

Reduktion i udtalen af grammatiske vs. leksikalske ord

Line Burholt Kristensen, Ruben Schachtenhaufen & Kasper Boye
Københavns Universitet

1. Indledning

Dansk *har* kan distinkt udtales [hɑ:[?]]. Imidlertid kan ordformen i sammenhængende tale udtales med forskellige grader af reduktion, fx som [hɑ], [ɑ], [ɐ] eller være helt bortreduceret uden udtale. Her argumenterer vi for at udtalen ikke kun afhænger af fonologiske processer, men også påvirkes af om ordformen repræsenterer et grammatisk eller et leksikalsk ord.

Der er både et grammatisk og et leksikalsk *har*. Fuldverbet *har* i *Liva har en hund* er leksikalsk. Hjælpeverbet *har* i *Liva har spist en hund* er grammatisk. Vi undersøger her om der er sammenhæng mellem et ords status som grammatisk vs. leksikalsk og hvor reduceret det udtales.

I en tidligere korpusundersøgelse af amerikansk engelsk har Bell et al. (2009) sammenlignet udtalen af grammatiske og leksikalske ord, men brugt betegnelserne funktionsord og indholdsord. Ved en sammenligning af de to grupper fandt de mere reduktion i udtalen af indholdsordene. Forskelle i udtale af funktionsord og indholdsord har bl.a. været brugt som argument for at de to typer af ord tilgås forskelligt under sprogproduktion (Garrett 1980; Bell et al. 2009). Forskelle i udtale stemmer også godt overens med grammatikaliseringsforskningens ide om at udviklingen fra leksikalsk til grammatisk ofte ledsages af fonologisk reduktion (Lehmann 2002:110-111).

Det er dog ikke uproblematisk at sammenligne en gruppe forskellige grammatiske ord med en gruppe forskellige leksikalske ord. Eksempelvis definerede Bell et al. (2009) indholdsord som sub-

stantiver, verber, adverbier og adjektiver, mens funktionsordene var ”discourse markers, conjunctions, existential *there*, pronouns and other proforms, prepositions, articles, quantifiers, demonstratives, verb auxiliaries, and verb particles” (Bell et al. 2009:97). Schachtenhaufen (2013:98) finder ligeledes at åbne ordklasser, som substantiver og adjektiver reduceres væsentligt mindre end lukkede ordklasser, som præpositioner, konjunktioner, pronomener og artikler. Men når man på denne måde sammenligner en gruppe af grammatiske ord som *og* og *dersom* med leksikalske ord som *lossepladsen* og *blå*, støder man på tre grundlæggende problemer: 1) grammatiske og leksikalske ord har forskellig frekvens og forudsigelighed 2) grammatiske og leksikalske ord har forskelligt reduktionspotentiale, og 3) grundlaget for sondringen mellem grammatiske og leksikalske ord er uklart. Vi gennemgår de tre problemer et ad gangen.

For det første er grammatiske ord samlet set mere frekvente og mere forudsigelige end leksikalske ord. Pga. denne korrelation kan man ikke vide om reduktionen af et ord som *og* skyldes ordets grammatiske status, ordets høje frekvens i korpusset eller ordets forudsigelighed. Bell et al. (2009) prøver at tage højde for forudsigelighed og frekvens i deres statistiske analyser. Men selv avancerede regressionsanalyser kommer til kort når inputtet til analysen i forvejen er skævvredent, dvs. når stort set alle funktionsord er markant mere forudsigelige og frekvente end indholdsordene.

For det andet må de to grupper af ord formodes at have forskellig fonologisk sammensætning og dermed forskelligt reduktionspotentiale. I dansk gælder det fx at /ə/ reduceres oftere end fuldvokaler, at approksimanter reduceres oftere end obstruenter, og at intervokaliske og finale konsonanter reduceres oftere end initiale konsonanter og konsonantklynger (Schachtenhaufen 2013:194 f). Hvis man sammenligner to grupper af ord med forskellige forekomster af reducerbare elementer, kan man risikere at de to grupper har forskelligt reduktionspotentiale.

