

Morfemformernes distribution

Præmisses for beskrivelse af morfemformers distribution på baggrund af en rent formalistisk analyse af udtrykssubstansen

1.0 Indledning

I især den ældre sprogvidenskab var grammatikken meget nært knyttet sammen med studiet af sprogformen og også stærkt præget af tidlige grammatikker for latin, men via latin også i høj grad præget af oldgræsk. Ligesom de latinske grammatikker var de tidligere danske grammatikker hovedsageligt koncentreret omkring ordformen, dvs. morfologi i en snæver forstand. I løbet af det 20. århundrede er fokus dog blevet flyttet fra morfologiske studier til syntaktiske studier hvilket her i landet specielt har været synligt med påvirkningerne fra Diderichsens *Elementær Dansk Grammatik*.

Skiftet fra fokus på ordet til fokus på sætningen har på den mere teoretiske side været ledsaget, og måske effektueret, af et skift fra fokus på sprogformen til et fokus på mere semantiske og pragmatiske sprogteorier. De rent formalistiske sprogteorier mangler en vis beskrivelseskraft på steder hvor de funktionelle grammatikker i højere grad har deres styrker (jf. afsnit 3.2).

Ikke desto mindre vil det være en formalistisk metode der i det følgende bliver analysegrundlaget, men deraf følger naturligvis ikke per se at den bagvedliggende sprogforståelse også er formalistisk (i en chomskyansk forstand). Det vil således være min forhåbning at den formalistiske metode kan være med til dels at pege på nogle styrker og dels nogle svagheder ved en sådan tilgang. Metoden er således valgt ikke som følge af en bestemt sprogforståelse, men som følge af en grundlæggende nysgerrighed efter hvor stor en del af et sprog der kan udledes ved en rent formalistisk analyse. En nysgerrighed jeg formentlig aldrig får tilfredsstillet.

Af de Saussures velkendte beskrivelse af sprogtegnet følger at relationen mellem det betegnende og det betegnede er arbitrær (de Saussure, 1916: 67). Dette er med omvej af pragerfonologerne, her især Trubetzkoy, baggrunden for Hjelmsslevs fonemdefinition hvor han bestemmer en metode til inventariering af invarianter således at "[e]n invariantforskel er til stede i udtryksplanet, naar der dær bestaar en korrelation (f. ex. korrelationen mellem *a* og *æ* i *lak* – *læk*), som modsvares af en korrelation i indholdsplanet (korrelationen mellem indholdsstørrelserne 'lak' og 'læk') paa en saadan maade at der bestaar en *relation* [i.e. en både-og-funktion (konjunktion)] mellem udtrykskorrelationen og indholdskorrelationen" (1943: 59).

At forholdet mellem form og indhold er arbitrært, medfører at det ikke kan lade sig gøre at slutte fra udtryk til indhold alene på baggrund af udtrykket og omvendt, men Hjelmsslevs definition af invarians (eller mere rigtigt af invariantforskel) medfører desuden at for at finde en invariant i udtryksplanet (en 'udtryksform'), må man nødvendigvis inddrage indholdsplanet. Den distinktive funktion sættes således foran alt andet.

Væsentligt er hertil at Hjelmsslevs definition i *Omkring Sprogteoriens Grundlæggelse* (herefter OSG) bygger på overvejelse over "udtrykkets invarianter af højeste grad", dvs. fonemer, og disses variant-eksemplarer, dvs. de fonetiske enkeltlyd (s. 56). Hvis man derimod antager at fonemanalysen af et sprog allerede foreligger, i hvilken udstrækning vil det så ud fra en potentiel infinit mængde fonemsekvenser være muligt at konstruere et udtrykssystem uden på noget tidspunkt at inddrage indholdsplanet? Dette er det centrale spørgsmål for nærværende undersøgelse.

Hvis det antages at Hjelmsslev har ret i sin betragtning, med henvisning til især Trubetzkoy's fonologiske arbejde som underbygning, og som følge heraf at fonemer kun kan inventarieres ved en hensyntagen til både udtryks- og indholdsplanet, og hvis det desuden antages (indtil videre uargumenteret, men som følge af Hjelmsslevs definition) at det vil være umuligt at konstruere en procedure der, givet fonemanalysen, vil være i stand til at opstille en fuldbyrdet grammatik for et sprog på baggrund af udtryksstansen alene, hvordan forholder det sig så med den grammatiske analyse mellem fonemniveau og sætningsniveau? – Dvs. det niveau man i bred, hjelmsslevsk forstand kunne kalde morfemniveauet (jf. Hjelmsslev, 1972: 37).

Da det er tilfældet at forholdet mellem udtryk og indhold er arbitrært, vil ordene *vand* og *mand* være lige så forskellige i en analyse af skriftsproget som ordene *konjak* og *cognac* eller *eksaminer* og *eksamener* – hvis igen vi ikke havde nogen viden om indholdsplanet. Forholdet mellem udtryk og indhold er således ganske vist arbitrært, men desuden helt usystematisk for de leksikalske morfemer. Også for de grammatiske morfemers vedkommende er den almenyldige regel, dvs. det arbitrære forhold mellem udtryk og indhold, stadig gældende, men modsat de leksikalske morfemers forekomst, er de grammatiske morfemers forekomst i langt højere grad systematisk – hvilket sandsynligvis har relation til det faktum at grammatiske morfemer ikke som leksikalske morfemer er potentielt infinitte i mængde, men derimod finitte i mængde – og i forholdet til de leksikalske morfemer endog meget stærkt begrænsede i mængde¹. Denne systematik giver en forhåbning om at man på det morfologiske niveau med større sandsynlighed vil kunne forudsige en del af udtryksformen på baggrund af udtrykssubstansen alene.

* * *

Det centrale spørgsmål er hvor stor en del af et sprogs grammatik der kan udledes rent formalistisk, dvs. ud fra en procedure der er uafhængig af andet end det foreliggende materiale. Den bærende tese, som vil blive forsøgt falsificeret, er:

- at det er en umulighed at inventariere et sprogs fonemer,
- at det er en umulighed at forudsige et sprogs topologi, og
- at det givet en fonologisk analyse er en mulighed at inventariere en række ledtypehelheder (helhed med en syntaktisk funktion i sætningen)

ud fra en rent formalistisk procedure. Den løbende falsificering vil dels problematisere og uddybe disse teser, og dels vil den munde ud i nogle mere konkrete overvejelser over muligheden for at inventariere nogle ledtypehelheder givet en række præmisser.

Indtil videre er en del begreber blevet anvendt med en vis indforståethed der ikke er helt berettiget. De er ikke alle uproblematiske hvorfor det vil være nødvendigt at indlede med en begrebsafklaring og en yderligere afgrænsning af problemet. Begrebsafklaringen såvel som problemafgrænsningen vil støtte sig op af nogle af de

koncepter Hjelmslev og Uldall skitserer for det grammatiske system, især på baggrund af de Saussure. Disse teoretikere vil også udgøre grundlaget i den følgende undersøgelse. Valget af de lidt ældre teoretikere som bærende i arbejdet skyldes flere forhold. Dels kan man argumentere for at de er fra en videnskabshistorisk periode hvor lingvistikken konsolideredes som selvstændig videnskab ud fra strukturalistisk inspirerede idéer om sprog, dels er de alle inspireret af den logiske positivismes videnskabskriterier der – uanset hvad man måtte mene om denne – med dens objektive verificeringskrav bærer vigtige grundelementer for en undersøgelse af relationen mellem substans og form, hvorimod de fleste moderne grammatiske teorier indeholder væsentlige kognitive aspekter af sprog (jf. afsnit 2.2). Det videnskabshistoriske islæt tillades således i det følgende at smitte en anelse af på undersøgelsen, men det er forhåbningen at dette ikke bliver på bekostning af en trods alt ikke uaktuel problemstilling.

2. 0 Problemer og begreber

Udgangspunktet for det omdrejningspunkt der indtil videre er skitseret i indledningen ovenfor, var en passage fra OSG hvor Hjelmlev skriver (s. 10):

”Paa forhaand skulde den sætning synes at maatte have generel gyldighed, at der til ethvert *forløb* svarer et *system*, gennem hvilket forløbet lader sig analysere og beskrive ved hjælp af et begrænset antal forudsætninger. Ethvert forløb maa antages at kunne anskues som sammensat af elementer af et begrænset antal, der stadig kommer igen i forskellige kombinationer. Disse elementer maatte da paa grundlag af forløbsanalysen kunne ordnes i klasser med ensartede kombinationsmuligheder, og der maatte videre paa grundlag heraf kunne opstilles en almindelig og udtømmende kalkyle over de mulige kombinationer. En saaledes anlagt historie maatte fra det blot primitivt skildrende stadium stige til en systematisk, exakt og generaliserende videnskab, i hvis teori alle begivenheder (mulige forbindelser af elementer) er forudset og betingelserne for deres realisation fastlagt.”

Det skal indledningsvist nævnes at Hjelmlev opererer med en mængde begreber han definerer formalt, herunder begreberne *analyse*, *forløb*, *system*, *element*, *klasse*. Begreberne defineres dels i OSG, dels i *Resumé of a Theory of Language* og dels af Uldall i *Outline of Glossematics*. En del af Hjelmlevs formale system behandles senere i undersøgelsen, men dog udelukkende hvor det vurderes at være gavnligt for arbejdet.

Hjelmlev indfører her to præmisser for den videre teori, nemlig (1) at ethvert forløb kan anskues som sammensat af elementer af et begrænset antal, og (2) at disse elementer kommer stadigt igen i forskellige kombinationer. For Hjelmlevs første præmis er det karakteristisk at han anvender formuleringen *kan anskues som*. For psykologen er ’forløbet’ noget ganske andet end for lingvisten, og denne forskel pointeres. Noget af det væsentligste for præmis 1 kan siges at være det finite – elementer af begrænset antal, mens noget af det væsentligste for præmis 2 kan siges at være infinittet – ”stadig ... i forskellige kombinationer”. Via en række kombinationer bør man altså vha. de finite elementer kunne anskue stadig forskellige kombinationer. Videre vil en forløbsanalyse kunne opdele de finite elementer i ”klasser med ensartede kombinationsmuligheder”, og alle de mulige kombinationer bør if. Hjelmlev kunne udledes af en opstillet kalkule.

Dette kan hævdes at være grundtanken i Hjelmslevs sprogteori (som den præsenteres i OSG) da han angiver at ”Sprogteoriens hensigt er at gennemprøve denne tesis om et system bag forløbet, en konstans bag fluktuationen” (ibid. s. 11). Et nærliggende spørgsmål ud fra en mere moderne vinkel ville være: ”I denne teori, hvad er da sprog?”. I stedet for selv at spørge er det fristende at lade en moderne teori tage ordet:

”In the terms just outlined, we can consider some of the classical problems of the study of language.

- a. What does Jones know when he has a particular language?
- b. How did Jones acquire this knowledge?
- c. How does Jones put this knowledge to use?
- d. How did these properties of the mind/brain evolve in the species?
- e. How are these properties realized in mechanisms of the brain?” (Chomsky, 1995: 17)

Selvom det måske ikke er vanskeligt ud af citatet at se at Chomsky er generativist, så er det på mange måder de samme spørgsmål der rejses af funktionalisterne:

”How does the natural language user (NLU) work? How do speakers and addressees succeed in communicating with each other through the use of linguistic expressions? How is it possible for them to make themselves understood, to influence each other’s stock of information ..., and ultimately each other’s practical behavior, by linguistic means?” (Dik, 1989: 1)

- Og nogle af de spørgsmål Dik rejser for at forklare sin teori, er:

- “What is the main function of a language?”
- “What is the psychological correlate of a language?”
- “What is the relation between the system of a language and its use?”
- “How do children acquire a natural language?”
- “How can language universals be explained?” (ibid. s. 5 ff.)

Ordene *acquire* og *use* er anvendt begge steder. Derudover er der visse diskursive forskelle på hvordan spørgsmålene formuleres, men spørgsmålene er meget ens, eksempelvis kan man sige at Chomskys spørgsmål *How are these properties realized in mechanisms of the brain?* modsvares af Diks *What is the psychological correlate of a language?*. Der er store forskelle på Chomskys og Diks grammatik, og det er næppe tilfældigt at Dik vælger ordet *psychological* frem for *mechanisms in the brain* eller *mind/brain* i Chomskys spørgsmål (d), men på mange måder er udgangspunktet for de

to nutidige teoretikere dog stadig det samme. Chomsky formulerer denne fælleshed som *explanatory adequacy* (Chomsky, 1965: 25f.):

”To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of *explanatory adequacy*. That is, to this extent, it offers an explanation for the intuition of the native speaker on the basis of an empirical hypothesis concerning the innate predisposition of the child to develop a certain kind of theory to deal with the evidence presented to him”.

Chomsky er hurtig til at indrømme at dette mål kun meget vanskeligt vil kunne nås. Men det er dog stadig et mål. Diks mål er en smule mere beskedent, men minder dog meget om Chomskys: ”how could we build a model of the natural language user (M.NLU) in such a way that M.NLU can do the same kinds of things that real NLUs can?”. Den store forskel mellem Chomskys og Diks tilgang kan på dette punkt siges at være at Chomsky efterspørger ”the innate predisposition of the child”, mens Dik omhyggeligt nøjes med at forlange en *model* for sprogbrugeren. Den store lighed er her at de begge søger en teori der afspejler sprogbrugernes mentale/psykiske kompetencer.

2.1 HJELMSLEVS TEORIFORSTÅELSE

Man kan lidt polemisk sige at en ’moderne’ sprogteori af denne type med interesse for sprogbrugeren på ingen måde er vedrørende for en strukturalistisk tilgang som Hjelmsslevs. Den de Saussure-inspirerede strukturalisme deler sine rødder med de forudgående franske sociologer, eksempelvis Durkheim, for hvem enkeltindividet som aktør kun havde en meget ringe status i undersøgelsen, om nogen, frem for de samfundsmæssige strukturerⁱⁱ. Med Chomsky og Dik som sammenligningsgrundlag bliver det da også med eksemplificering i Hjelmsslevs egen tekst noget tydeligere at hans teoriforståelse er en anden:

”1. Teorien fremtræder i sig selv som uafhængig af enhver erfaring. Den udsiger i sig selv ikke noget som helst om hvorvidt den i sine anvendelser kan faa relation til erfaringsdata eller ikke. Den indebærer i sig selv intet existenspostulat. Den udgør hvad man har kaldt et rent deduktivt system i den forstand at den i sig selv ene og alene muliggør en beregning af de muligheder der følger af de indførte forudsætninger.

2. Teorien indfører visse forudsætninger, om hvilke teoretikeren ud fra forudgaaende erfaring véd at de opfylder betingelserne for anvendelse paa visse erfaringsdata. Disse forudsætninger er af saa stor almenhed som muligt og faar derved chance for at opfylde betingelserne for anvendelse paa et højt antal erfaringsdata.

