor erkendelsesmæssig gevinst og for en egentlig anvendelse.

Syntaks har traditionelt spillet en central rolle i udforskning af den menneskelige sprogeevne. Det skyldes at syntaks netop synes at være et internt sprogligt fænomen, altså et fænomen der ikke som fx semantikken umiddelbart grænser op til kognitive systemer, eller som fonetik/fonologi til de motoriske systemer. Men formel lingvistik er i sig selv på ingen måde begrænset til syntaks, eller begrænset til en antagelse om at syntaksen skulle være fuldstændig uafhængig af semantik og pragmatik. Nærværende fremstilling vælger dog en noget traditionel tilgang og koncentrerer sig om syntaks.

DATAMATISKE GRAMMATIKKER

Datamatiske grammatikker skal typisk - ligesom et menneske - både kunne analysere og generere. Ved analyse tager grammatikken en natursprogsætning som input og producerer en mere eller mindre abstrakt repræsentation af den pågældende sætning. Ved generering tager den en (abstrakt) repræsentation som input og giver en natursprogsætning som output. Hvor abstrakte repræsentationerne er, og hvordan de er udformet, afhænger dels af grammatikkens anvendelse (og ambitionsniveau), dels af de grammatiske specifikationer som er lagt til grund for grammatikken. Afhængigt af anvendelsen kan man i nogle tilfælde være interesseret i rene semantiske repræsentationer, mens man i andre vil være tilfreds med syntaktiske repræsentationer. Her vil det primært handle om syntaktiske repræsentationer som udgangspunkt for hhv. analyse og generering.

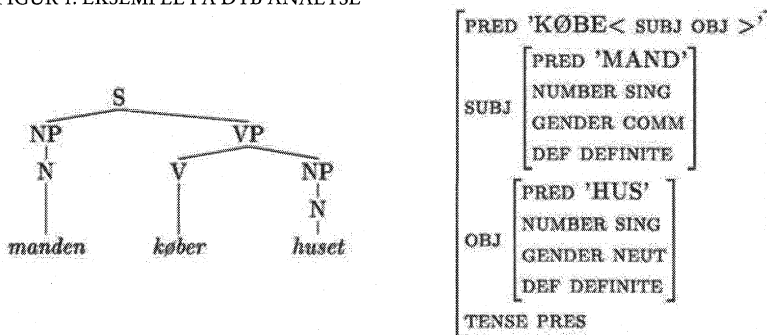
En datamatisk grammatik består af en mængde regler og et leksikon. Reglerne og leksikon processeres af en parser som er et program der foretager analyse og generering ved hjælp af en grammatik. Parseren stiller bestemte krav til grammatikken. For eksempel vil en parser baseret på Government & Binding processere de træstrukturer som er kendetegnende for denne teori, mens en parser for en HPSG-grammatik vil processere regler baseret på attribut-værdi-matricer som netop kendetegner HPSG (Butt & King 2003). At skrive en grammatik og at udvikle en parser er således to forskellige ting, og parsere kan være forskellige i hvor fleksible udtryksformer de tillader når lingvisten skriver grammatikker. Grammatikudviklingen er altså på den ene side bundet

til den bagvedliggende teori (eller de sproglige antagelser der ligger bag grammatikken) og på den anden side til udformningen af den konkrete parser. En af de p.t. mest avancerede platforme til grammatikudvikling er Xerox Linguistic Environment (XLE) som anvendes i det tidligere omtalte internationale projekt om udvikling af parallelle LFG-grammatikker. XLE er en omfattende grammatikudviklingsplatform som ikke kun indeholder en parser til avancerede LFG-grammatikker, men også faciliteter til morfologisk analyse, opbygning af leksika, prioritering af leksika i forskellige anvendelser mv. Endvidere kan XLE både bruges til analyse og generering.

Resultatet af den sproglige analyse kan være meget forskelligt, ikke mindst afhængigt af hvad analysen skal bruges til. Typisk skelner man mellem dyb analyse (*deep parsing*) og overfladisk analyse (*shallow parsing*). Ved dyb analyse forsøger man at finde en repræsentation for hele det sproglige input (fx en hel sætning), og analysen sigter mod at afdække så meget af inputtets syntaks som muligt. Det kan være konstituentstrukturen, de syntaktiske relationer mellem konstituenterne og eventuelle særlige ledstillingsmæssige forhold. Ved overfladisk analyse kan man være tilfreds med fx at få identificeret ordklasserne for de enkelte størrelser i inputteksten, eller med at få identificeret konstituenterne uden nødvendigvis at finde ud af hvordan de enkelte konstituenten forholder sig til hinanden (*chunking*). Jo dybere analysen er, jo mere vil man typisk være forpligtet på en bestemt teori om syntaktisk repræsentation, og en grammatik baseret på LFG producerer typisk en ret dyb analyse idet den som output giver både en konstituentstruktur og en repræsentation af de syntaktiske relationer som vist i figur 1. For overskuelighedens skyld er konstituentstrukturen i eksemplet vist som den velkendte "S består af NP og VP"-analyse, om end netop NP-VP-analysen har vist sig at være ret problematisk for dansk da den forudsætter at subjektet ikke kan stå imellem et verbum og dets eventuelle øvrige komplement. I de tilfælde hvor en anden konstituent end subjektet står i forfeltet, står subjektet imidlertid imellem verbet og dets øvrige komplement (*i dag [vasker Peter sin bil]*), og her er NP-VP-analysen altså ikke adækvat. På dansk, hvor stort set alle typer konstituenten som bekendt kan stå i forfeltet, er det altså nødvendigt at finde en anden strukturel beskrivelse (se senere). Det afgørende er imidlertid her at der

i denne type analyse (som er baseret på LFG) er to syntaktiske repræsentationsniveauer.

FIGUR 1. EKSEMPEL PÅ DYB ANALYSE



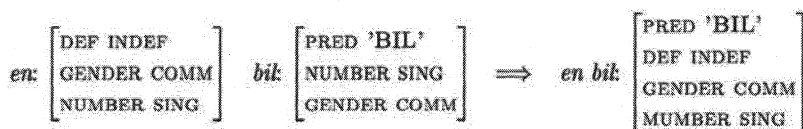
Det er inden for formel lingvistik meget almindeligt at beskrive konstituentstruktur ved hjælp af såkaldte kontekstfrie grammatikker. Kontekstfrie grammatikker er kendetegnet ved at bestå af en mængde regler hvor ét non-terminalt symbol på venstresiden (en moderknode i træstrukturen) genskrives som (eller forgrener sig i) et eller flere non-terminale eller terminale symboler på højresiden. For eksempel beskriver nedenstående lille (forenklede) kontekstfrie grammatik konstituentstrukturen (træstrukturen) i eksemplet i figur 1:

- S → NP VP
- NP → N
- VP → V NP

Det har længe været kendt at kontekstfrie grammatikker dels er vanskelige at bruge til mere indviklede sproglige fænomener, dels ikke kan beskrive alle aspekter af sprogs syntaks. I LFG udvides grammatikken derfor med træk-strukturer. Det er således en træk-struktur der i eksemplet i figur 1 viser de syntaktiske relationer mellem konstituenterne, altså fx at de to NP'er fungerer som hhv. subjekt og objekt for prædikatet. Træk-strukturer er mængder af attribut-værdi-par (fx TENSE PRES), og det er en meget fleksibel datastruktur til repræsentation af mange forskellige typer sproglig information.

Fleksibiliteten består bl.a. i at træk-strukturer kan kombineres med hinanden ved unifikation hvorved der opbygges information fra forskellige kilder. Unifikationen af to træk-strukturer er den træk-struktur som indeholder præcis al information fra begge de to træk-strukturer, hverken mere eller mindre. For eksempel vil unifikationen af en træk-struktur for artiklen *en* med en træk-struktur for substantivet *bil* give en samlet træk-struktur for hele frasen *en bil* som vist i figur 2.

FIGUR 2. UNIFIKATION AF TRÆK-STRUKTURER



Den samlede træk-struktur kombinerer således informationen fra de forskellige sproglige elementer i frasen og samler den til et hele. Substantivet er fx i sig selv underspecificeret for bestemthed, og denne oplysning leveres af artiklen. Til gengæld har både artikel og substantiv oplysninger om køn (GENDER) og tal (NUMBER), og kombinationen af disse oplysninger fungerer som et kongruenstjek idet unifikation kun kan kombinere træk-strukturer uden modstridende information. Hvis artiklen bærer specifikationen [GENDER NEUT] som i nominalfrasen **et bil*, vil informationen ikke kunne kombineres da trækket GENDER så både skal være COMM og NEUT på én gang.

I LFG-grammatikker indgår altså både kontekstfrie grammatikregler og træk-strukturer med tilhørende unifikation idet de kontekstfrie grammatikregler er annoteret med oplysninger om de tilhørende træk-strukturer. En LFG-grammatik kan derfor betydeligt mere end en grammatik der kun består af kontekstfrie regler, og en væsentlig del af at forstå LFG er at forstå samspillet mellem disse to strukturer som samtidigt er samspillet mellem konstituentstruktur og funktionel struktur. Konstituentstrukturen kan fx være Diderichsens feltskema, men feltskemaet skal jo suppleres med nogle begrænsninger såsom "hvis det direkte objekt står i forfattet, må pladsen for det direkte objekt i indholdsfeltet ikke samtidig være udfyldt" – og her kommer den funktio-

nelle struktur så netop ind i billedet. Præcist hvordan det foregår, vises senere.