For det tredje er grundlaget for sondringen mellem grammatiske og leksikalske ord mere intuitionsbaseret end teoretisk funderet. De to slags ord går som sagt ofte under betegnelserne funktionsord og indholdsord (fx Bell et al. 2009; Kallay & Redford 2014; Rosenberg et al. 1985). Disse betegnelser vidner om en forestilling om at nogle

ord har indhold, men andre har funktion. En sådan forestilling går stik imod en funktionel grammatisk tilgang, hvor alle sproglige udtryk har en indholdsside, og hvor indholdet netop er en funktion. Symptomatisk nok er sondringen mellem funktion og indhold da også for vag til at muliggøre en udmøntning i præcise kriterier for hvornår et ord tilhører den ene eller den anden kategori. Derfor defineres den undertiden ved en anden sondring mellem hhv. lukkede og åbne ordklasser (fx Segalowitz & Lane 2000; Harley 2006:118). Denne sondring er empirisk funderet, men som et forsøg på at udmønte sondringen mellem grammatiske og leksikalske ord er den ikke mindre problematisk. For det første er det tværsprogligt variabelt hvilke ordklasser der er lukkede, og hvilke der er åbne. I germanske, romanske og slaviske sprog udgør verber fx åbne klasser, mens de i visse andre sprog udgør en lukket klasse (Pawley 2006). For det andet tyder agrammatismeforskningen på at der inden for lukkede ordklasser kan skelnes mellem leksikalske og grammatiske ord. I en undersøgelse af pronomproduktionen i fransk agrammatisk tale argumenterer Ishkhanyan et al. (under udgivelse) således for at nogle franske pronomener er grammatiske (fx *y*, *je*, *me*), mens andre er leksikalske (fx *moi*). Friedericis (1982) undersøgelse af præpositioner i tysk afatisk tale kan tages til indtægt for en lignende sondring mellem grammatiske og leksikalske præpositioner. Der er altså intet teoretisk grundlag for, som Bell et al. (2009) gør det, at kategorisere pronomener og præpositioner *en bloc* som grammatiske ord (funktionsord). Og med en teoretisk uklar sondring mellem grammatik og leksikon er det heller ikke klart hvorfor denne sondring skulle påvirke reduktionsgraden.

For at imødegå de to første af de tre ovenfor nævnte problemer koncentrerer vores undersøgelse sig specifikt om ord med dobbelt status, ord som *har* der både kan være grammatiske og leksikalske. Reduktionspotentialet og frekvensen af ordformen er det samme for den grammatiske og den leksikalske version, så vi kan isolere effekten af grammatik fra effekten af reduktionspotentialer og ordformsfrekvens. Forudsigeligheden af ordet vil dog fortsat være forskellig for den grammatiske og leksikalske variant. For at imødegå det tredje problem bygger denne undersøgelses skelnen mellem grammatiske og leksikalske ord på et stærkere teoretisk fundament og anvendelige kriterier.

2. Definition af grammatik vs. leksikon

Manglen på en teoretisk forankring af sondringen mellem grammatiske og leksikalske ord skyldes svagheder ved de teorier der i øjeblikket dominerer grammatikforskningen: generativ grammatik og konstruktionsgrammatik. For generativ grammatik er sondringen mellem grammatik og leksikon ganske vist helt central, men med deres fokus på syntaktisk struktur er generative teorier dårligt udrustede til at håndtere ordsondringen. Omvendt er konstruktionsgrammatikken med sit fokus på ord og andre lagrede størrelser i princippet godt udrustet til at håndtere denne sondring, men den understreger lighederne mellem lagrede størrelser i en grad så kontrasten mellem grammatik og leksikon ignoreres (jf. fx Trousdale 2014).