Det første af disse momenter vil vi kalde teoriens *vilkaarlighed*; det andet vil vi kalde teoriens *hensigtsmæssighed*. Det forekommer nødvendigt at indoptage begge disse to momenter i tilrettelæggelsen af en teori. Men det følger af det sagte at erfaringsdata aldrig kan bekræfte eller afkræfte teorien selv men kun teoriens anvendelighed” (Hjelmsslev, 1934: 14)

Hertil kommer at for Hjelmslev er sproget givet for sprogteoretikeren. Sproget er undersøgelsesmaterialet og intet andet. Hjelmslev beskæftiger sig ikke med hvordan mennesker producerer sætninger, eller hvordan vi forstår dem. I god positivistisk ånd er udgangspunktet (1) en formal teori, (2) et materiale hvorpå teorien kan efterprøves – og intet andet. Det er dette udgangspunkt der gør Hjelmslevs teori til et godt afsæt for nærværende undersøgelse hvor det er morfemernes udtryksform der søges opstillet på baggrund af udtrykssubstansen, og hvor spørgsmålet om korrelation mellem sprogbrugere og grammatik kun er interessant på et overfladeniveau.

I forhold til Chomskys *explanatory adequacy* kan man sige at Hjelmslevs sprogteori ud fra traditionelt naturvidenskabelige, positivistiske kriterier søger at opstille en deskriptiv teori – en sprogteori der kan beskrive systemet bag empirien, men som ikke påstår at forklare noget om sprogbrugerens forhold til sproget. Som Hjelmslev skriver i citatet ovenfor er teorien delt i to. Den ene del er erfaringsuafhængig¹ (teoriens vilkårlighed), og den anden del er erfaringsafhængig (teoriens hensigtsmæssighed). Relationen mellem det erfaringsuafhængige og det erfaringsafhængige er muligvis værd at belyse historisk i første omgang. Den samtidige Alfred Ayer indleder sit positivistiske 'programskrift' *Language, Truth and Logic* med følgende i forordet til første udgave:

"Like Hume, I divide all genuine propositions into two classes: those which, in his terminology, concern 'relations of ideas', and those which concern 'matters of fact'. The former class comprises the *a priori* propositions of logic and pure mathematics, and these I allow to be necessary and certain only because they are analytic. That is, I maintain that the reason why these propositions cannot be confuted in experience is that they do not make any assertion about the empirical world, but simply record our determination to use symbols in a certain fashion. Propositions concerning empirical matters of fact, on the other hand, I hold to be hypotheses, which can be probable but never certain. And in giving an account of the method of their validation I claim also to have explained the nature of truth." (Ayer, 1936: 41)

Der er her en korrelation mellem Hjelmslevs vilkårlighedskrav og Ayers a priori udsagn. Fælles for begge er uafhængigheden af den menneskelige erfaring. Ayer formulerer det således "that they do not make any assertion about the empirical world" samtidig med at han præsupponerer at de ikke kan gendrives gennem erfaring, mens Hjelmslev bruger formuleringen "[teorien] udsiger i sig selv ikke noget som helst om hvorvidt den i sine anvendelser kan faa relation til erfaringsdata eller ikke. Den

¹ Jeg vil ikke inlade mig på en erkendelsesteoretisk diskussion af hvorvidt der eksisterer noget erfaringsuafhængigt overhovedet da en sådan diskussion vil være et unødigt og i et vist omfang ligegyldigt angreb på teorien uden for dens egne præmisser.

indebærer i sig selv intet existenspostulat”, og i øvrigt tilføjes ”[m]en det følger af det sagte at erfaringsdata aldrig kan bekræfte eller afkræfte teorien selv”.

Hjelmslev anerkender dog at en del af teorien hviler på erfaringens grundlag, nemlig den del af teorien hvor ”visse forudsætninger” indføres ”om hvilke teoretikeren ud fra forudgående erfaring véd at de opfylder betingelserne for anvendelse på visse erfaringsdata”. Her står specielt ordet ”visse” i modsætning til den ellers meget præcise og formalt formulerede sprogteori. Det viser sig senere i OSG at de forudsætninger der er tale om, er Hjelmslevs principper der alle er deduceret ud fra empiriprincippet² som han anser for ikke at være af specifik aksiomatisk natur (ibid. s. 15). Empiriprincippet defineres i OSG således at ”[b]eskrivelsen [af sprogsystemet] skal være modsigelsesfri, udtømmende og den simplest mulige” (s. 12). Dette har således en vis lighed med den anden type af ægte udsagn som Ayer henviser til, nemlig de udsagn der angår empiriske kendsgerninger.

Bl.a. med skelen til empiriprincippet kan sprogteorien anses for at være aksiomatisk, hvilket Hjelmslev for så vidt heller ikke benægter. Det han hævder i OSG, er derimod at teorien ikke indfører *specifikke* aksiomer (s. 15, 20, 32). Hjelmslev kredser lidt tættere om adledet *specifikke* i sin definition af konstanter og variable (s. 32):

”Ved en *konstant* vil vi forstaa et funktiv hvis tilstedeværelse er en nødvendig betingelse for tilstedeværelsen af det funktiv som det har funktion til; ved en *variabel* vil vi modsat forstaa et funktiv hvis tilstedeværelse ikke er en nødvendig betingelse for tilstedeværelsen af det funktiv som det har funktion til. Disse definitioner forudsætter visse ikkespecifikke indefinable: *tilstedeværelse*, *nødvendighed*, *betingelse*, og definitionerne af funktion og af funktiv” (min understregning, forfatterens kursiveringer).

Her anvendes ordet *ikkespecifikke* om en række undefinerede ord. Et normalt argument for at al videnskabelig teori er af aksiomatisk natur, er at for at kunne slutte fra *b* til *c*, må man kende *b* – man kan kun definere *b* ved at lade *b* være resultatet af en slutning, fx fra *a* til *b* hvorved problemet blot er skubbet over på en definition af *a*, eller ved at fastsætte *b* som aksiom – dvs. som en ubeviselig grundsætning man antager er gældende. Med tilføjelse af ordet *specifik* angiver Hjelmslev at inden for et system, her sprogteorien, er der ikke tale om at han opstiller aksiomer; de aksiomer han opstiller, er

² Jf. bl.a. ”Valget af inddelingsgrund [i.e. ’analysegrundlag’] vil ud fra hensigtsmæssighedshensyn (dvs under hensyntagen til de tre i empiriprincippet indgående krav) kunne blive forskelligt over for forskellige tekster” (s. 21).

derimod af erkendelsesteoretisk natur og henhører ikke under sprogvidenskaben, de er dog ikke desto mindre aksiomer - men aksiomer af almen, og ikke specifik, naturⁱⁱⁱ.

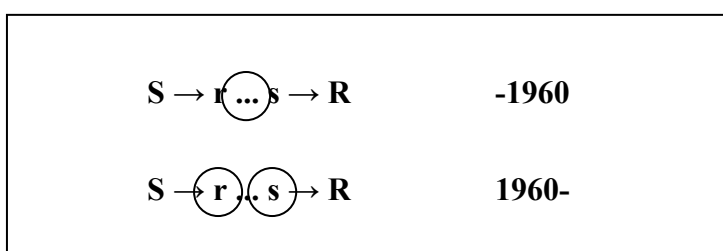
2.2 SPROGFORSTÅELSER

Hvis man skal sige noget om Hjelmslevs sprogforståelse – eller sprogontologi, kan det konkluderes at hans sprogteoretiske skrifter af princip ville være et dårligt sted at lede. Hjelmslev beskæftiger sig ikke i sprogteoretisk sammenhæng med hvad sprog *er*, da det er uden relevans for teorien. Sproget foreligger, og Hjelmslevs antagelse – og dette er netop kun en antagelse (OSG, s. 11) – er at sproget forekommer som substans, men at man kan afdække et system der er bestemmende for denne substans' forekomst, - *men* denne teori udgøres af et i ordets strengeste forstand rent *formalt* system der kun udgøres af relationer uden 'betydning'. Når Hjelmslev kalder noget for *forløb* og noget andet for *system* osv., understreger han at ordene *forløb* og *system* kun er valgt af hensigtsmæssige årsager – og udelukkende har en formal definition. Det ville ikke ændre noget ved teorien om han kaldte begreberne ved numre eller gav dem en alfanumerisk kode (som han gør i *Résumé of a Theory of Language*). Man kunne således karakterisere Hjelmslevs teori som en matematisk inspireret lingvistisk kombinatorik.

Heri ligger endvidere at relationen mellem 'indholdsplan' og 'udtryksplan' (begreberne defineres senere) er formal. Hjelmslev angiver at indholdsplan og udtryksplan gensidigt forudsætter hinanden ('interdependens'), men hvordan indholdsplanet og udtryksplanet kædes sammen i den faktiske sproglige produktion og analyse, - og om begreberne overhovedet har betydning for en mere kognitiv teori om sproglig produktion, er irrelevant i Hjelmslevs sprogteori. I sprogteorien behandles sproget som et isoleret undersøgelsesmateriale uden hensyntagen til menneskelig aktivitet. Parallellen til matematik bliver endog helt tydelig hos Uldall (1957) hvor glossematikken konsekvent bliver henvist til som en algebra.

Forskellen til de nyere sprogteoretikere, ovenfor repræsenteret ved Chomsky og Dik, kan anskues ved Leonard Bloomfields berømte model for sproglig interaktion (Bloomfield, 1933: 26). Bloomfield opererer med begreberne *stimulus* og *reaktion* hvorimellem der eksisterer et kausalitetsforhold således at når en person påvirkes af en stimulus, afføder dette en reaktion. En ikke-sproglig handling kan således beskrives som

$S \rightarrow R$. Ved en sproglig handling derimod har mennesker mulighed for at påvirke andre mennesker til at udføre en sådan reaktion. Den sproglige handling består dels af en talen og af en høren, og disse repræsenterer Bloomfield ved de små bogstaver s og r hvorved den sproglige handling kan formaliseres som $S \rightarrow r \dots s \rightarrow R$. Overordnet set kan man hævde at de ældre grammatikere interesserer sig for 'sproget i sig selv' hvilket måske nok er en uheldig frase, men ikke desto mindre er meget sigende, for mens grammatikkerne i den første halvdel af århundredet havde fokus på sprogsystemet, har man i sprogteori i tiden efter været stærkt koncentreret om selve den sproglige produktion og analyse. Dette kan skitseres i Bloomfields model således:



Der skal nødvendigvis tages en række forbehold her. Året 1960 er naturligvis for stift, men det er valgt ud fra at drejningen mod den talende for alvor tog til efter Chomskys *Syntactic Structures* (1957) og *Aspects* (1965). Herudover har der været tale om en glidende overgang. Mange tidligere, prominente forskere var meget fokuserede på den talende, fx Roman Jakobson, mens mange deskriptivister i dag stadig behandler sprog som man if. modellen gjorde førhen da de mere moderne teorier ofte kan være for tidskrævende at arbejde med hvis fokussen er deskriptivt og ikke teoretisk. Sidst kan det nævnes at der i modellen for 1660- er sat to ringe, men man må sige at det klare fokus var været på produktionen af sprog, og ikke på analysen. De to ringe er således en tillempning.

Når der på denne vis tages nogle forbehold, kan udviklingen dog også ses. Som argumenteret ovenfor er Chomsky og Diks fokus centreret omkring hvad en person der taler, gør – uanset om man kalder personen *Jones* eller NLU. En af de mere moderne danske teorier findes eksempelvis i Togeby's *PRAXT* hvor han bl.a. forkaster Diderichsens grammatik som grundlag for en automatisk syntaksanalyse ud fra en argumentation for at denne ikke har stor nok lighed med den faktiske sproglige produktion (Togeby, 1993: 105).

Den tidligere sprogvidenskab har ikke fokuseret nær så meget på den enkelte talers produktion og analyse af sætninger. Dette gælder med ovenstående argumentation for Hjelmslev, men er også karakteristisk for fx Zellig Harris der konstruerer formale procedurer til at beskrive sprog med ud fra et deskriptivt mål, for Bloomfield selv, og her i landet er det vidt gældende for fx Kr. Mikkelsen, Byskov, H. G. Wiwel, Aage Hansen (med mange flere) at sproget behandles som et materiale der registreres og analyseres systematisk og uden hensyntagen til den sproglige produktion^{iv}.

2.3 SPROGVIDENSKABELIGT UDGANGSPUNKT

Med en karakteristik af Hjelmslevs teori som en slags kombinatorik udgør den et solidt fundament for nærværende undersøgelse. I projektets titel, *Morfemformernes distribution*, er der med ordet *distribution* indlagt et ønske om at afdække regler for morfemformernes mulige forekomster i et sprogs sætninger, dvs. en kombinatorik. Dette ønske kan med største sandsynlighed ikke indfris i nærværende undersøgelse, hvorfor det væsentlige her vil være at kortlægge nogle teoretiske muligheder i et praktisabelt fundament.

Af denne grund vil der også blive taget udgangspunkt i Hjelmslevs begreber, og endsige i hans eksplicite, formale sprogteoretiske tilgang. Begreberne vil blive introduceret løbende som de bliver anvendt. Det vil forsøges at tage kritisk stilling til de metodiske og teoretiske spørgsmål ved i udgangspunktet at lade argumentationen foregå på teoriernes egne præmisser.

2.3.1 Det formalistiske udgangspunkt

Hjelmslevs teori er formal i det omfang at dens definitioner hviler på et formalt system bestående af en række begreber (der kan udtrykkes vha. symboler) der er så præcist definerede at de vil kunne anvendes af en computer. En formaliseret teori beskrives af Chomsky, 1975: 83, således:

”A formalized theory, then, is one that is formulated in accordance with certain clear canons of rigor and precision; definitions are given explicitly in such a way that defined terms are always eliminable, and the axioms and methods of proof are precisely stated. We can thus have a formalized theory purporting to be about form or about meaning.”

Til betegnelsen ”formal” kan man bestemme en sprogteori som formalistisk hvis den alene behandler sprogets form uden hensyntagen til noget betydningsniveau. I denne forstand er Hjelmslevs teori formal, men ingenlunde formalistisk, idet Hjelmslev

konsekvent gør opmærksom på at han også ønsker at behandle indholdsplanet – men stadig formalt (fx OSG s. 63 f. der minder meget om den funktionelle grammatiks 'selection restrictions' – cf. Dik, 1989: 91-97), og endvidere at udtryksplanet og indholdsplanet indgår en interdependensfunktion og dermed er gensidigt afhængige. I nærværende undersøgelse anvendes betegnelsen *rent formalistisk* om det forhold at undersøgelsens eneste empiri er udtryks substansen – med de teoretiske indskrænkninger der sker heraf (jf. afsnit 3.2.2).