REPRÆSENTATIONSNIVEAUER I LFG

LFG skelner altså mellem hvad man kan kalde “ydre” syntaks (c-struktur for *constituent-structure*) og “indre” syntaks (f-struktur for *functional structure*) som to parallelle dimensioner af syntaktisk repræsentation. Denne skelnen mellem en c-struktur og en f-struktur er bl.a. begrundet i typologiske forskelle: Mens sprog er meget forskellige i konstituentstrukturen (overfladestrukturen), synes de at ligne hinanden i behovet for at kunne identificere relationelle kategorier som subjekt, objekt osv. For eksempel har mange sprog anaforer som skelner mellem subjekter og ikke-subjekter når de skal finde deres antecedent, jf. forskellen mellem *Peter gav ham sin billet* og *Peter gav ham hans billet*. Anaforen *sin* viser tilbage til et subjekt (her *Peter*), og iflg. LFG er denne reference til et subjekt uafhængig af hvordan et subjekt realiseres i c-strukturen, om det er med en fast position eller med kasusmarkering. Der er med andre ord syntaktiske processer som angår egenskaber i f-strukturen, mens der omvendt også er processer som angår forhold i c-strukturen. For eksempel kan man i Korpus2000 se at når det danske verbum *at bestå* kombineres med et præpositionalobjekt indledt med *af*, er styrelsen næsten altid en nominalfrase som i (1) og ikke en ledsætning som i (2). Denne skelnen angår syntaktisk materiale og hører til i c-strukturen og ikke i f-strukturen som interesserer sig for de syntaktiske funktioner (i både (1) og (2) er der tale om et præpositionalobjekt uanset hvordan styrelsen for præpositionen ser ud).

- (1) *opgaven består af tre dele*
- (2) **opgaven består af at du skal løse tre dele*

Netop fordi sprogs c-strukturer er meget forskellige, mens f-strukturer neutraliserer overflademæssige forskelle, er det afgørende at man kan nå frem til disse mere generelle f-strukturer på forskellige måder. Et klassisk eksempel er identifikationen af de syntaktiske funktioner. I nogle sprog er syntaktiske funktioner som subjekt og objekt bundet til bestemte positioner i c-strukturen, og ledstillingen er dermed mere fast

(såkaldte konfigurationelle sprog). I andre sprog markeres subjekt og objekt morfologisk, og ledstillingen er i reglen friere (ikke-konfigurationelle sprog). C-strukturerne vil her igen være forskellige, mens f-strukturerne vil ligne hinanden idet de identificerer subjekt og objekt uden hensyntagen til hvordan disse funktioner kodes i det pågældende sprog.

C- og f-strukturerne eksisterer parallelt i kraft af en korrespondens-relation. Der er et nøje defineret, ikke-arbitrært forhold mellem c- og f-strukturen, og det er lingvistens opgave dels at definere et sprogs c- og f-strukturer dels at definere afbildningen fra c-struktur til f-struktur. LFG's arkitektur med afbildning af strukturer på andre strukturer åbner mulighed for at definere andre strukturer end kun c- og f-strukturer. Man kan for eksempel også definere en m(orfologisk)-struktur, en s(emantisk)-struktur og en i(nformations)-struktur.

Ud fra et parsing-synspunkt forsøger en LFG-grammatik altså at finde et sprogligt udtryks c- og f-struktur, mens generering typisk tager udgangspunkt i en f-struktur og afbilder den på en c-struktur. Hvis man er interesseret i en semantisk fortolkning, vil udgangspunktet typisk være en f-struktur med dens prædikat-argument-relationer, og inden for fx maskinoversættelse kan f-strukturen anvendes om en slags interface-repræsentation mellem kilde- og målsprog.

KONSTITUENTSTRUKTUREN (C-STRUKTUR)

Konstituentstrukturer er meget forskellige, og LFG åbner mulighed for at c-strukturer netop kan være af forskellige typer. Det er et mål for teorien at afdække den mulige variation inden for c-strukturer fordi det i sidste instans siger noget om sprog generelt, men beskrivelsen må bero på syntaktisk argumentation på det pågældende sprogs præmisser. Der er således ud fra LFG-teorien principielt ikke noget til hinder for at antage at feltskemaet er en adækvat repræsentation af dansk konstituentstruktur. Når det alligevel sjældent ses at feltskemaet anvendes som c-struktur, skyldes det at der er beskrivelsesmæssige fordele ved at antage at dansk har en endocentrisk frasestruktur i overensstemmelse med X-bar-teorien (som fx i Vikner 1995 – se endvidere nedenfor). Ved en

endocentrisk frasestruktur forstås at fraser har kerner som giver dem deres karakteristiske syntaktiske egenskaber. De væsentligste forskelle mellem en endocentrisk frasestruktur og for eksempel feltskemaet er at feltskemaet ikke har nogen entydig kerne-position. Verber kan stå på flere forskellige pladser. Dertil kommer at feltskemaet tildeler sætninger en flad struktur hvor der fx ikke forekommer en VP (altså en enhed af verbet og dets komplement eksklusiv subjektet). På den måde er det meget vanskeligt at indplacere VP-koordination som i *Peter [[gør rent] og [rydder op]]* i feltskemaet.

X-bar-teorien er en teori om at alle fraser, uanset om det er NP'er, AP'er, PP'er osv., har den samme grundlæggende struktur. Hele frasen er bygget op om en kerne, der eventuelt vælger nogle komplement, og som derudover kan forekomme med en specifikator, jf. følgende eksempler.

- (3) NP: *en aflysning af mødet*
- (4) AP: *dybt afhængig af piller*
- (5) PP: *ude i skoven*

Kernen i nominalfrasen (NP) i eksempel (3) er substantivet *aflysning*. Til højre for kernen forekommer komplementet *af mødet* og til venstre for kernen forekommer en specifikator, nemlig determinativet *en*. Mønsteret er præcist det samme i adjektivfrasen (AP) i (4). Kernen er adjektivet *afhængig*, til højre for kernen forekommer komplementet *af piller* og til venstre specifikatoren *dybt*. På samme måde er præpositionen kerne i præpositionalfrasen (PP) i (5) med komplementet *skoven* og specifikatoren *ude*. X-bar-teorien antager at når en leksikalsk kerne kombineres med sine eventuelle komplement, danner den en frasetype af bar-niveau 1, også skrevet X', mens den leksikalske kerne selv er af bar-niveau 0, ofte angivet som fx N⁰ (for nomen) eller A⁰ (for adjektiv). Argumentet for at antage et bar-niveau 1 er at disse konstituent for eksempel kan koordineres, som i *en [aflysning af mødet] og [udsættelse af beslutningen] kommer ikke på tale*. Her er to N'-konstituent blevet koordineret (*aflysning af mødet* og *udsættelse af beslutningen*) mens determinativet *en* er determinativ for dem begge. Kombinationen af substantivet *aflysning* og komplementet *af mødet* giver altså en N' (N-bar): *aflysning af mødet*. Kombinationen af adjektivet *afhængig* og et

komplement *af piller* danner en A': *afhængig af piller*, og kombinationen af *i* og *skoven* danner en P': *i skoven*. En konstituent på bar-niveau 1 kan så kombineres med en specifikator (*specificier*) og danne en maksimal frase af bar-niveau 2, angivet som hhv. N'', A'', P'' etc. – eller hyppigere som NP, AP eller PP. N''-strukturen *aflysning af mødet* kombineres således med en specifikator *en* og danner NP'en *en aflysning af mødet*. A''-strukturen *afhængig af piller* kan kombineres med en specifikator som adverbiet *dybt* og danne AP'en *dybt afhængig af piller*, og P''-strukturen *i skoven* kan kombineres med specifikatoren *ude* og danne PP'en *ude i skoven*. Da disse strukturer er bygget over samme skelet eller skabelon, og da denne struktur som sagt antages at gælde for alle frasetyper, anvendes ofte betegnelsen X-bar, hvor X forstås som en variabel over mulige syntaktiske kategorier som netop N, A eller P. Det særlige ved X-bar-teorien er altså at den antager en fælles skabelon for alle frasetyper som vises skematisk i figur 3.

FIGUR 3. X-BAR-STRUKTUREN

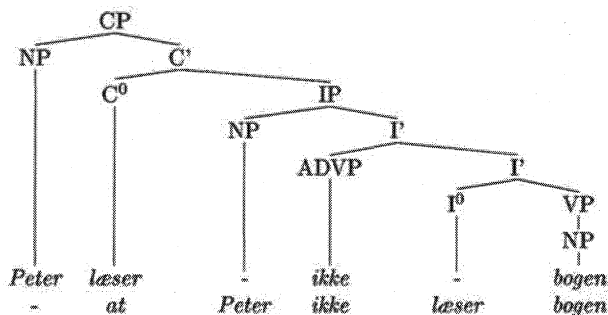


X-bar-strukturen som skildret ovenfor tager ikke højde for at der i (næsten) enhver frasetype kan forekomme et uforudsigeligt antal frie angivelser som i *en aflysning af mødet [i morgen] [på grund af utilstrækkelig deltagelse]*. For at tage højde for sådanne frie angivelser (også kaldet *adjuncts*) indføres i strukturen en særlig regel (*adjunction rule*) der ikke "tæller op" i bar-niveauerne. Som vist i figur 3 giver kombinationen af en adjunct og en X' en ny X'. I NP'en *en aflysning af mødet [i morgen] [på grund af utilstrækkelig deltagelse]* vil de to frie angivelser altså blive tilføjet strukturen som søstre til N'-knuder.