Et nyt alternativ til disse teorier er Boye & Harders (2012) teori om grammatisk status. Ligesom konstruktionsgrammatikken er den funktionel og usus-baseret, men ligesom generative teorier insisterer den på sondringen mellem grammatik og leksikon. I modsætning til begge indebærer den præcise kriterier for klassifikationen af ord og andre størrelser som grammatiske eller leksikalske. Teoriens kerne er den antagelse at kontrasten mellem grammatik og leksikon må forstås som en konventionaliseret mekanisme til at prioritere dele af komplekse sproglige meddelelser. Når vi ser specifikt på leksikalske og grammatiske ord, handler forskellen om at leksikalske ord har potentialet til at være en del af den diskursive forgrund, hovedsagen i ytringen. Leksikalske ord står ikke altid i forgrunden, men de kan stå i forgrunden. Grammatiske ord, derimod, udgør per konvention den diskursive baggrund. I eksempel (1) udtrykkes hovedsagen gennem de leksikalske elementer *Liva*, *har* og *hund*, og i (2) er den diskursive forgrund *Liva*, *spis*, *hund*.

(1) **Liva har en hund**

(2) **Liva har spist en hund**

De grammatiske ord *en* i (1) og (2) og *har* i (2) er derimod ikke en del af den diskursive forgrund – de udtrækker ikke den vigtigste information. Grammatiske elementer kan i særlige tilfælde kontrasteres med metalingvistik fokus, som i (3), hvor de grammatiske endelser *-er* og *-e* er prosodisk prominente.

(3) Jeg sagde bruger, ikke bruge

Grammatiske og leksikalske elementer adskiller sig derved i måden hvorpå de kan fokuseres. Grammatiske ord kan kun fokuseres selvstændigt med prosodi (og altså kun i metalingvistiske kontekster), mens leksikalske elementer kan fokuseres selvstændigt med prosodi såvel som ved kløvning, ved placering i fokusposition, med fokuspartikler eller ved reference i den efterfølgende diskurs (jf. Kristensen & Boye 2016). Ved at undersøge ordets fokuseringspotentiale kan man afgøre om det er grammatisk eller leksikalsk. Typisk repræsenterer ordformer entydigt enten grammatiske eller leksikalske ord, men former som *har* repræsenterer både grammatiske og leksikalske ord. I (1) ovenfor er *har* som sagt leksikalsk. I overensstemmelse hermed kan det fokuseres. I (4) kan det således læses som værende i skopus af den fokuserende nægtelse, dvs. benægtelse af ejerskab.

(4) **Liva har ikke en hund**

I (2) ovenfor er *har* derimod grammatisk. Dette *har* kan ikke (uden for metasproglige kontekster) fokuseres. I (5) kan det således ikke (uden metasproglig kontekst) læses som værende i skopus af den fokuserende nægtelse: *Ikke* vil typisk blive læst som havende skopus over *spist* eller over *en hund*.

(5) **Liva har ikke spist en hund**

Man kan selvfølgelig lægge tryk på *har* – også uden for metasproglige kontekster, fx hvis man som svar på (5) siger: *Jo, hun har*. Men det er ikke ensbetydende med at *har* er i fokus. Med *Jo, hun har* bringer man hele propositionen 'Liva har spist en hund' eller i hvert fald prædikatet 'spist en hund' i fokus. Tryk er i sig selv upålideligt som kriterium for fokusering (Jensen 2011).

Boye & Harders (2012) teori om grammatikken som baggrund understøttes af en række læsestudier hvor grammatiske ord oftere overses end leksikalske ord (Christensen 2015; Rosenberg et al. 1985; Vinther et al. 2015). En undersøgelse af reduktionsforskelle for grammatiske vs. leksikalske ord kan vise om grammatikken ikke bare op-

fattes som baggrund af modtageren, men også hos afsenderen, der kan markere den diskursive baggrund ved at reducere udtalen af den. Hvis grammatiske varianter på denne måde er mere reducerede end leksikalske varianter, kan reduktion ses som en indikation på grammatikalisering (jf. Lehmann 2002).

Formålet med vores korpusstudie er at undersøge om der er forskel på reduktionsmønstre for grammatiske og leksikalske ord der ligesom *har* indgår i homonyme ordparvarianter. I sådanne ordpar er ordformsfrekvens og reduktionspotentiale konstante. På baggrund af Boye & Harders (2012) teori om grammatikken som baggrund formoder vi at grammatiske ord vil være mere reducerede end leksikalske ord.