Jeg vil således i det følgende antage et formalistisk udgangspunkt. Ikke i den tro at en formalistisk analyse alene kan beskrive et sprogs morfemformers distribution, eller i den tro at en formalistisk teori har større beskrivelseskraft eller –ret i forhold til en mere semantisk eller pragmatisk teori, men med det formål at undersøge den formalistiske teoris muligheder og begrænsninger inden for emneområdet. Hertil kommer at det formalistiske udgangspunkt i nærværende undersøgelse adskiller sig væsentligt fra den generative transformationsgrammatik (hvilken i sprogvidenskaben til tider anvendes synonymt med formalisme), bl.a. deri at der i sagens natur ikke vil blive anvendt transformationer af nogen art.

2.3.2 Glossematiske definitioner

Før den videre undersøgelse påbegyndes, vil der blive redegjort for de mest grundlæggende definitioner hos Hjelmslev og Uldall. Det er ikke alle definitionerne der senere vil blive anvendt i den mere praktiske undersøgelse, men der vil jævnligt være henvisninger til dem. Da de desuden udgør det teoretiske fundament, vil det desuden være passende at redegøre mere indgående herfor.

Udtryksplanet

Det er karakteristisk for den historiske periode at begreberne *indholdsplan* og *udtryksplan* anvendes i sprogteoretiske arbejder. I nutidige teorier, der i højere grad ikke bare anerkender sproget som mentalt forankret, men også teoretiserer over sproget ud fra en sådan optik, støder man sjældent på begreberne indhold og udtryk. I stedet er det ord som materiale, form, funktion og semantik der forekommer. Det er igen væsentligt at bemærke at Hjelmslev ikke i sin teori beskæftiger sig med hvad 'indholdsplanet' og 'udtryksplanet' er. Derimod definerer han deres relation formalt således at de er gensidigt afhængige, dvs. er solidariske. I *Résumé of a Theory of Language* skriver

Hjelmslev: "The designations CONTENT PLANE or PLEREMATIC (symbol: γ°) and EXPRESSION PLANE or CENEMATIC (symbol: g°) are assigned arbitrarily as distinct names for Planes whose number within a Semiotic is two and only two" (s. 95, jf. desuden OSG, s. 55).

Selv om der er tale om en rent formal definition, kan man dog sige, at for at definitionen er hensigtsmæssig (i Hjelmslevs egen forstand), må det antages at tegnsystemet deles i to og kun to 'planer'. Hermed er det også antaget at tegnsystemet ikke består af mere end to overordnede planer (hvis teorien skal være anvendelig på det sproglige materiale). Indholdsplanet og udtryksplanet består af to dele: et forløb og et system. Til eksempel kan man tage et tilfældigt ordpar: *bag* og *lag*. De to ords forløb udgøres af sameksistensen mellem henholdsvis bogstaverne b-a-g og l-a-g. Der eksisterer således en *både-og*-funktion mellem delene i et forløb. De to ord adskiller sig dog også fra hinanden ved at bogstavet *b* indleder det ene ord, mens bogstavet *l* indleder det andet ord. Endnu et ord var fremkommet hvis bogstavet *s* havde indledt det sproglige forløb: s-a-g, osv. Mulighederne for kommutation udgør et system, og funktionen mellem størrelserne i systemet er en *enten-eller*-funktion. *Både-og*-funktionen benævner Hjelmslev *relation*, mens han benævner *enten-eller*-funktionen *korrelation*.

Funktion og funktionstyper

Den sidste del af Hjelmslevs teori der skal fremstilles her, er hans funktionsbegreb og de funktionstyper han skitserer. På baggrund af inddelingen af udtrykket i et forløb (udtrykssubstans) og et system (udtryksformen) opregner Hjelmslev 3 funktionstyper. Alle funktionstyperne er fremkommet som relationer mellem en konstant og en variabel. Konstanter og variable definerer Hjelmslev således:

"Ved en *konstant* vil vi forstå et funktiv hvis tilstedeværelse er en nødvendig betingelse for tilstedeværelsen af det funktiv som det har funktion til; ved en *variabel* vil vi modsat forstå et funktiv hvis tilstedeværelse ikke er en nødvendig betingelse for tilstedeværelsen af det funktiv som det har funktion til" (OSG: 32).

Ud fra definitionerne på konstant og variabel defineres *interdependens* som en funktion mellem to konstanter, *determination* som en funktion mellem en konstant og en variabel (og udtrykkes ved at variabelen (*v*) determinerer (\rightarrow) konstanten (*c*)) og *konstellation* som en funktion mellem to variable. Denne funktion er knyttet an til et krav om

tilstedeværelse således at hvis to funktiver er interdependente, er deres tilstedeværelse indbyrdes afhængige – eller med andre ord: hvis c_1 forekommer, forekommer c_2 . Hjelmlev benævner videre de tre funktionstyper, interdependens, selektion og konstellation, forskelligt alt efter om der er tale om en funktion inden for forløbet, inden for systemet eller i al almindelighed. Denne sondring vil jeg ikke medtage her da den ikke som sådan har relevans i nærværende sammenhæng (men se afsnit 4.33 for en uddybning).

Opsummerende vil jeg hæfte mig ved Hjelmlevs skitsering af det syntagmatiske (forløbet) i relation til det paradigmatiske (systemet) og ved hans 3 funktionstyper hvis relevans for analysen senere vil blive søgt retfærdiggjort.

3.0 Repræsentationsniveauer

Som tidligere omtalt vil jeg argumentere for at en ren formalistisk procedure til at opstille fonemer ikke er mulig. Tilsvarende vil jeg argumentere for at en ren formalistisk procedure til at klarlægge sætningsleddenes syntaktiske funktioner i et sprog ikke er mulig. Idet jeg derimod ikke kan finde afgørende argumenter for at det ikke skulle være muligt i et vist omfang at udfinde et sprogs morfologiske grammatiske system ved hjælp af en rent formalistisk procedure, vil jeg i det videre gøre nogle indledende overvejelser over denne problematik.

3.1 SPROGETS FONEMNIVEAU

Fonemer defineres ofte som de mindste betydningsadskillende tegn. De er således de mindste tegn med en distinktiv funktion. I Hjelmsslevs definition af det sproglige tegn angiver han at udtrykket og indholdet er interdependente. Dette gælder således også for fonemerne. Hjelmsslev henviser til Pragerskolens (Trubetzkoy's) fonemdefinition som han fremhæver på bekostning af London-skolens (repræsenteret ved Sweet). Trubetzkoy definerer *fonologiske oppositioner* (eller 'fonologiske distinktive oppositioner') som lydmodsatninger der i et givent sprog kan differentiere to ords betydning ('intellektuelle Bedeutungs', Trubetzkoy 1939: 30). Ud fra denne definition definerer Trubetzkoy begrebet *fonologiske enheder* som de enkelte led der indgår i en fonologisk opposition (s. 33), og slutteligt definerer Trubetzkoy *fonemer* som fonologiske enheder der i det givne sprog ikke kan forekomme som kortere på hinanden følgende fonologiske enheder (s. 34).

Det grundlæggende for fonemer er således if. Trubetzkoy dels den distinktive opposition og dels deres længde (jf. fonemdefinitionen). Det oppositionelle ligger som en forudsætning for den distinktive. Hvis ikke der eksisterer en opposition (i den definerede forstand) til en lydrække, har den ikke i sig selv nogen distinktiv funktion.

Derimod kan der godt forekomme en opposition som ikke har nogen distinktiv funktion, nemlig i form af frie såvel som bundne fonemvarianter. Til fonemdefinitionen mangler således kun ordet *distinktiv* at blive defineret. Den distinktive funktion benævner Trubetzkoy *betydningsadskillende* (s. 29). Det betydningsadskillende består i at alene én lyd (fonologisk enhed) adskiller en betydningsbærende enhed fra en anden. Idet det grundlæggende for fonemer således er deres distinktive funktion, er de som udtryksform i Hjelmslevs forstand arbitrært forbundet i en interdependensfunktion med det sproglige indholdsplan. Og den eneste måde at udlede fonemer på i et sprog er ved at undersøge hvornår lyd er betydningsadskillende.

Hvis man antager opdelingen af udtryk i invarianter og varianter, er fonemer invarianter, mens varianterne er fonemernes forekomst i udtrykssubstansen som frie og/eller bundne varianter, dvs. som enkeltlyd. Invarianterne (fonemerne) er således en del af et sprogs udtryksform, mens varianterne (enkeltydene) er en del af et sprogs udtrykssubstans. Idet fonemer er defineret ved at være distinktive (jf. Trubetzkoy og Hjelmslev), følger det at for på baggrund af udtrykssubstansen at udlede fonemer, er det en nødvendighed at kende til indholdsplanet. Et velkendt eksempel er det franske sprogs alveolære frikativer der veksler mellem [\pm STEMT] i ordparret [*pwazɔ̃ŋ*] *gift* og [*pwazɔ̃ŋ*] *fisk*. En rent formalistisk analyse vil derfor altid være nødt til at registrere hver eneste variant som invariant. Idet mængden af varianter går mod uendelig for mængden af sætninger gående mod uendelig³, og idet mængden af sætninger er givet (naturligt) sprogs grammatik kan generere, anses for at være infinit, må en ren formalistisk analyse (uden inddragelse af indholdsplanet) af enkeltlyd i fonemer anses for umulig hvis intet andet er givet end en udtrykssubstans. Dette er også en følge af Hjelmslevs tidligere citerede passage:

”En invariantforskel er til stede i udtryksplanet, naar der dær bestaar en korrelation ... som modsvares af en korrelation i indholdsplanet ... paa en saadan måde at der består en *relation* mellem udtrykskorrelationen og indholdskorrelationen” (OSG: 59)

3.1.1 Modargumentation

Med denne teoretiske argumentation skal det sidste ord dog ikke være sagt. Som gengivet ovenfor angiver Trubetzkoy at fundamentet for fonemdefinitionen er de

³ $M(\text{var}) \rightarrow \infty$ for $M(\text{stn}(\text{var})) \rightarrow \infty$

distinktive oppositioner. Men med til funktionen hører også lydlængden. Med 'lydlængde' henvises her til fonemerne som fonologiske enheder der i et givet sprog ikke kan forekomme som kortere på hinanden følgende fonologiske enheder. Det er denne sidste del af fonemdefinitionen Zellig Harris lægger vægt på i begyndelsen af *Methods in Structural Linguistics*. Harris forsøger at opstille en procedure hvorved lingvisten kan udfinde fonemer på baggrund af en lydrække. Den øvre længde af fonemerne kan udfindes ved at observere en sætning hvis segmenter er analyseret som *DEF*. Lingvisten konstruerer nu en ytring bestående af segmenterne *DA'F*, hvor *A'* er en gentagelse af segment *A* i en ytring repræsenteret ved *ABC*. Hvis en informant accepterer *DA'F* som en gentagelse af *DEF*, eller hvis en informant siger *DA'F* i en situation der tillader lingvisten at vurdere at ytringen er ækvivalent med *DEF*, og hvis lingvisten tilsvarende kan observere *E'BC*, hvor *E'* er en gentagelse af *E*, som ækvivalent til *ABC*, så kan det konkluderes at *A* og *E* er ækvivalente som frie varianter (Harris: 30 f.). Ethvert segment der kan undergå denne test vil være maksimalt lige så langt som ét fonem (ibid. s. 35). Hvis nu testen gælder ikke bare for to segmenter, men for alle segmenter der indeholder samme lyd, da kan det if. Harris udledes at der er tale om et fonem.

Problematikken er dog forskellig fra den nærværende hvor udgangspunktet er at den eneste viden om det givne sprog stammer fra sprogs substansen og intet andet. Harris' udgangspunkt er deskriptivt lingvistisk, og han forudsætter således eksistensen af en informant, dvs. nogle ekstralingvistiske informationer. Selv om det således er længden der ligger til grund for inventarieringen af fonemer, er det en præmis at der er nogle ekstralingvistiske informationer til stede i proceduren.

På trods af denne forskel i præmisser, er det dog stadig en mulighed at Harris' procedure kan anvendes. Hvis man kan udfinde en mængde enkeltlyd og kategorisere dem som fonemer ud fra en procedure der minder om Harris', så vil disse ganske vist ikke være fonemer, men spørgsmålet er om det vil være muligt at udfinde et finit antal, fx 500 potentielle fonemer. Det vil naturligvis vanskeliggøre proceduren, men det vil ikke umuliggøre den – hvis det altså overhovedet er muligt at finde et finit antal. Ud fra Trubetzkoy's fonemdefinition vil spørgsmålet tilsvarende være om man så at sige kan springe det første niveau over (niveauet med inventariering af oppositionelle distinktioner), og forsøge direkte at opstille "fonemer". Idet disse ikke er funderet i en

oppositionel distinktion er der dog ikke tale om egentlige fonemer, men blot om enkeltlyd bestående af varianter af en række mulige fonemer.

Denne udvej vil det dog være mere nærliggende at afprøve i praksis frem for teoretisk. I det videre vil det ud fra den tidligere argumentation være antagelsen at det ikke er en mulighed.

3.2 SÆTNINGSNIVEAUET: SYNTAKTISK FUNKTION

På sætningsniveau kædes en analyseret udtrykssubstans sammen med en række syntaktiske funktioner. I Diderichsens grammatik er udtrykssubstansen analyseret og reduceret til ledtyperne *nominal*, *verbal* og *adverbial*. I forhold til deres indbyrdes forekomst og placering (topologi) tilskrives leddene en syntaktisk funktion. Spørgsmålet er nu om denne procedure for tilskrivning af syntaktisk funktion til helheder i sætningen kan udledes formalistisk. Hvis problematikken anskues inden for Hjelmlevs begrebsramme, er der tale om at funktionen mellem en udtryksform og en korrekt indholdsform skal kortlægges. Hjelmlev angiver at funktionen mellem udtryk og indhold er interdependens, så selve funktionen er der ikke tvivl om ud fra denne optik. Problematikken består derfor i at tilskrive den korrekte indholdsform.

I den skitserede terminologi ovenfor er det angivet at det er en udtryksform der er udgangspunktet. Det indebærer bl.a. at det er givet som præmis at der skal foreligge en 'færdiganalyseret' udtryksform (fonemstreng) på baggrund af en udtrykssubstans. Som argumenteret for i afsnit 3.1 vil det her være antagelsen at dette ikke er en mulighed. Hvis denne problemstilling imidlertid ignoreres, kan der dog fokuseres på andre udfordringer.