Hvis man følger X-bar-strukturen slavisk, vil en sætning som *Peter ser filmen* faktisk være en VP. Verbet *ser* danner med objektet *film* en V' og kombineres med subjektet *Peter* på specifikationspladsen til en VP.

Spørgsmålet er så hvor fx underordnende konjunktioner skal stå som i sætningen *at Peter ser filmen*. Netop disse konjunktioner er motivationen for at antage en særlig sætningsfrase “oven over” VP hvor konjunktionen er kerne. Denne frasetype kaldes CP ud fra den engelske betegnelse for konjunktioner (*complementizer*). Konjunktionen står i C og kombineres med sit komplement (sætningen *Peter ser filmen*) til en C'. I reglen er specificier-pladsen i sådanne ledsætninger så ikke udfyldt, men det kan forekomme, fx hvis ledsætningen indeholder et hv-ord som i *jeg ved ikke [CP [DP hvem] [C Peter har mødt]]*. Her mangler så til gengæld en konjunktion i C, men det kan faktisk også forekomme, fx i det noget talesprogsprægede: *jeg ved ikke [CP [DP hvem] [C [C at] [Peter har mødt]]]*. Ud over CP-frasen er det så også almindeligt at antage en særlig IP-frase (*inflectional phrase*) som vist i figur 4 nedenfor. Her er IP kategorien for finite ledsætninger således at det finite verbum i ledsætninger står i I (med subjektet på specificier-pladsen). Generaliseringen er så at et finite verbum altid står i enten C eller I, mens V er forbeholdt infinitive verbalformer. Der findes imidlertid også argumenter for at antage at I aldrig udfyldes i dansk (Vikner 1995).

FIGUR 4. DANSK SÆTNINGSSTRUKTUR



Som bekendt antager man i reglen to basale ledstillingstyper i dansk: ledfølgen i hovedsætninger og ledfølgen i ledsætninger (der dog også omtales som det deklarative og det neutrale mønster da ledfølgerne ikke er bundet til hoved- eller ledsætninger (Heltoft & Hansen 2003)). Forskellen ses tydeligst i placeringen af sætningsadverbialer i forhold til det finite verbum: I det deklarative mønster står sætningsadverbialer til højre for det finite verbum, i det neutrale mønster står de til venstre for det finite verbum (jf. placeringen af *ikke* i eksemplet i figur 4). Ved at

antage en CP-struktur følger de to mønstre faktisk af hvordan man udfylder kerne-pladsen C. Enten har man en underordnende konjunktion, og vi får så det neutrale mønster med verbet efter et eventuelt sætningsadverbial, eller også har man det finitte verbum i C, og vi får så det deklarative mønster med det finitte verbum før et eventuelt sætningsadverbial. Den underordnende konjunktion og det finitte verbum konkurrerer så at sige om den samme plads, og derfor udelukker de gensidigt hinanden: Enten har man det ene mønster, eller også har man det andet (Vikner 1995). Strukturen for hoved- og ledsætninger er illustreret i figur 4.

Der er altså en række forhold som gør en X-bar-orienteret, endocentriske struktur attraktiv ved beskrivelsen af dansk c-struktur. For det første motiverer ovenstående struktur at dansk har det finitte verbum på 2. pladsen i fremsættende hovedsætninger. X-bar-strukturen åbner kun mulighed for én position til venstre for kernen (nemlig specificer-positionen i CP, som altså her svarer til forfeltet), og når der kun er én position, kan der kun være én konstituent til venstre for det finitte verbum. For det andet er det muligt at passe både hoved- og ledsætningsstrukturen ind i én og samme struktur. Forskellen er blot hvad der står i C-knuden, som derved får en ganske afgørende rolle ved bestemmelsen af hvad det er for en type sætning der er tale om. Hvis C udfyldes af et finit verbum, får vi hovedsætningsordstilling. Hvis C udfyldes af en konjunktion, får vi ledsætningsordstilling. For det tredje giver X-bar-strukturen også en struktur for andre frasetyper end lige sætningen, nemlig AP, PP mv. I den danske datamatiske LFG-grammatik er der derfor netop valgt en X-bar-orienteret struktur (Wedekind & Ørsnes 2004).

Der er imidlertid nogle forskelle mellem X-bar-strukturen som den anvendes i LFG og som den anvendes i andre teoretiske rammer som fx Government & Binding. C-strukturerne i LFG er orienterede mod overflade-realisationen, og af den grund forekommer der ingen tomme knuder (eller kun i meget sjældne tilfælde) som det ellers kendes fra Government & Binding. Det betyder blandt andet at man kan få noget uortodokse C-strukturer. I eksemplet i figur 4 forekommer fx et VP som direkte dominerer en NP (*bogen*). VP'et mangler så at sige en V-datter og dermed altså en kerne idet hovedverbet (*læser*) befinder sig i C. Denne

type af kerne-løse fraser er tilladt i LFG, men de er underlagt strenge restriktioner (Bresnan 2001).

FUNKTIONSSTRUKTUREN (F-STRUKTUR)

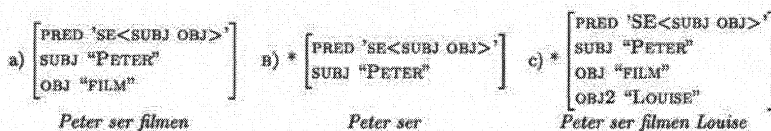
F-strukturen er en repræsentation af de basale syntaktiske relationer mellem konstituenterne. I og med at f-strukturen abstraherer fra sprogspecifikke overfladerealisationer, er f-strukturen ikke repræsenteret ved en træstruktur (som viser en sprogspecifik linearisering af konstituenterne), men derimod ved en mængde af træspecificationer. F-strukturen indeholder træk som SUBJ(ekt) og OBJ(ekt), og LFG skylder en del af sit navn (*functional*) at grammatiske funktioner forstås som en integreret del af teorien. Subjekt og objekt er altså ikke blot andre navne for bestemte positioner i en træstruktur, således som det fx kendes fra Government & Binding. Denne opfattelse af de grammatiske funktioner som primitiver er netop motiveret af at nogle sprog ikke synes at have en konfigurationskodning af grammatiske funktioner, men netop identificerer funktionerne på andre måder, fx ved morfologisk markering eller ved en kombination af ledstilling og morfologisk markering.

F-strukturen med dens mængde af attribut- og værdi-specificationer er underlagt en række velformethedsbetingelser. En f-struktur skal opfylde betingelserne om UNIQUENESS, COMPLETENESS og COHERENCE. Ved UNIQUENESS forstår man at et attribut kun kan tilskrives én værdi. Et eksempel herpå blev givet ovenfor hvor frasen **et bil* netop ikke havde nogen velformet f-struktur fordi man forsøger at tilskrive to værdier til samme træk, nemlig [GENDER NEUT] og [GENDER COMM]. Det vil også være UNIQUENESS der udelukker at en sætning indeholder to (ikke-koordinerede) subjekter eller to (ikke-koordinerede) direkte objekter.

Betingelserne COMPLETENESS og COHERENCE sikrer at en prædikators valenskrav overholdes. En prædikators valenskrav angives i PRED-værdien, der forstås som en semantisk værdi kombineret med en række syntaktiske funktioner som prædikatoren skal kombineres med for at give en grammatisk frase (typisk en sætning). LFG antager således

at en delmængde af de grammatiske funktioner er såkaldte *governable relations*, altså funktioner der forekommer i PRED-værdier. Den leksikalske indgang for verbet *at se* specificerer at verbet skal kombineres med et SUBJ(ekt) og et OBJ(ekt). Det vil sige at f-strukturen hvori *se* indgår som prædikator, skal indeholde et subjekt og et objekt for at opfylde COMPLETENESS, mens f-strukturen omvendt ikke må indeholde flere syntaktiske funktioner end krævet af verbet for at opfylde COHERENCE. F-strukturen i a) i figur 5 er således velformet, mens f-strukturen i b) i figur 5 er INCOMPLETE og f-strukturen i c) i figur 5 er INCOHERENT. Anførselstegn viser f-strukturer hvor attribut-værdi-notationen er udeladt for overskuelighedens skyld.

FIGUR 5. VELFORMETHEDSBETINGELSER FOR F-STRUKTURER



Netop på grund af disse velformethedsbetingelser skelner LFG (som mange andre) skarpt mellem valensbundne og ikke-valensbundne funktioner (eller i LFG-terminologi: styrede og ikke-styrede funktioner). Valensbundne funktioner optræder i PRED-værdien for prædikatorer og er underlagt principperne om COMPLETENESS og COHERENCE. Ikke-valensbundne funktioner, typisk frie angivelser kaldet ADJUNCTS, er netop ikke underlagt COMPLETENESS og COHERENCE da de ikke kræves af prædikatoren. ADJUNCTS-trækket er stadig underlagt UNIQUENESS, men da der som bekendt kan forekomme flere frie angivelser, er værdien for dette træk en mængde af f-strukturer som vist i figur 8 hvor ADJUNCTS-trækket er en mængde med to elementer: to f-strukturer for to frie angivelser.