Det skal bemærkes at bestemmelsen af hvornår en ordform repræsenterer to ord, er teoretisk betinget. Når vi som omtalt ovenfor skelner mellem grammatisk og leksikalsk *have*, antager vi betydnings- og distributionsforskelle som afgørende kriterier: Det *have* vi anser for leksikalsk, forekommer med nominale objekter, mens det vi anser som grammatisk, kræver et perfektum participium. Med andre kriterier (fx baseret på forskelle i ortografi) var man nået til andre resultater (fx at *have* er et ord, der varierer i betydning og distribution). Vores resultater afhænger naturligvis af de ordpar vi har inddraget og dermed af specifikke bestemmelser af hvornår en ordform repræsenterer to ord. Hvis fx *have* var blevet regnet som ét ord, ville det være blevet klassificeret som leksikalsk pga. anvendelser som dem i (1) ovenfor, hvor det potentielt udtrykker hovedsagen i en ytring. Men da der så ikke ville være noget homonymt grammatisk ord at kontrastere det med, ville det slet ikke være blevet inddraget i undersøgelsen. Det skal imidlertid understreges at teorien om grammatikken som baggrund i sig selv er uafhængig af sådanne bestemmelser. Den forudsætter at man kan identificere et ord, men siger intet om kriterierne for denne identifikation.

3. Fremgangsmåde for korpusstudie af ord med dobbelt status

Korpusundersøgelsen blev udført efter følgende fremgangsmåde: 1) Udvælgelse af mål-ordformer, 2) Udarbejdelse af “forventet distinkt udtale”, 3) Korpus søgning, 4) Klassificering af korpusforekomster på baggrund af ortografi, 5) Udregning af reduktionsværdi og 6)

Analyse af reduktion for grammatiske vs. leksikalske forekomster. Trin 4 (klassificering af korpusforekomster) var adskilt fra trin 3 (korpussøgning) og trin 5 (udregning af reduktionsværdi) for at sikre at viden om ordenes udtale ikke (bevidst eller ubevidst) kunne påvirke om ordene blev klassificeret som grammatiske eller som leksikalske. Kun en person vidste hvordan ordene var blevet udtalt i korpusset, og denne person (medforfatter RS) deltog ikke i klassificeringen af forekomsterne. Detaljer om de enkelte trin fremgår af følgende underafsnit.

3.1 Trin 1: Udvælgelse af mål-ordformer

Medforfatterne LBK og KB udformede to lister over ordformer der repræsenterer både grammatiske og leksikalske ord. Den ene liste bestod af former der dels repræsenterer leksikalske pronominer, dels grammatiske ord såsom artikler (Tabel 1). Den anden liste bestod af former der dels repræsenterer leksikalske fuldverber, dels grammatiske hjælpeverber (Tabel 2).

Pronominal-former	Eks. på leksikalsk variant	Eks. på grammatisk variant
<i>den</i> <i>det</i>	demonstrativ brug: <i>den sø der</i>	som bestemt artikel, ikke-demonstrativ: <i>den samme</i>
<i>en</i> <i>et</i>	som numeral eller ubestemt pronomen <i>et af vinduerne</i>	som ubestemt artikel <i>et par centimeter</i>
<i>hvad</i>	som interrogativt pronomen <i>hvad hedder sådan noget?</i>	som relativpronomen <i>hvad tid på døgnet det er</i>
<i>nogen</i> <i>noget</i> <i>nogle</i>	som ubestemt pronomen: <i>har du nogen?</i>	med artikelfunktion: <i>nogen gardiner</i>

Tabel 1 Oprindelig liste over pronominer med dobbelt status

Verbale former	Eks. på leksikalsk variant	Eks. på grammatisk variant
	Som fuldverbum	Som hjælpeverbum
<i>blive, bliver, blev, blevet</i>	<i>hvis man <u>bliver</u> på parkens side</i>	<i>det <u>bliver</u> placeret</i>
<i>få, får, fik, fået</i>	<i><u>få</u> den på din venstre hånd</i>	<i><u>få</u> lokaliseret den</i>
<i>have, har, havde, haft</i>	<i><u>har</u> grønne gardiner</i>	<i><u>har</u> låst din cykel op</i>