Udtryksformen kan i dette tilfælde kun udledes på baggrund af kortlagte, mindre udtryksformstørrelser. Hvis dette kan lade sig gøre i et sprog, vil det resultere i nogle formmodeller. En forsimplet model for dansk kunne på baggrund af Diderichsens beskrivelse af helheder forelægges som:

NOMINAL:	ARTIKEL	ADJEKTIV	SUBSTANTIV
NOMINAL:	NUMERALE/PRONOMEN		
ADVERBIAL:	PRÆPOSITION		NOMINAL
ADVERBIAL:	ADVERBIUM		
VERBAL:	VERBUM		

Ordene *artikel, adjektiv, substantiv, præposition, adverbium og verbum* er valgt på baggrund af deres hensigtsmæssighed. I beskrivelsen vil de være rent formalt defineret på baggrund af deres forekomst i helheden. Hvis følgende sætning analyseres efter disse helheder, fås nedenstående segmentering:

1789 døde mange franskmænd
den dag døde mange franskmænd
1789 døde mange franskmænd
den dag døde mange franskmænd

Begge med følgende ledtyper:

*N-V-N

Der resterer dog stadig et spørgsmål om hvorvidt man formalistisk vil kunne udlede at *1789* og *den dag* ikke er nominaler selv om de er opbygget som de fleste nominaler, men blot mangler en foranstillet præposition.

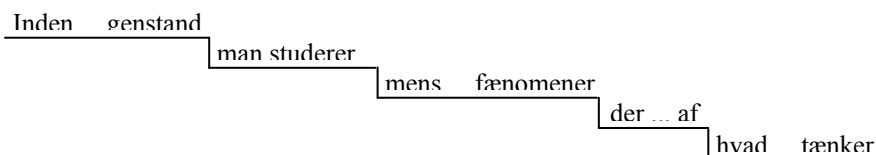
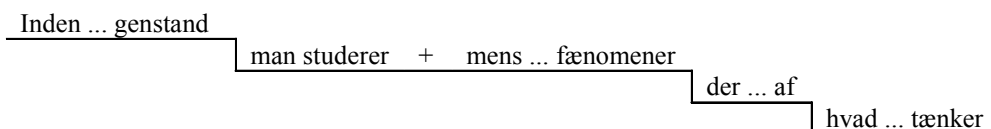
Et andet problem for den syntaktiske analyse er analysen af adverbielle adled til nominalhelheder, eksempelvis i sætningerne:

jeg spiser kartofler med rødkål
jeg spiser kartofler med bestik

Sætningerne er formalt set opbygget af samme størrelser, men analysen er alligevel forskellig. På sætningsniveau kan slutteligt endnu et problem illustreres. Der kan sættes spørgsmålstejn ved hvorvidt man rent formalistisk dels kan afgøre ledsætningers funktion og dels kan afgøre hvilken grad de er af. Den første problematik aktualiseres af sætninger med homonyme bisætningsindledere. Eksempelvis kan *hvor* indlede både adverbielle (*hvor vejen slår en bugt, ligger et lille hus*), substantiviske (*jeg ved hvor vejen slår en bugt*) og adjektiviske (*vi købte grunden hvor vi senere byggede huset*) bisætninger. En anden problematik optræder i følgende sætning:

Inden for samfundsvidenskaben er tankegangen en del af den genstand man studerer, mens naturvidenskaben beskæftiger sig med fænomener, der er afhængige af, hvad nogen tænker (Soros, 1998: 9)

To forskellige periodeanalyser for sætningen kan gengives således:



Den øverste analyse må anses for at være den korrekte. Det forekommer os meget sandsynligt, og normalt vil man formentlig ikke overveje analyse nummer 2. Analyse 1 kan endvidere underbygges hvis sætningen omformuleres: *inden for samfundsvidenskaben er tankegangen en del af genstanden, mens naturvidenskaben beskæftiger sig med fænomener der er uafhængige af hvad nogen tænker*. Her vil de fleste sprogbrugere formentlig blive enige om at meningen af den oprindelige sætning svarer til meningen af den omskrevne sætning. Dette ændrer dog ikke på at ud fra rent formalistiske kriterier vil det være vanskeligt at afgøre hvilken af de to analyser der er den korrekte. Endnu et problem opstår ved Ellipse, fx:

”Den anden [hånd] bruge han til små gestus, der blev forstået og adlydt af plejemoder og de andre trænede sygeplejersker” (Hans Scherfig: *Det forsømte forår*, min understregning).

Ordet *trænede* er homonymt mellem en præteritum-form og en adjektivisk participium-form (kongruerende med *sygeplejersker* i numerus). Ved den første analyse bliver understregningen en selvstændig sætning, mens den ved den anden analyse bliver en del af en elliptisk konstruktion hvor præpositionen er udeladt. Igen vil den formalistiske analyse være i vanskeligheder.

Alle disse problemer er fremlagt som problemer specifikke for dansk, men uden at gå i detaljer her vil man kunne finde lignende problematikker i formodentlig alle sprog.

3.2.1 Generelle problemer

Ud over disse konkrete problemer for dansk er der også andre, mere overordnede problemstillinger hvis proceduren skulle gælde for flere, potentielt alle, sprog. Hvis man ønsker at lave en procedure for at finde de syntaktiske funktioners distribution i en sætning, kunne man angive følgende grove udledelsesprocedure^v:

- (1) Det sætningsled der kan udgøre en sætning alene, er verbalet
- (2) Det sætningsled der alene med verbalet kan udgøre en sætning, er subjektet
- (3) Det sætningsled der kan alene med subjektet og verbalet kan udgøre en sætning, er objektet
- (4) De led der kun kan forbinde sætninger, er konjunktioner
- (5) Alle øvrige sætningsled er adverbialer

Her er problematikken mht. indirekte objekt og prædikativer udeladt. Hele definitionssystemet hviler her på punkt (1). Den første indvending kunne være at der if. nogle teoretikere ikke eksisterer verbaler i alle sprog. Desuden kan man i en mængde sprog udelade kopulaverbum i subjekt-prædikatsforbindelser, fx i nytestamentligt græsk: αἰρετον τ+μαχτ* ισα γινῃται ωἰ οἰδιδασκαλ οἱ αυτου (tilstrækkelig:AKK discipel:DAT:DEF at bliver:3SG som mesteren:NOM hans) *∴ det må være tilstrækkeligt for disciplen hvis han bliver som sin mester* (Matth. 10,25), og i arabisk (moderne standard): *al-bint jamīl* (pige:DEF smuk) *∴ pigen er smuk*.

Punkt (2) kan også drages i tvivl. I flere europæiske sprog indeholder verbalet et personmorfem hvorfor subjektet i sætningen kan udelades, eksempelvis på latin:

Amitti	merito	proprium	qui alienum	adpetit ⁴
V:PRÆS:3SG	ADV	ADJ:AKK	K ADJ:AKK	V:PRÆS:3SG
<i>stræber-han</i>	<i>med rette</i>	<i>egen</i>	<i>som fremmed</i>	<i>stræber-efter-han</i>
<i>∴ han mister med rette sig selv, som stræber efter at tilhøre en anden</i>				

Her forekommer subjektet ikke som selvstændigt ord, hvorfor regel (2) i proceduren undermineres (tilsvarende eksempler forekommer i en række morfologiske sprog). Den tredje regel udsættes dog også for problemer ud over de allerede skitserede. I nominativ-akkusativ-sprog er der en direkte lighed mellem kasus (udtryksform) og syntaktisk funktion. I (absolutiv-)ergativ-sprog derimod er denne direkte lighed ikke tilstede. Et eksempel fra Comrie (1981: 106) er:

<i>balan d'ugumbil</i>	<i>banḡul yarḡḡu</i>	<i>balḡan</i>
kvinde-ABSOLUTIV	mand-ERGATIV	slår
'manden slår kvinden'		

⁴ Phaedrus: *Canis per Fluvium Carnem Ferens*

bayi yara *banin^u*
mand-ABSOLUTIV kom-her
'manden kom her'

I første sætning er det objektet, *kvinde*, der står i absolutiv, mens det i anden sætning er subjektet, *mand*, der står i absolutiv. Dette fænomen forplumres yderligere af at det kun er få sprog der udviser fuldstændig konsistens i deres morfologiske ergativ-systematik (if. McGregor, 2002: 295).

3.2.2 Den syntaktiske analyse

Det synes at være tilfældet at det vil være en særdeles vanskelig proces at konstruere en procedure der rent formalistisk og ud fra udtryksformen kan forudsige de syntaktiske funktioner i et sprog. Hvis man vil acceptere et lidt lempeligere krav, kunne man spørge om det da er muligt rent formalistisk ud fra udtryksformen at konstruere en procedure der kan generere en infinit mængde sætninger på et givent sprog. Herved omgås problematikken også mht. hvad der er den 'korrekte' syntaktiske analyse. Kan det med andre ord lade sig gøre givet udtryksformen og via en rent formalistisk analyse at opstille en procedure der vil forudsige helhedernes forekomst i en infinit mængde sætninger på det givne sprog? – Sætninger der godt nok ikke vil være semantisk korrekte, men vil leve op til et krav om grammatisk korrekthed. Hvis dette vil kunne lade sig gøre, vil man også have en formalistisk definition af hvad en sætning er, og hvordan man udfinder en sætning på et hvilket som helst sprog (cf. Prebensen, 2003: 12).

Problemet med om et adverbial tilhører et foregående nominalet eller ej lader til at blive løst hermed idet proceduren ikke vil se nominalet og adverbialet som ét led, men blot vil registrere i en kombinatorik at der kan følge et adverbial med en bestemt form efter ethvert nominal, sætningen *kartofler med bestik spiser jeg* er jo ikke ugrammatisk i denne forstand, selv om den er semantisk underfundig. Det er heller ikke usandsynligt at ergativ-problematikken undgås idet proceduren i sin kombinatorik kan angive de mulige forekomster af sætningsled i en sætning.

Dog lader der til stadig at være uløste problematikker. Hvis det er tilfældet at der i sprog findes sætningsled hvis funktion ikke kan bestemmes formalistisk, men må bestemmes med inddrages af semantik eller pragmatik, og hvis den formale procedure

ikke kan udlede hvilke helheder der er tale om, vil proceduren være afgørende fejlagtig. Dette er muligvis tilfældet med de danske adverbialkonstruktioner der formalistisk set er opbygget som nominaler (fx *den dag, 1789*) eller ved homonymivanskeligheder (*I lømler skal ikke stjæle mine æbler* vs. *I guder hvor jeg er træt af at de stjæler mine æbler*). Hvis det er tilfældet at sådanne problemer ikke kan lade sig løse efter ovenstående kriterier, vil det ikke være muligt at konstruere en sådan procedure.

3.2.3 Morfologi og syntagmatik i sætningsledene

Hvis man antager at man kunne svare bekræftende på alle de skitserede spørgsmål, ville man få en analyse der begyndte ved enkeltlydene. Herudfra ville man vha. en given procedure kunne opdele enkeltlydene i fonemer. Fonemstrengene ville man kunne opdele i morfemer, man ville kunne sige noget om morfemernes distribution i sætningsled i form af deres forekomst og deres lineariseringsmuligheder, og man ville endelig være i stand til vha. proceduren at analysere sætninger syntaktisk på ethvert givent sprog. Indtil videre er fonemniveauet og sætningsniveauet blevet behandlet, og på basis af de ovenforstående argumenter er synspunktet her at det lader til ikke at være muligt her at opstille en sådan procedure.

Hvis man antager at den fonologiske analyse af et sprog er givet, eller undersøger maskinskrevet skriftsprog frem for talesprog, er det nærliggende at rejse det videre spørgsmål om hvor stor en del af sætningsledene det er muligt at konstruere på baggrund af de foreliggende fonemstrengene. Det er denne problematik der rejses i undertitlen til nærværende undersøgelse. Af den tidligere argumentation for umuligheden af at udlede fonemer på baggrund af en udtrykssubstans og ud fra rent formalistiske kriterier kan det udledes at det spørgsmål der rejses indebærer en umulighed. Væsentligt for argumentationen var at fonemerne er invarianter, mens enkeltlydene var varianter af invarianterne, men idet man ikke kan slutte fra varianter til invarianter vil man få et fonemantal der går mod uendeligt (for antallet af undersøgte sætninger gående mod uendeligt).

Derfor er det allerede nu udelukket at man for morfemernes vedkommende vil kunne slutte fra varianter til invarianter ved den rent formalistiske analyse. Man kunne antage at ordene *café* og *cafe* forekom. Disse to monomorfemiske ord er absolutte synonyme og vil derfor kunne anses for at være to forskellige udtrykssubstanser af

samme udtryksform og med samme indhold. Dette vil proceduren imidlertid ikke kunne registrere. Noget tilsvarende gælder for præteritum på dansk. På dansk er det kun verber der bøjes i tempus, og to eksempler på verber bøjet i præteritum er *legede*, *spiste*. Her må *-ede* og *-te* anses for at være absolutte synonymmer (allomorfer), eller formuleret anderledes: to invarianter i udtrykssubstansen der har samme udtryksform gengivet ved [PRÆTERITUM]. Man kunne argumentere for at verbernes distribution i sætningen ville afgøre at de to former er varianter af samme invariant – de har jo nøjagtigt de samme forekomstmuligheder. Men ikke nok med at de har nøjagtigt samme forekomstmuligheder i sætningen, de har også nøjagtigt samme forekomstmuligheder i sætningen som verber i præsens. Derfor kan man ikke på baggrund af distribution slutte fra x variant(er) til y invariant(er), hvor $x \geq y$. Denne begrænsning er derfor indlejret i procedurens effektivitet, og den kan formuleres generelt således at enhver variant må inventarieres som invariant i den rent formalistiske analyse, ikke fordi den er invariant ud fra sprogvidenskabelige kriterier, men fordi det er den eneste mulighed hvis analysen skal være succesfuld.

4.0 Morfemformernes distribution

Problemfeltet er nu afgrænset i forhold til nogle teoretiske og nogle generelle overvejelser over mulighederne. Det videre problem lyder således: Givet en fonologisk analyse af et sprog, vil det da være muligt at opstille en procedure som rent formalistisk er i stand til at inventariere så mange paradigmer i sproget, så proceduren kan opstille en grammatik for syntagmerne (helhederne), hvor et syntagma er en tekststreng der har en syntaktisk funktion i en sætning på det givne sprog. Denne problemstilling har nogle konsekvenser i forhold til de tidligere præmisser. De parametre proceduren vil være opbygget ud fra, vil være (a) inventariering af *ord*, (b) en beskrivelse af ordenes forekomstmuligheder og lineariseringsmuligheder og (c) en inventariering af *paradigmer* på baggrund af (b). Men selvom problemet nu er afgrænset, er det endnu ikke helt præcist defineret. Før overvejelser over selve proceduren kan det derfor være nødvendigt med nogle definitioner af de sproglige niveauer der skal behandles af proceduren, men for at nå til disse definitioner vil deres grundlag først blive tydeliggjort. Det følgende vil således være opdelt i (1) glossematisk algebra som grundlag for en formalistisk inventariering, (2) de sproglige repræsentationsniveauer som nødvendighed for beskrivelse af syntagmerne, (3) fremstilling af principper for selve proceduren med inddragelse af (1) og (2).