For at være grammatisk skal en sætning med andre ord have en velformet c-struktur og en velformet f-struktur. Netop fordi valenskrav først undersøges i f-strukturen, kommer f-strukturen til at fungere som en

slags filter på c-strukturer. En sætning som *Peter ser filmen Louise* kan godt tilskrives en c-struktur, men ikke en f-struktur da der er for mange konstituenten i forhold til prædikatets krav (jf. c) i figur 5). Man kan på den måde sige at også feltskemaet mangler en f-struktur der forhindrer at for mange af felterne udfyldes. Intet i feltskemaet forhindrer at det udfyldes med både direkte objekt, indirekte objekt, objektsprædikat og præpositionalobjekt, andet end at ingen prædikator sandsynligvis accepterer så mange led.

En væsentlig del af arbejdet i LFG går således ud på at afgøre hvordan f-strukturer ser ud, og hvilke kategorier der er relevante for f-strukturer – ikke mindst fordi f-strukturer antages at være fælles for alle sprog.

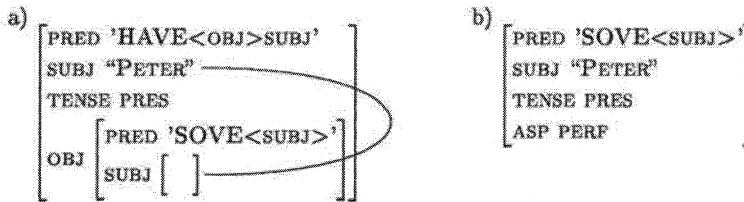
Et eksempel på at det kan være svært at trække grænsen mellem sprogspecifikke og sprogneutrale f-strukturer, er repræsentationen af komplekse verbalformer som i

(6) *Peter har sovet*

Her kan man analysere den infinitte verbalform *sovet* som en slags objekt for hjælpeverbet *har*. Både verbet *har* og verbet *sovet* kræver et subjekt, men de deler det samme subjekt (*Peter*) således at hjælpeverbet så at sige gør hovedverbets subjekt til sit eget (også kendt som *raising*). Denne analyse giver en kompleks f-struktur som vist i a) i figur 6 hvor hjælpeverbet leverer den “øverste” PRED-værdi i sætningens f-struktur. Man kan også vælge at lade den infinitte verbalform levere den “øverste” PRED-værdi (den er jo trods vigtigst ud fra et semantisk synspunkt) og gøre hjælpeverbet til en funktional kategori der kun bidrager med oplysninger om tempus og aspekt. En sådan flad f-struktur er vist i b) i figur 6. Strukturen i a) har den fordel at den tager højde for at den infinitte verbalform opfører sig som én konstituent ud fra de velkendte kriterier om permutation og substitution. Strukturen i b) har den fordel at den kun repræsenterer de egentlige indholdsord som PRED-værdier og dermed kan fungere som input til en semantisk komponent. Dertil kommer at nogle sprog realiserer temporale forhold med morfologiske/syntetiske former frem for perifrastiske former som den danske

perfektum, og strukturen i b) vil således være en fælles repræsentation på tværs af forskellige sprogs eventuelle præferencer for syntetiske eller perifrastiske verbalformer. Spørgsmålet er altså hvor universelle og hvor semantiske f-strukturer er eller bør være.

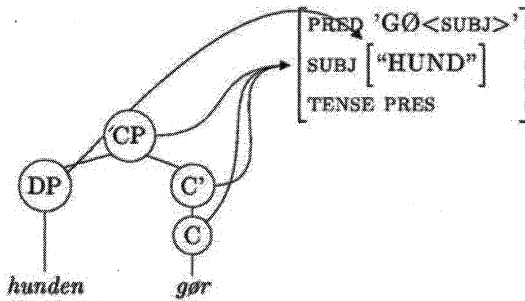
FIGUR 6. F-STRUKTUR-REPRÆSENTATIONER FOR KOMPLEKSE VERBALFORMER²



FORHOLDET MELLEM C- OG F-STRUKTUR

C-strukturen og f-strukturen står i et forhold af korrespondens til hinanden. Figur 7 viser hvorledes c-strukturen afbildes på f-strukturen i en ganske enkel dansk hovedsætning.

FIGUR 7. AFBILDNINGEN AF C-STRUKTUR PÅ F-STRUKTUR



F-strukturen for den spidsstillede nominalfrase (der nu i reglen analyseres som en determinatorfrase, en DP) indgår som værdi for SUBJ(ektet), og pilen peger således fra DP-knuden til f-strukturen der er værdi for SUBJ. Knuderne CP, C' og C er alle kerner i en struktur hvor DP'et står på specificier-pladsen (jf. figur 3), og derfor bidrager de med alle deres oplysninger til den samlede f-struktur for hele sætningen. Pilene pe-

ger således fra kerne-elementerne (CP, C' og C) til den yderste kantede parentes i sætningens f-struktur.

Selve afbildningen af c-struktur-knuder på f-strukturen gives (primært) som en form for instrukser (funktionelle annotationer) på de grammatikregler der beskriver c-strukturen. En (foreløbig) version af CP-reglen fra eksemplet i figur 7 kunne se ud som vist nedenfor:

$$\begin{array}{ccc} \text{CP} & \rightarrow & \text{DP} \quad \text{C}' \\ & & (\uparrow\text{SUBJ})=\downarrow \quad \uparrow=\downarrow \end{array}$$

Pilene beskriver hvordan der opbygges en f-struktur svarende til den c-struktur reglen beskriver. \uparrow -pilen refererer til f-strukturen for hele træet (med CP-knuden som rodkategori) idet pilen "peger på" og dermed refererer til moderknuden CP. \downarrow -pilen refererer til f-strukturen for den aktuelle knude. \downarrow -pilen under DP (det første symbol på højresiden af genskrivningspilen " \rightarrow ") refererer således til f-strukturen for DP'et, mens \downarrow -pilen under C' (det andet symbol på venstresiden) refererer til f-strukturen for C'. Den første annotation læses altså således at der i f-strukturen for hele sætningen CP indgår et subjekt, og det subjekt udgøres af f-strukturen for den aktuelle knude, nemlig DP. I en sætning som $[_{CP} [_{DP} \textit{hunden}] [_{C'} [_{C} \textit{gør}}]]]$, vil f-strukturen for $[_{DP} \textit{hunden}]$ således indgå som subjekt (som vist i figur 7).

Annotationen på C'-knuden ($\uparrow=\downarrow$) kan (mere proceduralt) læses som en instruks om at sende alle oplysninger videre til f-strukturen for moderknuden, altså til f-strukturen for hele sætningen CP. Den betyder at f-strukturen for moderknuden CP er den samme som f-strukturen for C', og dermed kommer alle oplysninger i f-strukturen for C' automatisk til at tilhøre f-strukturen for CP. I og med at C' giver alle sine særlige kendetegn videre til hele frasen (CP) bliver den altså til kerne i CP'et. I eksemplet med $[_{CP} [_{DP} \textit{hunden}] [_{C'} [_{C} \textit{gør}}]]]$, betyder det at verbet under c' sender alle sine oplysninger videre til hele sætningens f-struktur, og det svarer til at verbet er kernen i sætningen. En knude med annotationen $\uparrow=\downarrow$ bliver på den måde en slags "transit"-knude.

Denne (forenklede) regel er et eksempel på at subjekt-funktionen identificeres konfigurationelt, altså som en knude i en bestemt position i c-strukturen. I sprog med morfologisk identifikation af subjektet (typisk markering med nominativ kasus) kunne en f-struktur-annotation se således ud:

$$(\downarrow \text{CASE}) = \text{NOM} \Rightarrow (\uparrow \text{SUBJ}) = \downarrow$$

Denne implikation er opfyldt hvis første del af den er sand, det vil sige at hvis f-strukturen indeholder trækspesifikationen [CASE NOM], indgår den pågældende f-struktur som SUBJ i moderknudens f-struktur.

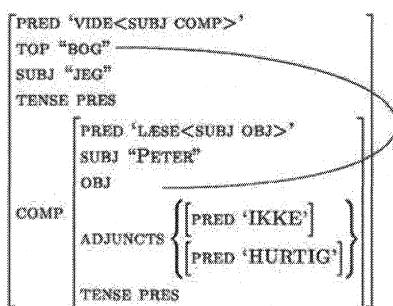
LFG-formalismen hvori LFG-grammatikker skrives, er et helt sprog til beskrivelse af f-strukturer og ikke mindst til beskrivelse af afbildningen af c-struktur på f-struktur. Det er lingvistens opgave at fastlægge forholdet mellem c- og f-struktur i et givet sprog og siden udtrykke det i den tilhørende formalisme. Formalismen har så den store fordel at den kan processeres af en parser.

Naturligvis kan man på dansk ikke automatisk gå ud fra at den første konstituent altid er subjekt som vist i eksemplerne nedenfor.

- (7) *Peter læser ikke den bog hurtigt* (subjekt i forfelt)
- (8) *den bog læser Peter ikke hurtigt* (objekt i forfelt)
- (9) *den bog ved jeg Peter ikke læser hurtigt* (objekt fra indlejret ledsætning i forfelt)

Eksemplerne viser topikalisering af forskellige konstituer, og det er f-strukturens opgave at repræsentere denne topikalisering. I eksempel (9) er den topikaliserede konstituent (TOP) samtidig objekt i den indlejrede sætning (COMP), og dette forhold vises ved at de to attributter deler værdi (vist ved buen i f-strukturen i figur 8).