Tabel 2 Oprindelig liste over verber med dobbelt status

Ordformerne i Tabel 1 og 2 er ikke tænkt som en fyldestgørende liste over danske ordformer med dobbelt status. De udgør snarere et udvalg af former der både som grammatiske ord og som leksikalske er hyppige i sprogbrugen. Et ord som *nu* har også dobbelt status, men er ikke inkluderet i listen. Mens leksikalsk *nu* er relativt hyppigt (*Han er i London nu*), formodes grammatisk *nu* at være relativt sjældent (*Han er nu rar*). *Nu* er derfor ikke en oplagt kandidat til en undersøgelse der sammenligner grammatiske og leksikalske ord – der ville formentlig ikke være tilstrækkeligt med forekomster i begge betingelser.

3.2 Trin 2: Udarbejdelse af ”forventet distinkt udtale”

For hvert af de 20 mål-ordformer på listen udarbejdede medforfatter RS en forventet distinkt udtale på basis af DanPASS-korpussets fonemiske annotering af ordet. Fx forventes *har*, der i korpusset er angivet /ha:ʔr/, udtalt [ha:ʔ] efter applikation af Grønnums (2005) fonem-til-allofon-regler. Vi har taget udgangspunkt i den maksimalt distinkte udtale, fx *bliver* der i korpussets fonemiske annotering er angivet som /bli:ʔvær/, som efter applikation af fonem-til-allofonreglerne forventes udtalt [bli:ʔvə], selvom det sjældent udtales mere distinkt end [bliɕʔ]. Udgangspunktet for analysen er dog ikke afgørende da vi er interesseret i den relative forskel mellem grammatiske og leksikalske ordformer, og ikke i hvor meget den enkelte udtale afviger fra det maksimalt distinkte. I den forventede distinkte udtale ser vi bort fra tryk da det behandles særskilt i vores analyse.

3.3 Trin 3: Korpussøgning

Medforfatter RS foretog en korpussøgning i talesprogskorpusset DanPASS (Grønnum, 2009) som består af monologer og dialoger fra 27 danske talere. Korpusset er opmarkeret med bl.a. fonetisk og ortografisk transskription. Det var derfor muligt at søge på alle forekomster af de 20 mål-ordformer og udtrække information om hvordan hver forekomst var udtalt samt i hvilken kontekst ordet var blevet ytret.

3.4 Trin 4: Klassificering af korpusforekomster

Hver korpusforekomst af de 20 mål-ordformer blev vurderet af to kodere. Koderne havde til opgave hver for sig at vurdere om den enkelte korpus-forekomst repræsenterede det grammatiske eller leksikalske ord. Medforfatterne LBK og KB klassificerede de pronomielle former, mens to studentermedhjælpere klassificerede de verbale. Ingen af disse fire kodere havde adgang til information om udtale, kun den ortografiske udskrift af korpusset. Koderne brugte Kristensen & Boyes (2016) fokustest til at afgøre om en korpusforekomst var grammatisk eller leksikalsk. Ifølge Boye & Harder (2012) har grammatiske elementer begrænsede fokuseringsmuligheder. Grammatiske elementer kan kun fokuseres metalingvistisk og kun med prosodisk prominens, fx med tryk på *har* i *han HAR spist æblet*.¹ Leksikalske elementer, derimod, kan fokuseres på en række etablerede måder, fx med kløvning eller fokuspartikel. Ved hjælp af Kristensen & Boyes (2016) fokustest kan man vha. fire kriterier afsøge danske ords fokuseringspotentiale og dermed afgøre om et ord er leksikalsk eller grammatisk. Hvis ordet opfylder et eller flere kriterier, regnes det for leksikalsk. Hvis det ikke opfylder nogen af kriterierne, regnes ordet for grammatisk.