4.1 GLOSSEMATISK ALGEBRA

Grundlaget for Hjelmslev og Uldalls sprogbeskrivelse er ”a theory of non-quantitative functions” (Uldall, 1957: 18). I *Outline of Glossematics* prætenderer Uldall således ikke at der er tale om en ’matematisk lingvistik’. Han sonderer på denne vis skarpt mellem kvantitative og ikke-kvantitative funktioner. a og b kan indgå i en kvantitativ funktion hvis de udgøres af tal, eksempelvis $a > b$ eller $f(a) = b$. Denne type af kvantitative afhængigheder er ikke relevante for den sprogvidenskabelige sprogbeskrivelse i

Hjelmslev og Uldalls forstand, eller rettere: ”a linguist designing an artificial language or a new scientific notation ... is only peripherally concerned with quantitative functions”. Uldall afviser således ikke at kvantitative funktioner kan være givtige i enkelte sammenhænge, men dog perifere i forhold til den samlede sprogteori. Uldall bruger ikke meget plads på at uddybe relationen mellem kvantitative og kvalitative funktioner hvorfor jeg her finder det nødvendigt med enkelte betragtninger.

Den kvantitative funktions område er grundlæggende set tal og tals indbyrdes relation. En fremgangsmåde man ofte støder på, er at lade mængdelæren være grundlæggende^{vi}. En mængde indeholder elementer hvis antal er større eller lig med nul. To elementer i den kvantitative mængde kan karakteriseres som værende større end, mindre end eller lig hinanden. Den kvantitative (matematiske) funktion bliver da en funktion der angiver en relation mellem tal i forskellige mængder. Relationen kan angives ved at tallene (elementerne) i mængden A projiceres over i mængden B udtrykts formalt ved $A \rightarrow B$. Hvis funktionen er givet ved $f(x) = x + 1$, og hvis $A = \{1, 2, 3, \dots, \infty\}$, så vil $B = \{2, 3, 4, \dots, \infty\}$.

Denne tankegang er allerede i et vist omfang anvendt sprogvidenskabeligt direkte. Chomsky skriver i *Syntactic Structures*: ”From now on I will consider a *language* to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements” (1957: 12). Chomsky definerer således sprog som en mængde af sætninger, dette nøjagtigt som Church’ og Posts anvendelse af mængdelæren i forhold til velformede formler (Prebensen, 2003: 14). Disse sætninger, skriver Chomsky, er (1) finite i længde og (2) konstrueret ud fra en mængde af finite elementer (en grammatik). Ud fra denne definition kan man se fx hele, positive tal (dvs. naturlige tal) som et sprog, såvel som man kan se dansk som et sprog^{vii}. Naturlige tals grammatik er konstrueret således at tallene er genereret af kombinationer af tallene fra 0-9. Cifret længst til højre i et naturligt tal ganges med 1, næste cifre (mod venstre) ganges med 10, næste igen med 100, osv. Denne regel kan reduceres til følgende procedure: hvis man ser tallet fra højre mod venstre, ganges første tal på venstre side af et eventuelt komma med 10^0 , de følgende mod venstre ganges med 10 opløftet i forrige potens + 1, mens de følgende tal mod højre ganges med 10 opløftet i forrige potens – 1, og tilslut lægges tallene sammen:

10^n	...	10^2	10^1	10^0	10^{-1}	10^{-2}	...	10^m
--------	-----	--------	--------	--------	-----------	-----------	-----	--------

Heraf navnet decimalsystem. Chomsky forstår da begrebet sprog meget bredt og argumenterer således for at alle disse bredt definerede sprog grundlæggende set består af samme type mængder.

Hvis man ser på elementerne i mængderne, er de dog væsensforskellige. Matematikerne beskæftiger sig i matematikken grundlæggende set med tal. Lingvister beskæftiger sig derimod grundlæggende set med fonemer, ud fra et datalogisk synspunkt en forskel mellem tal og strenge. – Eller udtrykt i Hjelmsevs optik: De mindste størrelser i matematikken er for matematikeren tal, mens de mindste størrelser i sproget for lingvisten er fonemer. Denne påstand er muligvis lidt grov, men det er i denne sammenhæng ikke så vigtigt. Det videre spørgsmål er nu hvad forskellen er på tal og fonemer – det Uldall benævner kvantitative og kvalitative funktioner.

I begge discipliner (matematik og lingvistik) er der tale om en række af elementer sammensat efter bestemte mønstre. I sprogvidenskaben giver det imidlertid sjældent mening at sige at $x < y$, mens det i matematikken ofte vil give mening. Uldall formulerer det således:

”We have noted that the prototype of statements in physics is ” a is greater than b ”; the corresponding statement in glossematics is ” a presupposes b ”, which in a similar way serves to establish an ordered dependence of great generality” (1957: 18).

En lighed mellem det matematiske og det sprogvidenskabelige udtryk er således ifølge Uldall to størrelser der indgår et afhængighedsforhold. Derfor er hans funktionsdefinition også:

”By a *function* is understood any dependence” (1957: 37)

Hjelmslev og Uldalls funktionsdefinition adskiller sig således en anelse fra den klassiske matematiske funktionsdefinition hvor en funktion er en afhængighed mellem konstante og variable størrelser (denne matematiske funktionsdefinition minder i øvrigt påfaldende meget om Hjelmsevs begreb *manifestation* som en selektion (specifikation) mellem hierarkier hvor formen er konstanten i specifikationen, mens forløbet er variabelen, jf. Hjelmsevs, 1975: 12).

Afhængigheden udtrykkes da i sprog som en afhængighed mellem (tekst- eller fonem-)strenge og ikke mellem tal. Talsystemet behandles i matematikken som et lukket system med universelt gyldige regler hvor man, så længe man holder sig til disse

regler, kan blive ved med at konstruere nye tal og talrelationer. I sprog derimod beskæftiger lingvisten sig med faktisk forekommende sætninger i form af fonemstreng. Mens matematikken således kan karakteriseres som en empirisk eller rationel videnskab, kan grammatikken karakteriseres som en empirisk videnskab. Det grundlæggende afhængighedsforhold mellem fonemer eller sekvenser af fonemer er således forekomst, et begreb det ikke giver megen mening at operere med i matematikken. Hvis man skulle anskue det matematiske udtryk $2 + 3$ ud fra en lingvistisk optik, kunne man sige at det relevante er (1) hvilken mængde 2 tilhører, (2) hvilken mængde 3 tilhører og (3) at forekomsten af '+' medfølger at der med nødvendighed må forekomme mindst 1 konstant eller 1 variabel på hver side af '+'-tegnet hvis udsagnet skal være grammatisk. Mens nummer (3) giver god mening at hævde, vil vi ikke nødvendigvis kunne afklare svaret på (1) og (2) idet problemet ikke løses ved en forekomstanalyse, men ved en analyse af den konkrete situation idet der ikke er noget naturgivent i at 2 og 3 tilhører denne eller hin mængde. Spørgsmålet (eller måske snarere svaret) bliver med andre ord meningsløst. I sætningen *jeg elsker legoklodser* giver spørgsmålene imidlertid god mening idet man ved kommutation kunne afgøre hvilke ord der kan stå på *jegs* plads, osv. Og man kunne desuden afgøre hvor i sætningen *jeg*, *elsker* og *legoklodser* kunne placeres.

Den grundlæggende forskel er således at det er forekomstkriteriet der er omdrejningspunktet for den formalistiske natursprogsanalyse. Derfor er den formalistiske natursprogsanalyse også en del af en empirisk videnskab, og kan som sådan (hvis man ønsker en sådan diskurs) ses i relation til fysikken, snarere end matematikken, idet matematikken beskæftiger sig med et lukket, deduktivt og tautologisk system. Dette leder også til at konkludere at den formalistiske inventariering på hvert enkelt sproglige niveau må være baseret på forekomstmuligheder. Dette indebærer nogle konsekvenser som i det videre skal belyses.

4.2 SPROGLIGE REPRÆSENTATIONSNIVEAUER

Det sproglige input består af tekststreng der udgøres af sekvenser af fonemer. Denne fonemforekomst er begrænset i en fonologisk inventariering der udgøres af et fonemalfabet for det givne sprog. Alt dette er givet for analysen idet den ellers ville være umulig. I stedet for fonemsekvenser vil jeg her anvende skriftsprog som input.

Inventareringen af skriftsproget udgøres i første ombæring af det skriftlige alfabet inklusive andre skrifttegn (fx , \$. : ; - samt blanktegn).

Bogstaverne forekommer i sekvenser adskilt af blanktegn, hvor hver sekvens adskilt med blanktegn udgør et ord (bestående af et antal rødder (r), derivativer (d) og fleksiver (f) større end eller lig 0, hvor $r + d + f \geq 1$). Ordene inventarieres, på baggrund af deres bøjning, i ordgrupper der reduceres til et antal ordtyper på baggrund af ordgruppernes forekomstmuligheder i sætninger. De enkelte ordtyper kan siges at indgå to slags afhængigheder, nemlig morfologisk og lineariseret afhængighed. I den morfologiske afhængighed er ordtypers forekomst på baggrund af morfologi inventariet, mens der i den lineariserede afhængighed er tale om en inventarering på baggrund af ordtypers forekomst i syntagmer.

I den morfologiske afhængighed indgår bøjningerne i henholdsvis paradigmatiske og syntagmatiske afhængighed. I den paradigmatiske afhængighed er der tale om et paradigme, P , der udgøres af en mængde af inventarierede elementer identificeret på baggrund af kommutation i paradigmet, mens der i den morfemsyntagmatikken er tale om en afhængighed mellem morfemer i syntagmet.

I lineariseringen er der tale om nogle forekomstregler der afgør enkelte ordtypers placering i en helhed. Helheden består af mindst 1 ordtype. Helheden kan indsættes i dels en paradigmatiske og i dels en syntagmatik. I paradigmatiske anskues helheden som en række mulige realiseringer ud fra kommutation, mens helheden i syntagmatikken anskues som en størrelse der indgår i en syntagmatiske afhængighed med andre syntagmer. Hvis man skal illustrere disse niveauer, kan det gøres i følgende skema:

(1) Morfemparadigmatik (morfologi/bøjning)	Ordtyper defineret ved en mængde af ord med (1) samme morfologi, (2) samme forekomst i et syntagme baseret på kommutation. En ordtype kunne udgøres af appellativer i bestemt form, singularis, som alle har samme morfologi og samme forekomst.
(2) Helhedsparadigmatik	En mængde af syntagmer der er kendetegnet ved kommutation. Eksempelvis indgår følgende blandt andre i en helhedsparadigmatik: {PRON:AKK-SUBST:SG:DEF, ARTKL:PLUR:DEF-SUBST:PLUR:DEF}
(3) Morfemsyntagmatik	Rektion/kongruens uden for syntagmerne defineret ved forekomstafhængighed mellem ordtyper. Eksempelvis er verber i passiv karakteriseret ved ikke at kunne efterfølges af infinitte verber.
(4) Helhedssyntagmatik (topologi)	Regler for forekomsten af selvstændige syntagmer i større helheder (sætninger).

Den første afhængighedstyper vedrører således morfemer der indgår i et og samme paradigme. Den svarer til den klassiske bøjningsslære eller hvad Brøndal (1932: 1) benævner ”*Ordklasser: Partes Orationis: almen Morfologi*”. Som redegjort for stemmer de her behandlede *ordtyper* dog ikke overens med den traditionelle grammatiks *ordklasser*.

Den videre grundtanke er at ethvert paradigme udgøres af en række elementer der kan aktualiseres i sætningen som syntagmer. (2) vedrører da længere helheder (mindst ét ord) som kan grupperes i et paradigme ud fra forekomst, herunder fx Diderichsens helheder (genstandshelheder og beskriverhelheder). (3) vedrører omvendt morfologiske afhængigheder mellem ordtyper i sætningen, fx i det danske verbalsystem og ved kasus. Denne gruppe svarer i nogen grad til Brøndals bøjning (ibid.): ”*Bøjning* (herunder Casus) ... speciel Morfologi”. Mens de fleste formodentlig ville indvilge i inden for denne undersøgelses egne principper at benævne gruppe (1) morfologi, ligger det mere åbent om gruppe (2) og (3) hører under morfologien eller syntaksen. Intentionen med den sidste gruppe er at denne skal indeholde regler for forekomsten af helheder i sætningen. (3) angiver således afhængighed mellem morfemer, mens (4) angiver afhængighed mellem sætningsled da helheder defineres som ledtypehelheder, dvs. en tekststreng uden neksus med en syntaktisk funktion i sætningen.

Det vil videre være tesen at den formalistiske analyse vil kunne afdække afhængigheder på alle niveauer. Spørgsmålet er så i hvilket omfang dette vil være tilfældet.

4.3 PRINCIPPER FOR PROCEDUREN

Det foreløbige arbejde vil nu være grundlaget for mere konkrete overvejelser over nogle mulige principper for proceduren. Principperne vil beskrives systematisk begyndende med en procedure for en inventariering og reduktion på baggrund af den paradigmatiske morfologi, og vil ende med nogle principper for inventariering og reduktion på baggrund af morfemsyntagmatik, og vil fortløbende i opbygningen følge skemaet ovenfor. Efter en fremstilling af proceduren vil den blive udsat for en kritisk evaluering.

4.3.1 Morfemparadigmatik

Som nævnt skal proceduren for inventariering af den morfemparadigmatikken følge dels en inventariering på baggrund af morfologi, og dels en reduktion på baggrund af kommutation. Inventarieringen på baggrund af morfologi vil kræve en sammenlignende analyse af en potentielt infinit mængde ord. Denne sammenligning vil føre til en inventariering af en række mulige morfemer. For at illustrere dette anvendes i første omgang 4 ord: *65billet*, *billetten*, *billet*, *65billetten*:

A	6	5	b	i	l	l	e	t			
B			b	i	l	l	e	t	t	e	n
C			b	i	l	l	e	t			
D	6	5	b	i	l	l	e	t	t	e	n

Fællesmængden af bogstaver er b-i-l-l-e-t. Idet der eksisterer en fællesmængde af bogstaver mellem de fire ord samtidig med at ordene ikke er identiske, består en mængde af ordene af bogstaver som er forskellige fra de øvrige ords bogstavinventar.