FIGUR 8. F-STRUKTUR FOR EKSEMPEL (9)



Denne identifikation af den topikaliserede konstituent med objektet i den indlejrede sætning sker i grammatikreglerne, men da det ikke er muligt på forhånd at opregne alle de funktioner en topikaliseret konstituent kan have (principielt er der ingen grænser for hvor dybt indlejrede disse funktioner kan være), er man nødt at udtrykke relationen med "funktionel usikkerhed" (*functional uncertainty*). Man overlader det så til COMPLETENESS og COHERENCE at finde ud af præcist hvilken funktion der kan være tale om. Annotationen på første konstituent i CP-reglen fra før ændres som vist nedenfor (XP er en slags variabel over mulige frasytyper: DP, VP, AP mv.).

CP → XP C'
 (↑TOP)=↓ ↑=↓

Annotationerne læses sådan at den første konstituent fylder TOP-trækket i sætningens f-struktur. Den topikaliserede konstituent fylder derudover en eller anden G(rammatisk) F(unktion) i f-strukturen som altså kan være indlejret i en ledsætning (en COMP). I eksempel (7) identificeres *bogen* entydigt som objekt i sin position til højre for negationen. Da verbet kræver et subjekt og et objekt, vil f-strukturen kun være COMPLETE hvis den topikaliserede konstituent er subjektet. I eksempel (8) kan *Peter* kun være subjekt i positionen til venstre for negationen og uden for forfeltet. For at opfylde COMPLETENESS må den topikaliserede konstituent altså være objekt. Kleene-stjernen (COMP*) beskriver en sekvens af træknave gennem f-strukturen idet den tillader et arbitrært antal forekomster af COMP. På den måde vil den tillade topikalisering ud af dybt indlejrede sætninger som i *den bog påstod han at alle vidste at han havde skrevet*. I sidste tilfælde fungerer den topikaliserede kon-

stituent som objekt i ledsætningen *han havde skrevet*, som igen er COMP for *vidste*, som igen er COMP for *påstod*. For at identificere den topikaliserede konstituents syntaktiske funktion som objekt for *skrevet* skal der etableres en relation mellem TOP-trækket og COMP-COMP-OBJ-trækket. Det er præcist hvad annotationen i reglen gør.

LEKSIKON

LFG tilhører familien af leksikalistiske teorier, dvs. syntaksteorier der tillægger leksikon en afgørende rolle i beskrivelsen af også regelbundne sproglige fænomener. Et eksempel herpå er aktiv-passiv-relationen der i derivationelle teorier som Government & Binding har været behandlet som en syntaktisk alternation bundet til forskellige repræsentationsniveauer (en dybde- og en overfladestruktur), men som i LFG behandles som en leksikalisk valens-alternation. Det betyder samtidig at leksikon i disse teorier ikke er en liste af ord med deres uforudsigelige egenskaber, men derimod et struktureret modul med egne muligheder for at udtrykke leksikalske og syntaktiske generaliseringer gennem de leksikalske enheder og ikke mindst gennem relationerne mellem dem. Generaliseringer udtrykkes typisk ved leksikalske regler der relaterer leksikalske indgange til andre leksikalske indgange, og ved organisering af leksikon som et nedarvningshierarki.

En leksikalisk indgang angiver den semantiske værdi for den pågældende leksikalske størrelse og ordets valens i form af en liste af styrede funktioner. Nedenstående leksikalske indgang viser således at verbet *flytte* skal kombineres med et subjekt og objekt.

flytte V (↑PRED)=’FLYTTE<SUBJ OBJ>’

Verbet *flytte* indgår sammen med en lang række andre transitive verber i en systematisk valensalternation med et præpositionalobjekt (i LFG-termer en OBL_θ hvor θ betyder at det oblique led udtrykker en tematisk rolle som fx tema eller agens).

- (10) *Peter flytter stolen*
 (11) *Peter flytter på stolen*

Denne generalisering kan udtrykkes i en leksikalsk regel der relaterer transitiver verber til en alternativ valensramme hvor tema-argumentet udtrykkes som et oblikt præpositionelt objekt (P er en variabel for den aktuelle PRED-værdi).

OBJ-to-OBL $(\uparrow\text{PRED})=\text{'P<SUBJ OBJ>'} \rightarrow (\uparrow\text{PRED})=\text{'P<SUBJ OBJ}_\theta\text{'}$

Eller i forkortet form:

OBJ-to-OBL $(\uparrow\text{OBJ}) \rightarrow (\uparrow\text{OBJ}_\theta)$

Reglen tager som input en valensramme og ændrer den således at der kræves et oblikt led i stedet for et direkte objekt.

Passiv-alternationen for personlig passiv kan tilsvarende beskrives som en leksikalsk operation der gør objektet i den aktive form til subjekt, og subjektet i den aktive form til en optionelt oblikt led, en PP med præpositionen *af*.

Leksikalske regler anvendes i vid udstrækning i implementeringer af LFG, men de har en tvivlsom teoretisk status. For eksempel stipulerer den leksikalske regel for passiv at subjekt i den aktive form (agens) bliver til et optionelt oblikt led, men den siger ikke noget om hvorfor det er den eneste mulighed. Hvorfor bliver agens for eksempel ikke objekt? Som en reaktion på de leksikalske reglers stipulative natur findes der egentlige teorier om koblingen mellem argumentstruktur og funktionel struktur. Et eksempel herpå er LFG's LEXICAL-MAPPING THEORY (LMT). LMT opdeler de grammatiske funktioner SUBJ, OBJ, OBL_θ og IOBJ (indirekte objekt) efter om de er associeret med en bestemt semantisk rolle (+/-r), og efter om de er objektive, dvs. om de kan forekomme som komplement for transitiver verber og præpositioner (+/-o). Det giver følgende klassifikation af funktionerne hvor " >> " indikerer en rangfølge efter hvor markerede funktionerne er:

SUBJ: [-r -o] >> OBJ: [-r+o] >> OBLØ: [+r -o] >> IOBJ: [+r +o]

LMT antager samtidig at tematiske roller er inhærent klassificerede for mulige syntaktiske funktioner. Agens er således inhærent klassificeret for [-o]. Konsekvensen heraf er at et agens-argument enten kan realiseres som subjekt eller som en oblik præpositionalfrase idet begge disse funktioner er [-o] som vist ovenfor. Det er præcist det mønster der findes i aktiv-passiv-alternationen hvor agens i aktiv realiseres som subjekt og i passiv som en præpositionalfrase med af (*katten fanger fisken/fisken fanges af katten*). Klassifikationen som [-o] betyder samtidig at agens-argumentet ikke kan realiseres som objekt da objektet er specificeret som [+o], og dermed beskrives det meget udbredte mønster at agens-argumenter meget sjældent optræder som objekter. Et tema-argument er derimod inhærent klassificeret som [-r], og det betyder at det kan alternere mellem SUBJ og OBJ der begge er [-r]. Konsekvensen heraf er at et tema-argument kan være objekt som vi ser i aktive transitive konstruktioner (*katten fanger fisken*), og at det kan være subjekt som det ses i passiv-konstruktioner (*fisken fanges af katten*). Opdelingen af de syntaktiske funktioner i primitive træk og den syntaktiske klassifikation af de tematiske roller med disse træk motiverer således hvorfor valensalternationer synes at følge bestemte mønstre.

Specielt ved argumenter der alternerer mellem objekter og oblikke led (som i *flytte*-eksemplet i (10) og (11) ovenfor) giver LMT imidlertid problemer. *Flytte* kunne have følgende argumentstruktur med tilhørende inhærent klassifikation:

<i>flytte</i>	<	agens	tema	>
		[-o]	[-r]	

Agens-argumentet vil så mappe til subjekt hvis vi antager at det mest prominente argument med [-o] mapper til subjekt, og tema mapper til objekt, som er den laveste funktion kompatibel med [-r]-specifikationen. Men tema-argumentet kan aldrig mappe til den oblikke funktion da tema er [-r] og den oblikke funktion har [+r], og det er ellers netop det der sker i *Peter flytter på stolen*. Man er derfor nødt til at antage at tema-argumentet får forskellig inhærent klassifikation, fx afhængigt af

verbets telicitet. I den teliske version (den transitive med det direkte objekt) er tema [-r] og kan derfor blive objekt, mens tema i den ateliske version (med præpositionalobjektet, det oblikke led) er [-o] og derfor kan mappe til OBL₀. Med forskellige klassifikationer af tema-argumentet er der imidlertid ikke så meget tilbage af en eventuelt universel in-hærent klassifikation af argumenterne og dermed heller ikke så meget forklaring på valensalternationernes udseende. Bl. a. af den grund lader teorier om koblingen mellem argumentstruktur og syntaks sig nu i højere grad inspirere af optimalitetsteorien med dens tanke om at universelle principper (constraints) er i indbyrdes konkurrence, og at den optimale kandidat er den der overtræder færrest (eller mindre tungtvejende) constraints (Frank, King & Kuhn 1998).