Herunder er de fire kriterier vist. Opfyldelse af kriterierne er eksemplificeret med ordet *familie*, et leksikalsk ord der opfylder alle fire fokuskriterier:

¹ Scheuer (1995:43) skriver at danske hjælpeverber altid er trykssvage foran deres fuldverbumbstyrelse, og Rischel (1983) og Hansen & Lund (1983) ser enhedstryk som det normale mønster for sekvenser af hjælpeverbumb og fuldverbumb på dansk. Men Hansen & Lund (1983) nævner også fænomenet *modsatningstryk* som svarer til vores metalingvistiske fokus.

1. Ordet kan selvstændigt fokuseres ved hjælp af en fokus-konstruktion, fx sætningskløvning eller sætningsknode. Eks.: *det var familien der gjorde det.*
2. og/eller ordet kan fokuseres selvstændigt ved hjælp af en fokus-operator, fx *kun, ikke, især og netop*. Eks.: *ikke familien*
3. og/eller ordet kan selvstændigt fokuseres ved placering i en fokusposition, fx i sætningsskemaets indholdsfelt eller ved fokusering i forfeltet (forfeltet kan i visse tilfælde indeholde fokuseret materiale, jf. Kristensen 2013). Eks.: *han mødte familien.*
4. og/eller elementet kan adresseres i den efterfølgende diskurs vha. et ja/nej-spørgsmål, et hv-spørgsmål eller vha. en anafor. Eks.: *Han mødte familien. Hvad mødte han? Familien.*

Ord som *at* eller *og* kan ikke fokuseres vha. nogen af testene og kan derfor regnes for grammatisk og sekundære per konvention, mens ord som *Ib* og *løbe* kan fokuseres på mindst en af de fire nævnte måder og derfor kan regnes for leksikalske og potentielt diskursprimære.

Tabel 3 illustrerer hvorledes leksikalske varianter af de 20 mål-ordformer alle kan opfylde min. 1 kriterium (nogle opfylder som *familien* flere kriterier) hvorimod de grammatiske varianter ikke kan opfylde nogen af kriterierne.

	Mål-ord	Eksempel på opfyldelse af kriterium
Opfylder 1. kriterium: ”Fokuseringskonstruktion”	den/det _{LEX} en/et _{LEX}	<i>Det var <u>den/det/en/et</u> jeg så</i>
Opfylder 2. kriterium (men ikke 1. kriterium) ”Fokuspartikel”	blive/bliver/blev/blevet _{LEX} få/får/fik/fået _{LEX} have/har/havde/haft _{LEX}	<i>Han vil ikke <u>blive</u>. Han forlader os. Hun vil ikke <u>have</u> en hund. Hun vil kun <u>ae</u> den. Du <u>får</u> den ikke. Du skal købe den.</i>
Opfylder 3. kriterium (men ikke 1. og 2. kriterium) ”Fokusposition”	hvad _{LEX} nogen/noget/nogle _{LEX}	<i><u>Hvad</u> så du? Der kom <u>nogen/noget/</u> <u>nogle</u></i>

Reduktion i udtalen af grammatiske vs. leksikalske ord

Opfylder 4. kriterium (men ikke 1.-3. kriterium) “Senere adressering”	-	-
Grammatiske varianter der ikke opfylder nogen af de fire kriterier.	den/det _{GRAM} , en/et _{GRAM} , hvad _{GRAM} , nogen/noget/nogle _{GRAM} , blive/bliver/blev/blevet _{GRAM} , få/får/fik/fået _{GRAM} , have/har/ havde/haft _{GRAM}	

Tabel 3. Konstruerede eksempler der viser hvordan leksikalske varianter opfylder min. 1 kriterium (nogle opfylder flere kriterier, men kun det ene er illustreret her), hvorimod de grammatiske varianter ikke opfylder nogen. Eksemplerne illustrerer brug af ordene, men ordene bruges også på andre måder end som vist i tabellen. Eksempelvis kan leksikalsk *den* både være kernen i et nominal (*den*) og være en del af et nominal (*den vej* eller *den her vej*). I begge tilfælde kan leksikalsk *den* fokuseres selvstændigt – i modsætning til grammatisk *den* (som i *den samme*).