Udtrykt formalt fås:

$$A \cap B \cap C \cap D = \text{'billet'}^5$$

$$A - A \cap B \cap C \cap D = \text{'65-'}^6$$

$$B - A \cap B \cap C \cap D = \text{'-ten'}$$

⁵ 'xxxxx' angiver at der er tale om en tekststreng

⁶ '-' angiver hvor $A \cap B \cap C \cap D$ forekommer i forhold til $X - A \cap B \cap C \cap D$

$$C - A \cap B \cap C \cap D = \emptyset$$

$$D - A \cap B \cap C \cap D = \text{'65-'} \ \& \ \text{'-ten'}$$

De således inventarierede forskelle mellem de enkelte mængder og $A \cap B \cap C \cap D$ bliver '-ten' og '65-' og \emptyset . Disse må derfor anses for at være *mulige* morfemer for ordet 'billet' ($A \cap B \cap C \cap D$). I praksis må man dog konstatere at denne procedure i naturlige sprog konsekvent vil producere et for højt antal morfemer. Teoretisk kan man argumentere for dette standpunkt ved at det følger af fonemdefinitionen at der vil forekomme mulige morfemer som faktisk er (ikkemorfologiske) fonemer:

A	v	a	n	D
B	m	a	n	D
C	s	a	n	D
D	l	a	n	D

Her gælder det at $A \cap B \cap C \cap D = \text{'and'}$. De mulige morfemer af denne analyse bliver så 'v-', 'm-', 's-', 'l-', hvilket er en ukorrekt analyse. Derfor må de egentlige morfemer udfindes blandt mængden af mulige morfemer.

Som nævnt tidligere adskiller de leksikalske morfemer og de grammatiske morfemer sig fra hinanden på ét væsentligt punkt. De adskiller sig nemlig derved at de grammatiske morfemer forekommer systematisk, mens de leksikalske morfemer ikke forekommer systematisk. Usystematikken gør sig desuden gældende for fonemernes forekomst (som i ovenstående tilfælde). For at noget skal kunne karakteriseres som et grammatisk morfem, må det derfor som minimum forekomme systematisk. Der vil ikke kunne udledes nogen systematik mellem forekomsten af det mulige morfemparadigme 'v', 'm', 'l' og 's' sammen med andre tekststrengene end 'and'. Man kunne indvende at eksempelvis en tekststreng som 'agt' dog også ligger under for samme systematik: *lagt*, *sagt*, *magt*, *vagt*, men tekststrengen er ikke systematisk. Således må begrebet systematik defineres før proceduren kan fortsætte, og de systematisk forekomne morfemer må udledes.

Indtil nu har proceduren udledt en maksimal opdeling af hvert ord i 4 tekststrengene: $a: \text{'a-'}, b: \text{'-b'}, c: \text{'-c-'} \text{ og } x: \text{'x'}$. Her er x givet som ordstammen, a er eventuelt præfiks, b er eventuelle suffiks, og c er eventuelle infiks. $a-c$ er ikke opdelt i mindre morfemer.

Derfor er affikser defineret som en tekststreng der forekommer umiddelbart før, efter eller midt i et ord.

Det må stilles som kriterium for systematikken at den skal kunne afdække en forekomstrelation mellem et morfemparadigme (en mængde af morfemer) og en ordgruppe (en mængde af ord). Desuden kan det stilles som kriterium at ordgruppen skal være potentielt infinit – i praksis kunne man sige at man skal kunne registrere mere end eksempelvis 500 forekomster. Hvis man griber til den matematiske terminologi, kan man formalisere problematikken således at man leder efter en funktion hvor hvert input- x har n output- y -værdier, hvor n er lig mængden af elementer i morfemparadigmet. Denne funktion skal desuden gælde for en potentiel infinit mængde af x 'er (hvis man antager at et sprogs ord er potentielt infinit), hvilket i praksis kunne sige at der skal forekomme eksempelvis 500 x 'er. Funktionen skulle konkret udtrykke at

$$x_z + m_n \in Y, \text{ hvor } z, n \in N \text{ og } Z = N^7$$

Dvs. at hvis ordstammen adderes (dvs. sammenføjes) med hvert enkelt element m i morfemparadigmet M fås $z \cdot n$ forskellige resultater for hvilke det gælder at $\{y_1, \dots, y_{zn}\}$ forekommer som elementer i mængden Y . Den potentielt infinite mængde af y 'er, der er et resultat af funktionen, udgør elementerne i mængden Y -ordgruppe. På dette tidspunkt er det nødvendigt at nævne at mængdebegrebet anvendes en anelse anderledes end i matematikken. Når to elementer adderes gælder det ikke at $A + B = C$, men at $a + b \in C$, hvor '+' angiver sammenføjning. Hermed menes at elementet a sammenføjes med elementet b følgende reglerne for affikser. Den tekststreng der er resultatet heraf, tilhører mængden C (det vil i ovenstående eksempel sige Y). Endvidere gælder det at funktionen *skal* gælde for en infinit mængde x .

De ordgrupper der indtil videre er lokaliseret, er dog større end mængden af ordtyper i et sprog som dansk. Eksempelvis vil ordene *spiste* og *legede* tilhøre forskellige ordgrupper. Disse ordgrupper må reduceres til én ordgruppe for at simplificere den grammatiske beskrivelse (det der tidligere er benævnt som ordtype). Dette kan kun ske ud fra kommutation (som følge af ordtypedefinitionen). Denne reduktion må dog, for at undgå fejlanalyse, ske efter at den samlede analyse er fuldført

⁷ hvor x = ordstamme, m = element i morfemparadigme, n = antal elementer i morfemparadigmet, $N = \{1, 2, 3, \dots, \infty\}$, dvs. de naturlige tal større end nul, og x = mængden af ord i en ordgruppe X .

idet den må basere sig på en analyse af identitet mellem ordgruppernes forekomst i helheder i sætningen.

4.3.2 Helhedsparadigmatik

Procedurens næste trin er den paradigmatiske undersøgelse af hvordan ordgrupper kan udgøre en samlet og sammenhængende helhed i sætningen. Den syntagmatiske helhed skal som minimum opfylde Diderichsens parentesprincip (EDG: 146):

”Det grundlæggende Princip for Leddenes Stilling bestaar deri, at Ord, der hører sammen, følger umiddelbart efter hinanden og ikke kan adskilles af uvedkommende Led. Man kan altsaa som Regel sætte Parenteser om de Ord, der hører sammen og danner relativt selvstændige Enheder, og disse Parenteser griber normalt ikke ind over hinanden: (*En overordentlig spændende Oplevelse havde (det tapre Kompagni) (i den lille Bjergby Andermatt)*”).

Det første spørgsmål er så hvilke muligheder der er for at udlede en procedure der er i stand til at sætte disse parenteser. En sådan procedure ville dog som tidligere argumenteret kræve en syntaktisk analyse hvilket ikke ligger inden for den formalistiske analyses muligheder. Idet den ydre funktion (i Diderichsens terminologi) således ikke kan være til nogen nytte i denne sammenhæng, må det være den indre bygning af disse helheds paradigmer der skal undersøges og om muligt formaliseres.

Helhedsparadigmatikken skal med andre ord søge at udfinde en række paradigmer bestående af n på hinanden efterfølgende ordgrupper der under visse regler altid forekommer sammen. For at eksemplificere dette, kan ét (ufuldstændigt) paradigme være:

ADVERBIAL

A	PRÆPOSITION ←	NOMINAL
B	ADVERBIUM	

$v \rightarrow c$ angiver selektion (*specifikation*⁸)

Ovenstående angiver at det topologiske paradigme *adverbial* består af to delmængder, A og B . Delmængde A består af endnu to delmængder, nemlig af ordgruppen *PRÆPOSITION* og af ordgruppen *PRÆPOSITION* efterfulgt af helhedsparadigmet *NOMINAL*, mens delmængde B består af én delmængde, nemlig af ordgruppen *ADVERBIUM*. Dette kan formaliseres således:

⁸ Begrebet *specifikation*, som er en præcisering af selektionstypen, defineres i 4.3.3.

$$\left. \begin{array}{l} \{PRÆPOSITION\} \\ \{PRÆPOSITION + NOMINAL\} \\ \{ADVERBIUM\} \end{array} \right\} \subseteq ADVERBIAL$$

↑↑

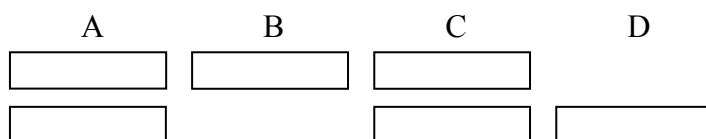
$$ADVERBIAL = \{PRÆPOSITION, PRÆPOSITION + NOMINAL, ADVERBIUM\}$$

Men den formale definition af den indre bygning mangler stadig, for hvad er det der ligger til grund for denne opdeling? Derfor må der udledes en procedure der kan udlede et skema svarende til det ovenfor.

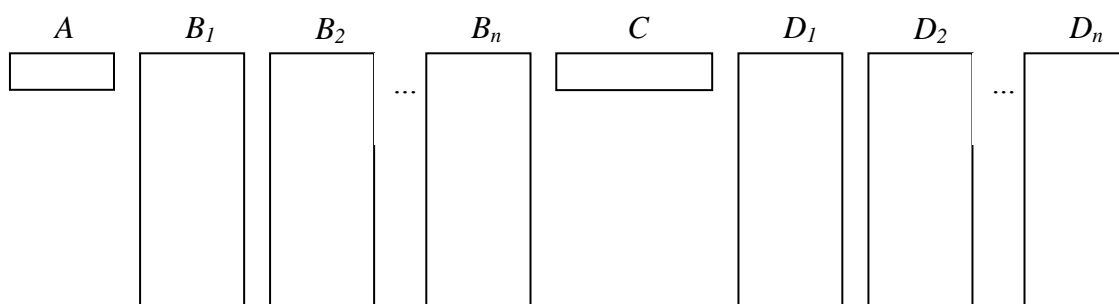
Hvis man tager udgangspunkt i Diderichsens definition og føjer dertil at for at der er tale om et selvstændigt syntagma skal det være tilfældet at syntagmet kan flyttes i sætningen. Det vil sige at i sætningen *jeg kender Peter*, er der tale om tre selvstændige syntagmer fordi leddene kan flyttes rundt: *Peter kender jeg*. Men det eneste givne er udtrykssubstansen, så hvis dette skal stilles som kriterium, må funktionen 'flytning' defineres. Denne funktion kan dog kun defineres hvis det der skal flyttes er defineret først, hvorfor der nu er tale om en ubrugelig tautologi: *Sætningsled x er selvstændigt sætningsled, hvis det kan flyttes, led x kan flyttes hvis det er selvstændigt sætningsled*. En måde at definere flytningen på er i forlængelse af begrebet *forekomst*. Hvis flytningsbegrebet omformuleres i overensstemmelse med de forekomstregler der er skitseret, bør følgende gælde: Tekststrengen *x* er et selvstændigt sætningsled hvis og kun hvis (1) de enkelte ord i *x* følger umiddelbart efter hinanden, (2) hvis *ABx* forekommer, forekommer også enten *AxB*, *xAB*, *BxA*, *xBA* i en potentiel infinit mængde af varianter, og (3) *A* og/eller *B* kan ikke indsættes i og dermed forekomme som del af *x*.

Regel nummer (3) angiver at hvis eksempelvis *ABx* og *BxA* forekommer systematisk, men aldrig eksempelvis *AxB*, så må det antages at *B* \in *X*. Det ligger også som grundlag for at proceduren skal have som kriterium at den skal afsløre den størst mulige længde af *X*, dvs. det størst mulige antal elementer, *x*, i *X*, som kan flyttes samlet.

Proceduren kan illustreres således:



Givet er sætningskomponenterne (ordene som ordgrupper) A , B , C og D . Hvert segment ovenfor er defineret som ét ord fra én ordgruppe, hvor $A \neq C$ og $A \neq B$ (og som logisk følge: $B \neq C$). Videre er det givet at *hvis* øverste række forekommer, dvs. hvis ABC forekommer umiddelbart efter hinanden, og *hvis* nederste række forekommer, dvs. hvis ACD forekommer umiddelbart efter hinanden, og *hvis* $B = D$, så er $B = D$, A og C selvstændige led i sætningen. På dette trin af proceduren inventarieres $B = D$ som x_0 . Det gælder for x_0 at der er tale om en variabel, og hvor ethvert ord der opfylder kriterier for at blive inventarieret som x_0 , bliver inventarieret som element i mængden X . Efter X er færdiginventarieret, dvs. der forefindes ikke flere x_0 end de allerede inventarierede, vil proceduren skulle udvides til segmenter defineret som a ord fra b ordgrupper, hvor $a \geq b$. Proceduren udvides da som illustreret:



Målet bliver således at inventariere det størst mulige antal ord der efterfølger hinanden, og som kan undergå 'flytningen' fra B til D , mens A og C er konstante.

Dette vil give en liste lignende den ovenstående hvor det er givet at $ADVERBIAL = \{PRÆPOSITION, ADVERBIUM, PRÆPOSITION+NOMINAL\}$. På denne vis ville muligvis hvert enkelt syntagma kunne opstilles.

På baggrund af denne analyse vil der nu kunne foretages en reduktion af ordgrupperne til ordtyper. Ordgrupperne ville være inventarierede på baggrund af hver enkelt variant således at eksempelvis *billet-ten* ville være inventarieret i en anden ordgruppe end *legoklods-en* fordi bestemtheds morfemet i det ene tilfælde er betegnet *-ten* (i den formalistiske analyse), mens bestemtheds morfemet i den anden analyse er betegnet *-en*.

Reglen for reduktion vil nu være at hvis alle de led i alle syntagmer hvori en ordgruppe indgår er identiske bortset fra to ordgrupper der opfører sig identisk mht. placering, så er disse to ordgrupper identiske og kan reduceres til én ordtype.

4.3.3 Morfemsyntagmatik

Morfemsyntagmatikken kan med inspiration fra Hjelmlev inddeles i 3 typer som kan kaldes *konstellativ syntagmatik*, *determinanssyntagmatik* og *interdependenssyntagmatik*. Når det teoretiske omdrejningspunkt ikke er generelt, men knyttes specielt an til syntagmer, benytter Hjelmlev 3 parallelbegreber til *konstellation*, *determination* og *interdependens*, nemlig *kombination*, *selektion* og *solidaritet*. Således vil jeg også anvende disse tre begreber frem for de andre, selv om de formalt set har samme definition.