OPTIMALITETSTEORI I TEORI OG PRAKSIS

Op gennem 1990'erne har optimalitetsteorien brudt med den tanke at et constraint er til for at blive overholdt, forstået på den måde at et sprogligt udtryk for at være velformet ikke må overtræde nogen constraints. I stedet repræsenterer optimalitetsteorien en mere fleksibel opfattelse af constraints som rangordnede størrelser, og et sprogligt udtryk kan godt overtræde et constraint uden at blive ugrammatisk hvis det vel at mærke sker i en bedre sags tjeneste, nemlig at undgå overtrædelse af et højere prioriteret constraint. At være grammatisk bliver så at være optimal, det vil sige til at være den realisation af et givet input der overtræder færrest af de dyre constraints. På den måde minder optimalitetstankegangen om den datamatiske lingvistik's regler for hvordan man vælger den bedste analyse, dvs. at man med heuristikker som "subjektet står i reglen først" vælger den mest plausible blandt flere mulige analyser.

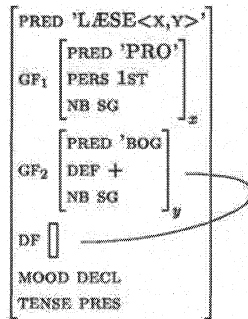
Optimalitetsteori repræsenterer en ny forestilling om den Universelle Grammatik. I optimalitetsteori genereres der til et givet input (en morfo-syntaktisk repræsentation af det man vil sige) en række strukturelle repræsentationer som alle har det til fælles at de bevarer informationen fra input. Men denne mængde af strukturelle repræsentationer (kandidaterne) er ikke bundet til et bestemt sprog, og derfor skal de enkelte sprog, fx dansk, altså også vælge blandt strukturelle kandidater som ellers ikke forekommer i dansk. De findes så ikke i dansk fordi de ikke er

optimale i forhold til den danske udformning af rangfølgen af constraints. Mængden af constraints er fælles for alle sprog, men sprogene prioriterer disse constraints forskelligt (ud over at de har et sprogspecifikt leksikon). Lingvistens opgave er følgende at finde frem til mængden af constraints og at afgøre rangfølgen af constraints for et givet sprog.

Optimalitetsteorien er ikke specielt bundet til syntaks. Oprindeligt stammer den fra fonologien, men anvendes også inden for morfologi, semantik, pragmatik og som nævnt syntaks. Dertil kommer at optimalitetsteorien ikke er bundet til en bestemt syntaktisk teori. Den har tværtimod været anvendt i forbindelse med flere forskellige teorier. Forskellene er her typisk hvordan input antages at se ud, og hvordan de strukturelle repræsentationer i kandidatmængden ser ud. For eksempel vil en optimalitetsteoretisk analyse inden for en derivational teori som Government & Binding antage at kandidaterne er X-bar-orienterede træstrukturer med spor (*traces*) der indikerer flytninger. I optimalitetsteoretisk LFG (fx Kuhn 2003) vil kandidatrepræsentationerne være par af c- og f-strukturer hvor c-strukturerne, som nævnt tidligere, ikke nødvendigvis er i overensstemmelse med X-bar-teorien. Den teoretiske ramme har dog undertiden også konsekvenser for hvilke constraints der antages. For eksempel gælder de ovenfor nævnte velformethedsprincipper COMPLETENESS og COHERENCE som meget højt rangerende constraints i optimalitets-baseret LFG (måske endda uovertrædelige), om end andre teorier typisk har lignende principper for at sikre valensopfyldelse.

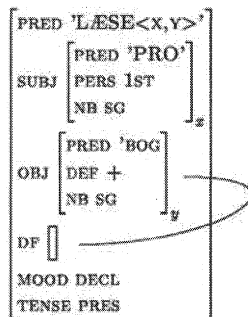
I OT-baseret LFG kan input være en underspecificeret trækstruktur som vist i figur 9 (Bresnan 2000), som repræsenterer et givent morfo-syntaktisk indhold uafhængigt af hvordan det realiseres udtryksmæssigt.

FIGUR 9. INPUT I OT-BASERET LFG

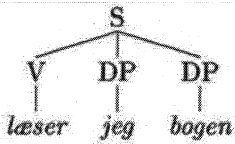


Input angiver en basal prædikat-argument-struktur, dvs. den angiver relationerne mellem prædikatet (*læse*) og de involverede argumenter (*jeg* og *bog*) i form af underspecificerede, men distinkte grammatiske funktioner (GF₁ og GF₂). Derudover angiver input tempus, modalitet, bestemthed og at *bog* er udhævet i diskursen (den indtager en diskursfunktion DF). Ud fra dette input genereres en mængde af c- og f-struktur-par. For overskuelighedens skyld vises nedenfor kun 4 mulige strukturelle kandidater som alle er relateret til f-strukturen i figur 10 (jf. Bresnan 2000). F-strukturen i figur 10 er i modsætning til input i figur 9 en parallel struktur til en c-struktur med tilhørende leksikalske elementer og er derfor mere specifik end input. Her er inputtets underspecificerede grammatiske funktioner således instantieret til SUBJ(ekt) og OBJ(ekt).

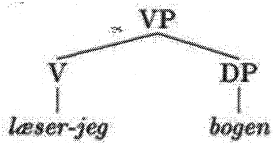
FIGUR 10. F-STRUKTUR FOR KANDIDATMÆNGDEN



Denne kandidat overtræder flere principper for dansk syntaks. For eksempel står det finitte verbum ikke på 2.-pladsen hvilket kan formuleres

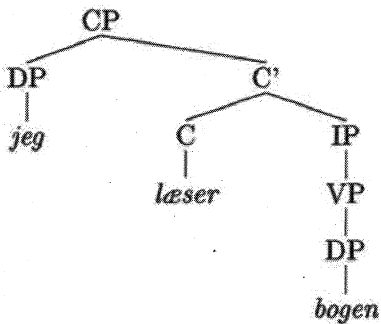


overveje om denne kandidat overhovedet overholder kravet om at bevare informationen fra input).

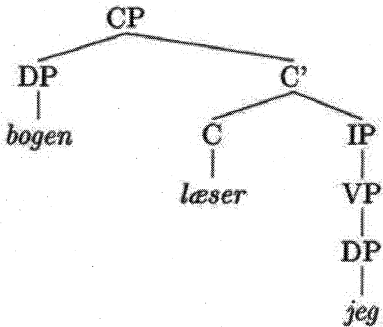


Denne kandidat overholder heller ikke kravet om V2 og om at diskursfunktioner står til venstre (TOP-L), og samtidig kategoriserer den subjektet som en verbal bøjningsendelse (som fx i italiensk).

Dermed overtræder den et krav om at der skal være et manifesteret subjekt (SUBJ), som nok er et højt prioriteret krav i dansk.



Denne kandidat overholder et krav om at der skal være et manifesteret subjekt (SUBJ), og at verbet står på 2.-pladsen (V2), men den overtræder kravet om at en diskursfunktion spidsstilles på dansk (TOP-L). Til gengæld overholder den et krav om at et subjekt typisk står til venstre på dansk (SUBJ-L).



Denne kandidat overholder ligeledes kravet om at der skal være et subjekt (SUBJ), og at verbet står på 2.-pladsen (V2). Den overholder tillige kravet om at en diskursfunktion spidsstilles (TOP-L), men den overtræder kravet om at subjekt typisk står til venstre på dansk (SUBJ-L). Da det imidlertid lader til at være sådan i dansk at subjektet

står først med mindre andet ønskes udhævet i diskursen (i dette tilfælde objektet), kunne dette være den optimale kandidat, og TOP-L vil således være vigtigere end SUBJ-L.

Evalueringen af disse kandidater er givet (i stærkt forenklet form) i skemaet i figur 11 hvor udråbstegnet angiver at en overtrædelse er "fatal" og altså fører til at en kandidat ikke kan komme i betragtning som optimal.

FIGUR 11. EVALUERING AF DE FIRE KANDIDATER

- "☞" UDPEGER DEN OPTIMALE KANDIDAT

		V2	SUBJ	TOP-L	SUBJ-L
	[S [V læser] [DP jeg] [DP bogen]]	*!		*	
	[VP [V læser-jeg] [DP bogen]]	*!	*	*	
	[CP [DP jeg] [C' [C læser] [IP [VP [DP bogen]]]]]			*!	
☞	[CP [DP bogen] [C' [C læser] [IP [DP jeg]]]]				*

Optimalitetsteori er således kendetegnet ved antagelsen om at constraints kan være indbyrdes modstridende, og at den grammatiske kandidat er "det bedste kompromis". Samtidig er optimalitetsteori et eksempel på at egentligt regelbaseret grammatik opgives til fordel for strukturelle repræsentationer.

OPTIMALE KANDIDATER OG DATAMATISKE GRAMMATIKKER

Tanken om at nogle analyser er mere optimale end andre, kendes i høj grad fra udvikling af datamatiske grammatikker. Her drejer det sig dog ikke nødvendigvis om at finde den grammatiske kandidat, men i højere grad om at vælge den mest plausible blandt flere mulige analyser. Jo dybere analyse en grammatik foretager, jo flere dobbeltydigheder vil den i reglen finde.