Hver korpusforekomst blev kategoriseret som grammatisk eller leksikalsk af to kodere der arbejdede hver for sig. Koderne havde ikke adgang til det fulde korpus, kun til en ortografisk transskription af hver forekomst af mål-ordene samt til de 25 ord der fulgte før mål-ordet og de 25 der fulgte efter. Koderne var overvejende enige, men i tilfælde af uenighed diskuterede koderne til slut klassificeringen, eller de konsulterede en tredje koder. Eksempelvis kan det være svært at afgøre om bestemmeren *den* i *den store klippe* er brugt demonstrativt eller ej. Den demonstrative brug er fokuserbar og dermed leksikalsk. Ligeledes kan det være svært at afgøre om talbetydningen af *en* i *en hvid firkant* er relevant eller ej – og dermed om ordet er leksikalsk.

I nogle tilfælde kan det ikke afgøres om en forekomst fra korpusset er grammatisk eller leksikalsk. I korpuseksempler med afbrydelse som *ja men jeg har okay* er det ikke tydeligt om ordet *har* er brugt grammatisk eller leksikalsk. Pga. selvafbrydelsen er der hverken en præsens participium der kan støtte en grammatisk læsning af *har* eller et objekt der kan støtte en leksikalsk læsning. I sådanne tilfælde hvor der hverken er argumenter for en grammatisk eller en leksikalsk fortolkning af korpusforekomsten, blev forekomsten

udeladt fra analysen. På dette grundlag blev 188 verbalforekomster og 18 pronomielle forekomster udeladt.

Endelig var der tilfælde hvor et ord blev gentaget, fx gentages mål-ordet *hvad* fire gange i *hvad hvad hvad hvad ligger der i nærheden*. I sådanne tilfælde indgår kun den sidste gentagelse i analysen. På dette grundlag blev 76 verbalforekomster og 209 pronomielle forekomster udeladt.

3.5 Trin 5: Udregning af reduktionsværdi

For hver korpusforekomst udregnede medforfatter RS en reduktionsværdi ved at sammenligne den forventede distinkte udtale (jf. trin 2) med den faktiske udtale i korpusset, dvs. den fonetiske transskription af forekomsten, og optælle antallet af afvigelser mellem de to. Hvis ordformen *har*, med den forventede distinkte udtale [hɑ:ʔ], i et konkret tilfælde udtales [hɑ], er reduktionsværdien 2, idet de to elementer [:] og [ʔ] er bortfaldet.

3.6 Trin 6: Analyse af reduktion for grammatiske vs. leksikalske forekomster

Af hensyn til den statistiske sammenligning af grammatiske og leksikalske varianter, var det væsentligt at alle mål-ordformer forekom hyppigt i begge varianter. Ordformer blev udeladt fra videre analyse hvis de forekom sjældnere end 5 gange i det samlede korpus i enten den grammatiske eller leksikalske variant. Dette frasorteringskrav påvirkede 10 af de 20 ordformer fra de oprindelige lister i tabel 1 og tabel 2: Ordformerne *blev*, *blevet*, *fik*, *haft*, *få* og *fået* havde alle under 10 forekomster i korpusset. Ordformerne *det*, *blive*, *får* og *havde* var hyppige i den leksikalske brug, men havde under 5 forekomster i den grammatiske brug. Den videre analyse fokuserede på mål-ordformer hvor både den leksikalske og den grammatiske variant var hyppige: De 7 pronomielle former *den*, *en*, *et*, *hvad*, *nogen*, *noget* og *nogle* samt de tre verbale former *bliver*, *har* og *have*.