Kombination symboliserer Hjelmlev ved $v-v$, selektion symboliserer Hjelmlev ved $v \rightarrow c$, mens han symboliserer solidaritet som $c \sim c$ ⁹. Man kan hævde at der eksisterer en morfemsyntagmatik mellem leddene i den paradigmatiske topologi, men denne er afdækket i proceduren for den syntagmatiske paradigmatiske topologi hvorfor den under alle omstændigheder er uinteressant i denne sammenhæng. Spørgsmålet her er hvorvidt morfemer i nogle led forudsætter tilstedeværelsen af morfemer andre led. De tre hjelmlevske muligheder ligger således op til tre forskellige typer af afhængighed. Disse tre typer kan endvidere hver især deles i to afhængighedstyper: afhængighed inden for et paradigme og afhængighed uden for et paradigme. Betegnelserne *inden for* og *uden for* kræver lidt præcisering. Definitionen angiver at (1) der foreligger et parvist afhængighedsforhold mellem alle i sætningen forekomne helheder (helhedssyntagmatik), hvilket vil sige (2) det kræves at der er mindst to syntagmer til stede for at der kan være tale om et afhængighedsforhold, og (3) afhængighedsforholdet kan da bestå i (a) at de to syntagmer tilhører samme paratagme, eller (b) at de to syntagmer tilhører hvert sit paratagme. Følgende konkretiserer Hjelmlevs tre afhængighedstyper parret med afhængighed (a) og (b) ovenfor.

⁹ c : konstant, v : variabel

Notation

Følgende notation følges: i $A(x)$ angiver A et helhedsparadigme, mens x angiver et led (dvs. en helhed) i paradigmet. Det gælder således at $x \in A$. Til notationen er der følgende bemærkninger: Tegnet for logisk negation (\neg) kan forekomme ved enten A eller x . Ved $\neg A$ angives at der forekommer et paradigme som ikke er lig A , og ved $\neg x$ angives at der forekommer et element i paradigmet A hhv. $\neg A$ som ikke er lig x . Det følger matematisk heraf at $A(x) \phi \neg A(x)$ (ϕ er tegn for en vilkårlig funktion (jf. Uldall, 1957: 36)) er en ulovlig funktion idet venstresiden er defineret som 'elementet x forekommer i paradigmet A ', mens højresiden er defineret som 'elementet x forekommer i paradigmet $\neg A$ '.

Leksikalsk afhængighed

Den leksikalske afhængighed er kendetegnet ved kombination. Kombinationen kan karakteriseres uformalt som 'ingen afhængighed', men i Hjelmsslevs terminologi er der alligevel formalt set tale om en afhængighed selv om den er defineret mellem to uafhængige variable. Den leksikalske afhængighed kan forekomme som enten:

$$A(x) \text{ — } \neg A(y)$$

$$A(x) \text{ — } A(\neg x)$$

Det vil med andre ord sige (1) at hvis elementet x forekommer fra paradigmet A , og hvis samtidig elementet y forekommer fra et hvilket som helst andet paradigme end A , så er afhængigheden mellem de to elementer defineret ved at begge variable kan forekomme uafhængigt af hinanden. (2) at hvis elementet x forekommer fra paradigmet A , og hvis samtidig et hvert andet element forekommer fra paradigmet A , så er afhængigheden mellem de to elementer defineret ved at begge variable kan forekomme uafhængigt af hinanden. Når ovenstående regler her benævnes *leksikalsk afhængighed* skyldes det at der er tale om en mængde forekomster der ingen betydning har for grammatikken i forhold til den mulige forekomst af noget som helst andet i sætningen.

Grammatisk afhængighed

Den grammatiske afhængighed er kendetegnet ved solidaritetsfunktionen. Den grammatiske afhængighed deles ligesom den leksikalske op i to:

$$A(x) \sim \neg(B(y))$$

$$A(x) \sim A(\neg x)$$

Heraf fås (1) at hvis der forekommer et x fra paradigmet A , så kan der ikke forekomme et paradigme med elementet y . (2) at hvis der forekommer et x fra paradigmet A , så kan et hvilket som helst andet element fra paradigmet A ikke forekomme i sætningen.

Den grammatiske afhængighed kan eksemplificeres ved kasus. Hvis der forekommer et nominal i akkusativ, kan der ikke forekomme andre nominaler i akkusativ fra enten samme paradigme og/eller fra et andet paradigme (groft sagt – i eksempelvis både latin og græsk kan nogle verber have flere objekter i akkusativ).

Morfologisk afhængighed

Man kan sige at der egentlig eksisterer to væsensforskellige former for grammatiske afhængigheder hvoraf afhængighederne ovenfor udgør den ene gruppe. Den første gruppe (ovenfor) er kendetegnet ved at forekomsten af noget udelukker forekomsten af noget andet. Derimod kan forekomsten af noget også forudsætte forekomsten af noget andet. For at skelne mellem disse to er den ene her benævnt *grammatisk* afhængighed, og den anden (nærværende) *morfologisk* afhængighed. Ordet *grammatisk* refererer her til en overordnet afhængighed, fx mellem kasus, der ofte vil være kædet sammen med noget af enten semantisk eller syntaktisk karakter – alt efter hvordan man ser det. Tilfældet med at der kun kan forekomme ét led i akkusativ i en sætning på nogle sprog, hænger sammen med at akkusativ markerer (direkte-/akkusativ-)objekt, og idet der ikke kan forekomme to direkte objekter i en sætning (hvilket kan begrundes syntaktisk/semantisk), træder den grammatiske afhængighed i kraft.

Den afhængighed der skal behandles her, den morfologiske afhængighed, er derimod kendetegnet ved ikke at have nogen tilknytning til et semantisk plan, men udelukkende at vedrøre sprogsystemet i en hjelmslevsk forstand. Den morfologiske afhængighed er kendetegnet ved selektion:

$$A(x) \leftarrow \neg A(y)$$

Det vil sige at to elementer fra to forskellige paradigmer indgår en funktion hvor det ene element selekterer det andet. Et eksempel på dansk er verbalformernes morfemsyntagmatik. Her gælder bl.a. følgende (lidt grove) regler:

Modalverbum \leftarrow Infinitiv

Modalverbum \leftarrow Hjælpeverbum i Infinitiv \leftarrow Perfektum Participium

Modalverbum \leftarrow Hjælpeverbum i Infinitiv \leftarrow Perf.ptc. af *blive* \leftarrow Perf.Ptc.

Modalverbum ← Verbum i passiv

m.v.

4.3.4 Helhedssyntagmatik

Udgangspunktet i nærværende undersøgelse har været at det ikke kan lade sig gøre at opstille en procedure der kan forudsige sætningers helhedssyntagmatik. Derfor vil jeg ikke inkludere mig på en yderligere diskussion af problematikken her, men lade spørgsmålet stå åbent. Proceduren afsluttes således her, og videre vil den blive behandlet i et noget mere kritisk lys.

5.0 Evaluering af proceduren

Indtil videre har det primære omdrejningspunkt været at fremstille en procedure der havde mulighed for alene givet udtrykssubstansen (en potentielt infinit tekstmængde) at konstruere en grammatik for et givent sprog. Opgaven er blevet indsnævret i flere omgange og er endt med at være et spørgsmål om hvorvidt det er muligt at konstruere sætningsled ud fra en foreliggende fonemanalyse eller en elektronisk bogstavskrift. Fokus har alene været på fremstillingen og udviklingen af den mulige procedure, men en væsentlig del af en undersøgelse af denne karakter, dvs. en konkret fremstilling af en procedure, må også være en vurdering af dens anvendelighed, altså en kritisk analyse af proceduren. I forbindelse med udarbejdningen af proceduren, findes følgende kritikpunkter nævneværdige:

- i) Det må først og fremmest tages som forbehold at forslaget til en procedure kun er en skitse. For at proceduren skulle være anvendelig ville det dels kræve en helt anden detaljeringsgrad, dels en større grad af formalisering og dels en prioriteret rækkefølge hvori de enkelte procedurer skulle anvendes. Først da ville der forelægges en egentlig algoritme der så kunne afprøves med henblik på falsificering.
- ii) Det ligger som en nødvendig forudsætning for proceduren at tekstinputtet er homogent. I samfund hvor der ikke eksisterer nogen formel retskrivning, vil proceduren vanskeliggøres yderligere, men mere alvorligt er muligvis dels stave- og skrivefejl, og dels lyrisk sprogbrug. Hvis proceduren indarbejdes i en computer, og hvis denne bliver præsenteret for en mængde dansk skriftsprog, herunder eksempelvis en større mængde poesi af en forfatter som Højholt, vil der opstå større og muligvis uovervindelige vanskeligheder. Der vil således være behov for en matematisk model der er i stand til at

levere kvalificerede gæt og derved frasortere tilfælde hvor der er sandsynlighed for fejl.

- iii) Der er argumenteret for at en forhåndenværende foneminventariering vil være nødvendig. Denne må da desuden indeholde nogle kriterier for ordafgrænsning. I proceduren har det været forudsat at denne inventariering kunne erstattes af et skriftsprog. Her må dog tilføjes tre problematikker:
- a. Der må som præmis også stilles en præcisering af hvilket tegn der anvendes til at afgrænse ord. I nutidige vestlige sprog hersker der ingen problemer idet mellemrummet (blanktegn) altid anvendes som ordafgrænsning, men dette er ikke tilfældet for alle sprog. I gotisk adskiltes ord ikke, mens andre skriftsprog har brugt et særligt tegn til at adskille ordene med, således fx i runeindskriftene og i persisk kileskrift (cf. Byskov, 1923: 16).
 - b. Skriftsproget er blevet anvendt som erstatning for en fonemskrift for at have en eksisterende sprogs substans (modsat fonemskrift), men for at undgå de samme problemer som fonemanalysen møder, må det med nødvendighed være en fuldstændig regulær skrift, dvs. en maskinskrift. Hermed kan der stadig opstå problemer i sprog hvor skriften ikke er fuldstændig regulær som i arabisk hvor vokalerne dels valgfrit kan udelades, og hvor bogstaverne kædes sammen og derfor ændrer udseende efter deres omgivelser. I mange vestlige sprog opstår der her problemer af mindre alvorlig karakter i form af anvendelsen af versaler.
 - c. Ved brug af forskellige skrifttyper, fed, kursiv, understregning, m.v. vil der opstå problemer. Hertil kommer at sprogs substansen skal stamme fra en homogen tidsperiode og fra en homogen dialekt.
- iv) Proceduren vil ikke have noget forhold til nogen indholdskategori. To helt identiske ord (absolutte synonymer eller allomorfer) vil derfor konsekvent blive regnet som værende forskellige. Mere generelt kan det siges at proceduren ikke vil kunne inventariere invarianter, men kun vil kunne inventariere varianter. Dette skaber et teoretisk, sprogvidenskabeligt problem. Det må desuden ud fra undersøgelsen konkluderes at dette problem

ikke kan løses. Hvis det kunne lade sig gøre at inventariere de enkelte helheders forekomst i sætningen, ville der potentielt set være mulighed for at gruppere morfemer med fuldstændig samme distribution som allomorfer, men som der er argumenteret for, antages der her at dette ikke er en mulighed så længe betydningsniveauet ikke kan inddrages idet fx PRÆTERITUM og PRÆSENS har nøjagtigt samme distributionsmuligheder i sætningen. Inddrages elliptiske konstruktioner, åbnes dog en mulighed idet det her tydeliggøres at de to former kan grupperes sammen, fx: *jeg synes at du kommer og går hele tiden*, men ikke: **jeg synes at du kommer og gik hele tiden*. Dette ville dog kræve en yderligere undersøgelse – og inddragelse af ellipser vil lukke op for nye fejlmuligheder.

- v) Det vil ikke ud fra den beskrevne analyse være muligt at inventariere nominalhelheder som et helhedsparadigme. Det fastholdes som en grundregel at elementer der kan flyttes i sætningen udgør et og kun et helhedsparadigme. Den største successivt forekomne ordmængde der kan flyttes i dansk, er PRÆPOSITION+NOMINAL. Dermed udfindes kun hvad man kunne kalde adverbialhelheder, dvs. helheder i sætningen der har adverbial funktion, men ikke nominalhelheder. For at nominalerne kan identificeres, vil det kræve en mulighed for at identificere adpositioner.
- vi) Et stort problem der hidtil er ukommenteret, opstår ved forskellige tituleringer. Fx *lektor Blomme*. Disse er de eneste tilfælde hvor to substantiver kan forekomme umiddelbart efter hinanden. Hidtil har problemet med en nominalform der forekommer som adverbial kun været berørt på sætningsniveau, men i eksempler som *en tolv alen bred vaage* og *de flere alen tykke mure* (fra Byskov § 171). En yderligere problematik her er i forhold til artsbetegnelser, fx *en flaske vin* o.lign. hvor flere substantiver står efter hinanden og kan flyttes i sætningen.
- vii) Proceduren har et tydeligt afsæt i det danske sprog, og den vil forventeligt blive problematiseret i en yderligere grad hvis den skulle afprøves over for andre sprogtyper. Her vil det dels være et problem med sprog hvor morfologien angiver den syntaktiske funktion (fx latin, finsk og græsk), hvor

der er en højere grad af agglutinerings (fx grønlandsk), og hvor der er en mindre grad af agglutinerings (fx vietnamesisk). Selvom proceduren således ideelt set skulle være et udkast til en universel procedure (det ville være en nødvendighed hvis den skulle kunne leve op til kriteriet om ikke at have forudgående viden om sproget), er dette ingenlunde tilfældet. Overvejelserne er overfladiske og vedrører stort set kun sprog lignende det danske.

- viii) Proceduren søger at udnytte at grammatisk morfemers forekomst er systematiske, og kræver som præmis for systematik at denne forekommer i en potentielt infinit mængde ord. Dette udelukker derfor også en beskrivelse af mindre ordklasser som pronominer.

Som følge af ovenstående indvendinger må det derfor konkluderes at det ikke i nærværende undersøgelse har været muligt at opstille en beskrivelse af sætningsledene ud fra et rent formalistisk grundlag. Det har dog været muligt at udfinde en række præmisser der, hvis de opfyldes, vil muliggøre en sådan procedure, samt at skitsere nogle tanker der vha. en udbygning kunne hjælpe til at konkretisere mulighederne såvel som umulighederne yderligere.

6.0 Konklusion

Efter at have forsøgt at indskrænke nærværende arbejdsområde til det der kunne anses for det teoretisk mulige, må det konstateres at det inden for nærværende rammer ikke har været en mulighed ud fra den rent formalistiske tilgang at opstille en procedure der kan beskrive og konstruere sætningsled på baggrund af en potentielt uendelig fonemstreng.