I datamatiske LFG-grammatikker kan man definere en særlig optimalitetsstruktur der hjælper med at identificere den mest plausible analyse. Denne o-struktur er en mængde af optimalitetsmarkeringer som tilskrives gennem grammatikreglerne eller leksikonposterne. Hvis der er flere analyser til et givet input, evalueres de enkelte analysers o-struktur i forhold til en defineret rangfølge af markeringer, og den optimale analyse vælges. Ud over en egentlig rangfølge af optimalitetsmarkeringer kan man også skelne mellem positive og negative markeringer. Positive markeringer betyder at en analyse med denne markering foretrækkes, mens en negativ markering omvendt betyder at en analyse kun bruges i mangel af en bedre. Positive og negative markeringer i kombination med rangfølge muliggør en meget subtil sorteringsmekanisme.

O-strukturen kan udnyttes til at identificere ledfunktioner på dansk. Som bekendt identificeres subjekt og objekt på dansk ved en kombination af ledstilling og selektionsrestriktioner. I (12) nedenfor kan *firmaet* kun være objekt da det står til højre for nægtelsen. Eksempel (13) vil derimod typisk producere to analyseresultater da man ud fra position ikke kan afgøre om *direktøren* eller *firmaet* er hhv. subjekt eller objekt. Da verbet *lukke* imidlertid i reglen tager et humant subjekt, vil man dog nok gøre *direktøren* til subjekt. I eksempel (14) kan man slet ikke støtte sig til selektionsrestriktionerne da ingen af substantivernes referenter er humane. Her vil man nok automatisk gå ud fra at det er subjektet der står først. Og det eksempel viser samtidig at selektionsrestriktioner næppe kan formuleres som absolutte krav da de netop meget ofte overtrædes.

(12) *direktøren lukkede ikke firmaet*

(13) *direktøren lukkede firmaet*

(14) *firmaet lukkede filialen*

Interaktionerne mellem position, selektionsrestriktioner og default-fortolkninger kan beskrives ved en kombination af positive og negative optimalitetsmarkeringer. Overtrædelse af selektionsrestriktioner kan angives ved markeringen SELRESVIOL (mnemoteknisk for “selectional restriction violation”) mens default-placeringen af subjektet som den første konstituent kan angives med markeringen SUBJFIRST.

SUBJFIRST kan gøres til et positivt træk (med præfigeret + som vist nedenfor) mens overtrædelse af selektionsrestriktioner gøres til et negativt træk.

OPTIMALITYORDER +SubjFirst SelResViol

Markeringerne for selektionsrestriktioner angives i den leksikalske indgang for verbet *lukke*:

lukke V (↑PRED)=’LUKKE<SUBJ OBJ>’
 {{(↑SUBJ HUM)= c+ |
 SELRESVIOL ∈ o(V)}

De to sidste linier indeholder en noget forenklet syntaks i forhold til standard-LFG, men de læses sådan at enten skal subjektet i f-strukturen indeholde trækket [HUM +], eller også tilskrives sætningen strafmarkeringen SELRESVIOL, der fortæller at sætningen har overtrådt selektionsrestriktionerne for det pågældende verbum ved ikke at have et human subjekt. Denne strafmarkering samles i en særlig o-struktur (optimalitetstruktur) der i sidste ende bruges til at vælge den bedste af flere mulige analyser. Præferencen for et initialt subjekt angives i selve sætningsreglen, sådan at hvis TOP-attributtet i f-strukturen er identisk med SUBJ-attributtet, tilskrives markeringen SUBJFIRST til sætningens o-struktur.

Sætningen *direktøren lukkede firmaet* får de to analyser vist i figur 12.

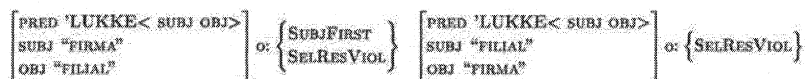
FIGUR 12. ANALYSER FOR ‘DIREKTØREN LUKKEDE FIRMAET’

$$\left[\begin{array}{l} \text{PRED 'LUKKE< SUBJ OBJ>} \\ \text{SUBJ "DIREKTØR"} \\ \text{OBJ "FIRMA"} \end{array} \right] \text{ o: } \{ \text{SUBJFIRST} \} \quad \left[\begin{array}{l} \text{PRED 'LUKKE< SUBJ OBJ>} \\ \text{SUBJ "FIRMA"} \\ \text{OBJ "DIREKTØR"} \end{array} \right] \text{ o: } \{ \text{SELRESVIOL} \}$$

Den første analyse, hvor *direktøren* er subjekt, har en positiv markering mens den anden analyse, der gør *direktøren* til et spidsstillet objekt, derimod har en negativ markering. Den første analyse vil altså blive foretrukket.

Sætningen *firmaet lukkede filialen* får de to analyser vist i figur 13.

FIGUR 13. ANALYSER FOR 'FIRMAET LUKKEDE FILIALEN'



I begge tilfælde er selektionsrestriktionerne overtrådt, og begge o-strukturer indeholder altså en negativ markering. Til gengæld har den første analyse med firmaet som subjekt en positiv markering, nemlig at subjektet står først. Derfor vælges den første analyse.

O-struktur og optimalitetsteori er altså ikke det samme, men eksemplet viser at tanken om optimale kandidater over for mindre optimale kandidater har vist sig at være meget frugtbar inden for både teori og praksis.

LFG OG DATA-DREVET PARSING (DOP)

Når man arbejder med datamatiske grammatikker, opdager man at sproget kan være mere dobbeltydigt end man tror. Computere finder analyser som nok reelt er mulige, men som ofte er meget lidt plausible ud fra semantiske overvejelser eller almindelig omverdensviden. Én måde at gøre computerne mere "performans"-orienterede på er som nævnt at anvende en form for præferencemarkeringer, en anden mulighed er at udvide grammatikken med statistiske oplysninger, som dermed indarbejder semantiske og videnskæssige hensyn. For eksempel kan man på basis af korpora undersøge hvor hyppigt bestemte grammatikregler anvendes, og når grammatikken har adgang til disse oplysninger, kan man ud fra de enkelte reglers sandsynlighed udregne sandsynligheden for en given analyse og altså vælge den mest sandsynlige. På den anden side er det ikke sikkert at grammatikkens enheder (fx de udsnit af træstrukturer som en kontekstfri regel kan beskrive, nemlig en rodkategori og de kategorier rodkategorien dominerer, jf. den kommenterede regel fra eksemplet i figur 7) er de samme enheder som an-

vendes ved menneskelig sprogprocessering. Tværtimod kan man forestille sig at den menneskelige sprogprocessering opererer med enheder af variabel størrelse, nogle gange repræsentationer for hele sætninger, andre gange repræsentationer for deltræer.

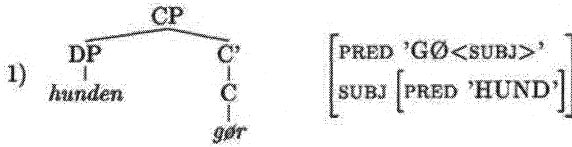
Data-orienteret Parsing (DOP) (Bod 1998) bygger på en forestilling om at sproglig viden ikke er baseret på regler, men derimod på syntaktiske repræsentationer af tidligere hørte ytringer. Disse repræsentationer kan være hele konstituentstrukturer, fx for hele sætninger, men også fragmenter af konstituentstrukturer af variabel størrelse. Nye ytringer analyseres ved at eksisterende fragmenter af syntaktiske strukturer kombineres til mulige analyser af det nye input, og sandsynligheden for en given analyse beregnes i modellen ud fra de repræsentationer sprogbrugeren allerede har adgang til. Jo mere en analyse af et nyt input ligner allerede eksisterende repræsentationer, jo sandsynligere er den pågældende analyse. Repræsentationer af nyt input er således en produktiv kombination af strukturer fra et eksisterende korpus af syntaktiske repræsentationer.

Udgangspunktet for en DOP-model er et korpus af syntaktiske repræsentationer af ytringer. Selve modellen består af

- En formel specifikation af den lingvistiske repræsentation
- Operationer til opdeling af repræsentationer i strukturfragmenter
- Operationer til komposition af strukturfragmenter
- En sandsynlighedsmodel

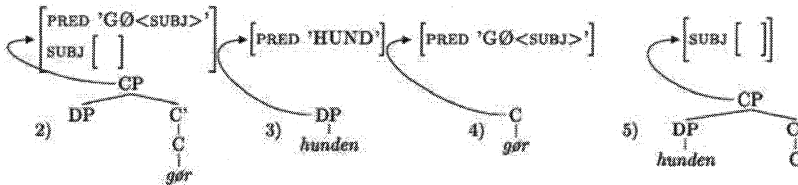
Der findes flere forskellige tilgange til data-orienteret parsing i LFG, men i én version af LFG-baseret DOP består den lingvistiske repræsentation af ytringer af c-strukturer med tilhørende f-strukturer (Bod & Kaplan udkommer). Hvis korpus indeholder ytringen *bunden gør*, kan den være associeret med c- og f-struktur-parret i figur 14.

FIGUR 14. C- OG F-STRUKTUR FOR 'HUNDEN GØR'



Modellens definitioner af strukturfragmenter beskriver hvordan denne repræsentation kan opdeles i yderligere strukturfragmenter. På basis af repræsentationen i 1) i figur 14 dannes (blandt andre) strukturfragmenterne i figur 15, ligeledes som del af korpus.