Det der kan konstateres at være belæg for, i det mindste teoretisk, at argumentere for muligheden af, er:

- (1) at opstille en procedure der for dansk i et vist omfang kan analysere adverbialers indre bygning, dvs. ords forekomst i adverbialhelheder
- (2) muligvis at opstille nogle afhængighedsforhold for dansk mellem forekomsten af verbalformer
- (3) muligvis at opstille nogle få afhængighedsforhold mellem kasusforekomst

Dette givet at følgende præmisser som minimum er opfyldt:

- (1) sproget er dansk
- (2) sproget er fejlfrit (dvs. anvender en homogen retskrivning)
- (3) sprogmaterialer der anvendes som input, skal stamme fra samme tid og fra samme dialekt
- (4) udtrykssubstansen er produceret ved en elektronisk maskinskrift
- (5) varianter behandles som værende invarianter
- (6) der gives på forhånd regler for ordafgrænsning, herunder specialtegns funktion
- (7) der gives en potentielt infinit mængde sætninger der alle lever op til præmisserne

- (8) der gives en procedure der kan behandle uoverensstemmelser og uregelmæssigheder
- (9) den beskrevne procedure konkretiseres, formaliseres og reglerne opstilles i en prioriteret rækkefølge hvori de skal anvendes

Problemerne for proceduren består primært i vanskeligheden med at beskrive helsætningsniveauet. Dog er der andre mulige elementer som kunne indlejres i proceduren for at effektivisere den. Eksempelvis kan adverbialer enten forekomme med præposition, med postposition eller som kasusbøjning. Man kunne derfor forestille sig at indlægge en specifik underprocedure der skal undersøge forekomsten af adpositioner. Herved kunne det muliggøres at udfinde nominalerne særskilt. Man kunne dog på den anden side forestille sig at problemet bliver endnu større ved at skelne mellem nominalhelheder i eksempelvis nominativ og nominalhelheder i ablativ.

Undersøgelsen har således både resulteret i en række negative og i en række positive resultater på trods af at de positive resultater dog ikke har været udsat for nogen evalueringsprocedure, men kun har været genstand for teoretiske overvejelser. De negative resultater angiver en række restriktioner for arbejdet som det er skitseret. De negative resultater er med til at pege i retning af at en del af sprogbygningen er påvirket af semantiske eller pragmatiske vilkår, således eksempelvis ledsætningernes funktion såvel som de enkelte sætningsleds funktion i sætningen. De positive resultater er medvirkende til at pege på at en stor del af sproget stadig er særdeles systematisk – og det kan endvidere hævdes at den i formalistisk forstand mest systematiske del af sproget er de grammatiske morfemer der muligvis vil kunne udledes rent formalistisk.

7.0 Litteraturliste

Ayer, 1936: Alfred J. Ayer: *Language, Truth and Logic*. 1946 (førsteudgave 1936). Penguin Books.

Bloomfield, 1933: Leonard Bloomfield: *Language*. 1961 (førsteudgave 1933). Holt, Rinehart and Winston. US.

Brøndal, 1932: Viggo Brøndal: *Morfologi og Syntax*. 1932. Bianco Lunos bogtrykkeri A/S. København.

Byskov, 1923: J. Byskov: *Dansk Sproglære*. 3. udgave, 1923. Det Schønbergske Forlag. København.

Chomsky, 1957: Noam Chomsky: *Syntactic Structures*. 1957, nærværende udgave er et fotografisk optryk fra 2002 med introduktion af Lightfoot. Mouton de gruyter. Berlin.

Chomsky, 1975: Noam Chomsky: *The Logical Structure of Linguistic Theory*. 1975. Plenum Press, New York. US.

Chomsky, 1995: Noam Chomsky: *The Minimalist Program*. 1995. The MIT Press. Cambridge, Massachusetts.

Comrie, 1981:

de Saussure, 1916: Ferdinand de Saussure: *Course in General Linguistics*. Charles Bally og Albert Sechehaye (red.). Førsteudgave 1916, nærværende udgave: 1966. McGraw-Hill Paperback Edition. US.

Diderichsen, 1946: Diderichsen, Paul: *Elementær Dansk Grammatik*. 3. udgave, 7. oplag 1976, førsteudgave 1946. Gyldendal, København.

Dik, 1989: Simon C. Dik: *The Theory of Functional Grammar. Part 1: The Structure of the Clause*. 1989. Nærværende udgave er second, revised edition, 1997. Redigeret af Kees Hengeveld. Mouton de Gruyter, Berlin.

Harris, 1951: Zellig Harris: Structural Linguistics (tidligere: *Methods in Structural Linguistics* (1947)). 1974 (førsteudgave 1951). The University of Chicago Press, Chicago & London.

Hjelmslev, 1943: Louis Hjelmslev: *Omkring Sprogteoriens Grundlæggelse*. 1993 (førsteudgave 1943, forfatterens engelske udgave udkom 1961). Travaux du Cercle Linguistique de Copenhague, Vol. 25. Reitzels Boghandel A/S, København.

Hjelmslev, 1975: Louis Hjelmslev: *Résumé of a Theory of Language*. 1975. Travaux du Cercle Linguistique de Copenhague, Vol. 16. Nordisk Sprog- og Kulturforlag, København.

Hjelmslev, 1972: Louis Hjelmslev: *Sprogssystem og sprogforandring*. 1972. Travaux du Cercle Linguistique de Copenhague, Vol. 15. Nordisk Sprog- og Kulturforlag, København.

Jespersen, 1913: Otto Jespersen: *Sprogets Logik*. 1913. J. H. Schultz A/S. København.

Jørgensen, 2000: Henrik Jørgensen: Studien zur Morphologie und Syntax der festlandskandinavischen Personalpronomina. Mit besonderer berücksichtigung des Dänischen. 2000. Aarhus University Press, Aarhus.

McGregor, 2002: William McGregor: "Ergative and accusative patterning in Warrwa". 2002. I Kristin Davidse (red.): *The Nominative & Accusative and their counterparts*. John Benjamins Publishing Company. Philadelphia, PA, USA.

Ng, 1999: Tze Beng Ng: *Calculus. An introduction*. 1999. Springer. Berlin.

Prebensen, 2003: Henrik Prebensen: "Indledning" i Peter Juel Henriksen og Henrik Prebensen (red.): *Sprog og Matematik*. 2003. Handelshøjskolens forlag. København.

Rehling, 1932: Erik Rehling: *Det danske sprog. Fremstilling for Lærere og Seminarier*. 1932. J. H. Schultz Forlag. København.

Shannon & Weaver, 1949: Claude E. Shannon og Warren Weaver: *The Mathematical Theory of Communication*. 12. oplag, 1971 (førsteudgave, 1949). University of Illinois Press. Chicago, US.

Soros, 1998: Georg Soros: *Den globale kapitalismes krise. Det åbne samfund i fare*. 1998, nærværende danske oversættelse: 1. udg., 1. oplag 1999. Rosinante Forlag A/S. København.

Togeby, 1993: Ole Togeby: *PRAXT. Pragmatisk tekstteori I*. 1993. Aarhus Universitetsforlag. Århus.

Trubetzkoy, 1958: N. S. Trubetzkoy: *Grundzüge der Phonologie*. 5. oplag 1971 (førsteudgave 1958). Vandenhoeck & Ruprecht, Tyskland.

Uldall, 1957: H. J. Uldall: *Outline of Glossematics*. 1957. Travaux du Cercle Linguistique de Copenhague, Vol. 10₁. Nordisk Sprog- og Kulturforlag. København.

Wiwel, 1901: H. G. Wiwel: *Synspunkter for dansk sproglære*. 1901. Det Nordiske Forlag, København.

ⁱ Udtrykt i forlængelse af Shannons matematiske kommunikationsteori kan man hævde at de grammatiske morfemer bærer en mindre information end de leksikalske. Shannon forstår begrebet information formalistisk som et statistisk mål for en talers frie valg af tekst i produktionen af en sætning (Shannon & Weaver, 1949: 9).

ⁱⁱ Blandt sociologer er specielt Durkheims indflydelse diskuteret. Frans Gregersen (1991: 106) argumenterer for at de Saussure ikke har været inspireret af Durkheim pga. kronologien, specielt i deres udgivelser. Durkheims *De la division du travail social* er fra 1893, *Les regles de la méthode sociologique* fra 1895, mens hans mest berømte undersøgelse, *Le suicide* er fra 1897. De Saussures *Cours* er som bekendt udgivet 1916 på baggrund af forelæsningsnoter fra tre kurser i almen lingvistik i perioden 1907-11 (cf. desuden Gregersen, 1991: 141). Uanset om de Saussure har været direkte inspireret af Durkheims tanker eller ej, synes den metodiske såvel som den teoretiske lighed at være slående mellem de to. Dette kan så enten tages som udtryk for at de Saussure har været direkte inspireret af Durkheim, eller det kan tages som udtryk for at de Saussure har været inspireret af et generelt samtidigt videnskabsyn, der så igen kan spores blandt mange andre hos Hjelmlev. Tak til Jørgen Holmgaard for at have påpeget denne mulige relation.

ⁱⁱⁱ Jf. Michael Rasmussen (1992: 170f.) fremlægning. Rasmussen citerer Hjelmlevs tekst "Et sprogteoretisk causeri" fra 1941: "Kravet om de færrest mulige og helst ingen indefinable og aksiomer er selvsagt ikke absolut: det vilde være en meningsløshed. Et hierarki af definitioner ender altid et sted, det vil sige: det ender uvægerlig i nogle indefinable, om ikke af anden grund, saa fordi det ellers måtte blive uendeligt ... Kravet kan kun gaa ud paa, at man inden for det fagomraade, man behandler, gaar saa vidt muligt i denne retning, og helst saa vidt, at de indefinable og de aksiomer, der til sidst bliver tilbage, er af en saadan art, at de naturligt hører hjemme i den rene erkendelsesteori, altsaa er af fællesvidenskabelig og ikke fagvidenskabelig karakter, at de er *almene* og ikke *specifikke*".

^{iv} Byskov (1923) synes næsten bevidst at undgå at tale om mennesker og brugen af sproget. I indledningen skriver han: "En dansk sproglære har til opgave at beskrive det danske Sprog. Dette foreligger dels som et Talesprog, dels som et Skriftsprog". Byskov omtaler sproget som noget foreliggende, og sprogbrugeren omtales hverken i lydlæren eller i verselæren hvor det kunne forventes. I lydlæren skriver han: "Sproglydene dannes ved, at en Luftstrøm på sin Vej fra Lunerne bliver sat i

Svingninger ved Hjælp af Taleorganerne. Disse svingninger meddeler sig til den Luft, der er udenfor den talendes Mund, og videre fra Luftdel til Naboluftdel”. Byskov anvender først en passiv-form med *sproglydene* som subjekt hvilket giver afsæt for at svingningerne i næste sætning bliver subjekt.

Erik Rehling (1932) omtaler derimod sprogbrugen: ”Grammatikken er Sproglærens almindelige Del; den skal undersøge de almindelige Love og Regler for Sprogbrugen. Den specielle Sproglære, Oplysninger om de enkelte Ord og deres Betydninger, maa søges i Ordbøgerne.” Han taler endog om relationer mellem tanke og sprog: ”Mens det saa nogenlunde er muligt at udtrykke en og samme Tanke paa en Række forskellige Sprog, er de Midler, hvorved den udtrykkes, ingenlunde de samme”, men sprogbrugeren der forbinder tanke og sætning undlades nævnt, som hos Byskov, og i stedet er de skjult i en passivkonstruktion (”de Midler, hvorved den udtrykkes” frem for *de midler hvorved taleren udtrykker den*). Rehling taler også i højere grad om logiske forbindelser mellem ordene, og hos Jespersen (fx *Sprogets Logik* (1913), men også *Philosophy of Grammar* og *Analytic Syntax*) argumenteres der for at ordenes indbyrdes logiske forhold (semantik) tilmed i vidt omfang er bestemmende for deres syntaktiske funktion, men stadig er sprogbrugeren unævnt: ”formålet for denne afhandling er nået, hvis det på nogle enkelte indbyrdes nær beslægtede områder er lykkedes mig at vise, at der bagved sprogenes uendelige afvexlinger i kasus- og andre former, i ordstillingslove og syntaktiske regler, dog ligger visse simple logiske grundforhold” (1913: 90).

Hos Wiwel (1901) findes også en modstilling mellem indhold og udtryk således at ”Ved ”sprog” forstår man et meddelelsesmiddel for vore tanker”, men med de karakteristiske passivkonstruktioner undgås endnu engang sprogbrugeren produktion af sætninger: ”Man har fingersprog, øjensprog, ja blomstersprog o. s. v., men de to fuldkomneste slags sprog er dog talesproget og skriftsproget. Talesproget (hvor man til tegn for tanker bruger de mangehånde for øresansen opfattede lydbølger, som den for åndedrættet nødvendige luft kan danne på sin vej igennem strubehovedet, mund og næse) er en naturlig virksomhed, der findes hos alle folkeslag; skriftsproget (hvor tankerne knyttes til synlige tegn) er en kulturopfindelse” (1901: 1). Wiwel er dog inde på at ”man til tegn for tanker bruger ...”, men fokus er stadig i høj grad på sproget som selvstændigt system.

^v Givet at forudsætningen for proceduren kan imødekommes, nemlig den at ethvert sprogs grammatik kan beskrives ved syntaktiske funktioner – og ikke eksempelvis ved semantiske roller. Til denne problematik skriver Henrik Jørgensen (2000: 53 f.): ”Auch dieses Axiom [at intet sprog mangler kasus] ist neuerdings durch die Untersuchungen der südostasiatischen Sprache Mlabri von Jørgen Rischels angefochtet worden. In dieser Sprache sind – so die Ergebnisse Rischels – weder Reihenfolgesyntax noch Flexion deutlich vorhanden, obwohl eine vage SVO-Syntax nachgewiesen werden konnte ... Um den Fall Mlabri und ähnliche Fälle zu berücksichtigen, muss man anerkennen dass das Phänomen Kasus sich noch auf andere Weisen ausdrücken kann”.

^{vi} Dette er eksempelvis tilfældet i Ngs *Calculus* (Ng, 1999). Der er, if. Prebensen, 2003: 11, tale om et aldrende matematiksyn med inspiration fra Georg Cantors mængdelære, stående i kontrast til Russell typeteori.

^{vii} Der synes dog at være en problematik i Chomskys argumentation. Chomsky argumenterer ud fra et sprog hvis alfabet udgøres af a og b , og hvis grammatik udgøres af at alle sætninger der består af n forekomster af a , efterfølges af n forekomster af b , og kun disse, og hvor alle sætninger begynder med a (Chomsky, 1957: 21). Som eksempel på sætninger i sproget angiver Chomsky følgende: ” ab , $aabb$, $aaabbb$, ...”. Men dermed kan dette ab -sprog ikke være et sprog. Hvis det er en definerende egenskab for sprog at det består af en mængde af sætninger, som hver især er finite i længde (ibid. s. 11) og ab -’sproget’ indeholder sætninger der er infinite i længde, kan der ikke være tale om et sprog i denne forstand. Dette får også konsekvenser for Chomskys bevisførelse for at engelsk ikke er et ”finite state language” (ibid. s. 21 ff.), idet Chomsky hævder at der eksisterer absolutte rekursive funktioner i engelsk (ibid. s. 22). Denne problematik ser jeg i det følgende bort fra, da den er gældende for både kunstsprog som matematik og naturlige sprog som dansk og engelsk.