FIGUR 15. EKSEMPEL PÅ STRUKTURFRAGMENTER I KORPUS



Fragmenteringen sikrer at de enkelte delrepræsentationer igen kan indgå i andre strukturbeskrivelser for andre ytringer. Kombinationen foregår fx ved at et strukturfragment kan kombineres med et andet strukturfragment hvis den venstre non-terminale knude i et træ matcher kategorien for topknuden i det andet deltræ, og hvis de to f-strukturer kan unificeres. Således vil strukturfragmenterne 2) og 3) i figur 15 kunne kombineres til repræsentationen i 1) i figur 14 da den venstre nonterminale knude, DP, i 2) matcher topknuden DP i 3), og da f-strukturerne kan unificeres således at [PRED 'HUND'] indgår som værdi for SUBJ(ekt)-attributtet i f-strukturen i 2).

Ved parsing/analyse på basis af den data-orienterede metode undersøges hvordan en analyse for en given ytring kan derivatives ud fra de allerede eksisterende strukturfragmenter i korpus. Man finder ud af hvilke strukturfragmenter der kan kombineres således at resultatet bliver den sætning der skal analyseres. Når man på den måde har fundet en mulig derivation af ytringen ud fra de eksisterende strukturfragmenter,

beregner man en sandsynlighed for denne derivation. Man finder sandsynlighederne for de enkelte strukturfragmenter ved at undersøge antallet af strukturfragmenternes forekomster i korpus i relation til antallet af forekomster af strukturfragmenter med samme rodkategori. Sandsynligheden for et strukturfragment [DP *hunden*] som i 3) i figur 15 er således antallet af forekomster af netop dette strukturfragment i relation til alle forekomster af strukturfragmenter med DP som rodkategori. Hvis [DP *hunden*] forekommer 10 gange ud af i alt 100 strukturfragmenter med DP som rodkategori, er sandsynligheden for dette fragment 0.1. Når alle strukturfragmenterne og deres sandsynligheder er fundet, beregnes derivationens samlede sandsynlighed som produktet af alle strukturfragmenternes sandsynligheder. Sandsynligheden for en analyse er med andre ord større jo færre strukturfragmenter der indgår. En ytring kan fx have to forskellige analyser i korpus, svarende til to forskellige derivationer. I den første analyse indgår to strukturfragmenter med sandsynlighederne 0.1 og 0.2, hvorved den samlede sandsynlighed for derivationen bliver 0.02. I den anden analyse indgår tre strukturfragmenter med sandsynlighederne 0.3, 0.1 og 0.2. Den samlede sandsynlighed bliver således 0.006. Den første analyse med to strukturfragmenter foretrækkes fordi den har den højeste sandsynlighed. Dette er igen udtryk for at sandsynligheden for en analyse er større jo mere analysen ligner tidligere strukturer som allerede ligger i korpus.

Ofte kan man nå frem til den samme analyse ad flere forskellige derivationer. For eksempel kan man nå frem til en analyse af *hunden gør* med strukturen i 1) i figur 14 ved enten at finde hele repræsentationen (da sætningen allerede forekommer i korpus), eller ved at kombinere strukturfragmenter som fx fragment 2) og fragment 3) i figur 15. I disse tilfælde vil sandsynligheden for analysen være summen af sandsynligheden for de enkelte derivationer der fører til denne analyse. Igen får man altså den effekt at jo større dele af input-ytringen der matcher noget der allerede kendes i korpus, jo mere sandsynlig er den pågældende analyse. I DOP foretrækkes med andre ord en analyse som man kan nå til med flere forskellige derivationer ud fra strukturfragmenter i korpus idet en sådan analyse vil have en højere samlet sandsynlighed. Mange forskellige derivationer forekommer netop når det er muligt at genbruge store fragmenter fordi store fragmenter samtidig kan dannes ud fra deres en-

keltdele. Netop derved opnår DOP den effekt at den allerede eksisterende sproglige erfaring udnyttes ved mødet med nyt input. Jo mere en ny ytring ligner de lagrede repræsentationer, jo større er sandsynligheden for at den analyseres i overensstemmelse med det allerede kendte.

Det afgørende ved DOP-modellen er dels at den søger at ophæve skellet mellem kompetens og performans idet den opererer direkte på lagrede data, dels at den opererer med repræsentationer og ikke regler. I stedet for at lede efter en Universel Grammatik leder man således efter et Universelt Repræsentationsformat, og her er LFG med sit system af korrespondens mellem forskellige sproglige dimensioner altså et bud på et sådant format.

LFG SOM MODEL FOR SPROGLIG VIDEN

LFG har i sin 25-årige historie været en overordentligt stabil teori, og samtidig har den som teori vist sig yderst anvendelig inden for både anvendelsesorienteret og teoretisk lingvistik: LFG-grammatikker har for eksempel været brugt i datamatiske applikationer, undersøgelser af sprogtilegnelse (Pienemann 1998), optimalitetsteori og data-orienteret parsing – ud over at LFG har været den teoretiske ramme for et væld af enkeltanalyser. Centralt står LFG's parallelle arkitektur med forskellige dimensioner af sproglig information forbundet ved afbildningsfunktioner. En lang række væsentlige emner har end ikke været nævnt: opbygningen af semantisk repræsentation, inkorporering af morfologi, samspil mellem syntaks og informationsstruktur, for slet ikke at tale om LFG-formalismens matematiske egenskaber. For lingvisten giver LFG ikke blot (som andre formelle modeller) "the advantages of amplifying, extending, and clarifying the analytic imagination" (Bresnan 2001: 43), men også en konsistent ramme for beskrivelsen af hvordan sproget virker som system, og som model et bud på hvordan sprog forarbejdes.

Bjarne Ørsnes
Copenhagen Business School
E-mail: boe.id@cbs.dk

NOTER

- 1 Jeg vil gerne takke Jørg Asmussen, Jürgen Wedekind og NyS-redaktionen for mange kritiske og skarpsindige kommentarer til tidligere versioner af denne artikel. Alle fejl og utilstrækkeligheder er naturligvis alene mit ansvar.
- 2 Notationen *PRED 'HAVE <OBJ> SUBJ'* viser netop at der er tale om raising. Hjælpeverbet *have* kombineres med et objekt og derudover et ikke-tematisk subjekt (uden for de spidse parenteser), det vil sige et subjekt som hører hjemme i argument-strukturen for en anden prædikator (i dette tilfælde hovedverbet).

LITTERATUR

- Bod, R. & R. Kaplan (udkommer): "A Data-Oriented Parsing Model for Lexical-Functional Grammar". Tilgængelig på <http://staff.science.uva.nl/~rens/>.
- Bod, R. (1998): *Beyond Grammar: An experienced-based Theory of Language*. Stanford: CSLI Publications.
- Bresnan, J. (2001): *Lexical-Functional Syntax*. Blackwell Publishers.
- Bresnan, J. (2000): "Optimal Syntax". J. Dekkers, F. van der Leeuw & J. van de Weijer (eds.): *Optimality Theory - Phonology, Syntax and Acquisition*. Oxford: Oxford University Press. 334-385.
- Bresnan, J. & A. Zaenen (1990): "Deep unaccusativity in LFG". K. Dziwirek, P. Farrell and E. M. Bikandi (eds.): *Grammatical relations: a cross-theoretical perspective*. Stanford: CSLI Publications. 45-57.
- Butt, M, T.H. King, M. Niño & F. Segond (1999): *A Grammar Writer's Cookbook*. Stanford: CSLI Publications.
- Butt, M. & T.H. King. (2003): "Grammar Writing, Testing, and Evaluation". A. Farghaly (ed.): *Handbook for Language Engineers*. CSLI Publications. 129-178.
- Dalrymple, M. (2001): *Lexical Functional Grammar*. 'Syntax and Semantics', vol. 34. San Diego: Academic Press.
- Frank, A., T.H. King, J. Kuhn & J. Maxwell (1998): "Optimality Theory style constraint ranking in large-scale LFG grammars". *Proceedings of the LFG '98 Conference*. University of Queensland: Brisbane.
- Heltoft, L. & E. Hansen (2003): *Grammatik. Syntaks*. 'Skrifter fra Dansk og Public Relations'. Roskilde Universitetscenter.
- Korpus2000*. København: Det Danske Sprog- og Litteraturselskab.
- Kuhn, J. (2003). *Optimality-Theoretic Syntax - A Declarative Approach*. Stanford CA: CSLI Publications.

- Pienemann, M. (1998): *Language processing and second language development – Processability theory*. Amsterdam, Netherlands: John Benjamins Publishing Company.
- Vikner, S. (1995): *Verb Movement and Expletive Subjects in the Germanic Languages*. New York: Oxford University Press.
- Vikner, S. (1999): “Ledstillingen i dansk og Government & Binding”. P.A. Jensen & P. Skadhauge (red.): *Sætningskemaet i generativ grammatik*. Kolding: Institut for Erhvervsproglig Informatik og Kommunikation. SDU. 83-110.
- Wedekind, J. & Bjarne Ørsnes (2004): “An Account of the Danish Verbal Complex and its Topicalization”. *Acta Linguistica Hafniensia*, vol. 36. 35-64.
- Ørsnes, B. (2004): “Automatisk opbygning af et LFG-baseret datamatisk leksikon for dansk”. Henrik Holmboe (red.): *Nordisk Sprogteknologi. Årbog for Nordisk Sprogteknologisk forskningsprogram 2000-2004*. København: Museum Tusulanums Forlag. 211-236.