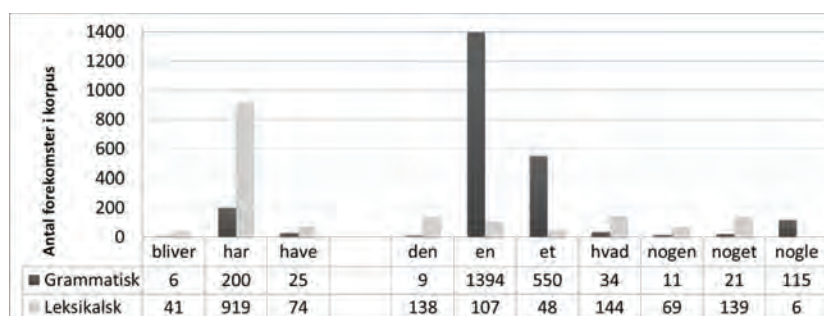
Den statistiske analyse blev udført i statistikprogrammet R version 3.1.1. (R Core Team 2014) med *lmer*-funktionen fra *lme4*-pakken. Med denne funktion kan man foretage *mixed model*-analyser, statistiske analyser hvor man tager højde for både systematisk og

tilfældig variation. Til at undersøge forskellen inden for de enkelte mål-ordformer brugte vi *glht*-funktionen i *multcomp*-pakken (Hothorn et al. 2008). For hver af de to ordlister (pronomener hhv. verber) blev der konstrueret to *mixed models*. Den ene var en *lmer-model* med den numeriske reduktionsværdi som afhængig variabel, mens den anden var en binomial *glmer-model* med en kategorisk afhængig variabel som indikerede om forekomsten var prosodisk prominent eller ej. For alle modeller var de uafhængige variable den tilfældige variabel taler, og de systematiske var ordformen, og kategoriseringen af ordformen (om den var grammatisk eller leksikalsk).

4. Resultater

Efter en opgørelse af antal korpusforekomster gennemgås forskelle i reduktion og i prosodisk prominens for leksikalsk vs. grammatisk brug af mål-ordformerne.

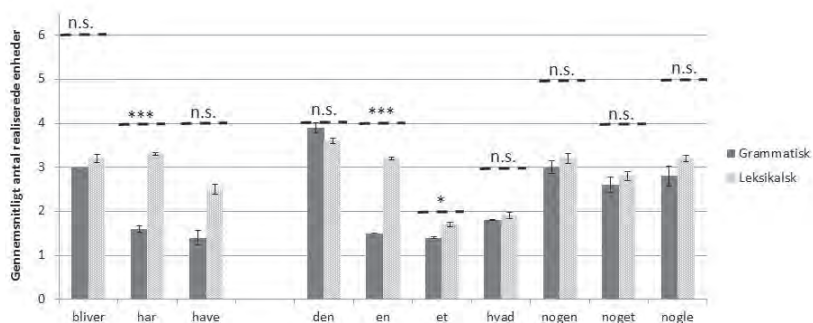
Som vist i figur 1, optræder de 10 mål-ordformer i DanPASS med forskellige frekvenser. Formerne *en*, *et* og *har* forekommer hyppigst. De øvrige 7 mål-former har hver under 200 korpusforekomster. Den grammatiske brug er hyppigere end den leksikalske for ordene *en*, *et* og *nogle*, mens den leksikalske brug er fremherskende for *bliver*, *have*, *den*, *hvad*, *nogen* og *noget*.



Figur 1. Frekvensen af hhv. grammatiske og leksikalske varianter af de 10 mål-former i analysen.

Det gennemsnitlige antal reducerede lydlige enheder for hhv. grammatiske og leksikalske varianter af de 10 mål-ordformer blev

opgjort og analyseret (Figur 2). Hvis man ser samlet på gruppen af verbale former, er de grammatiske varianter mere reducerede end de leksikalske ($z = -4,9$ $p < 0,001$). Ligeledes er de grammatiske varianter af pronomielle former samlet set mere reduceret end de leksikalske ($z = -3,1$ $p < 0,01$). Analysen viser individuelle forskelle mellem mål-ordformerne (der også kan have afsæt i de forskellige antal forekomster, jf. figur 1). For tre af mål-ordformerne (*har*, *en* og *et*) er den grammatiske variant signifikant mere reduceret end den leksikalske (*har*: $z = -3,3$ $p < 0,001$; *en*: $z = -7,7$ $p < 0,001$; *et*: $z = -2,1$ $p < 0,01$). For de øvrige 7 mål-ordformer er der ikke signifikante forskelle mellem grammatiske og leksikalske varianter.

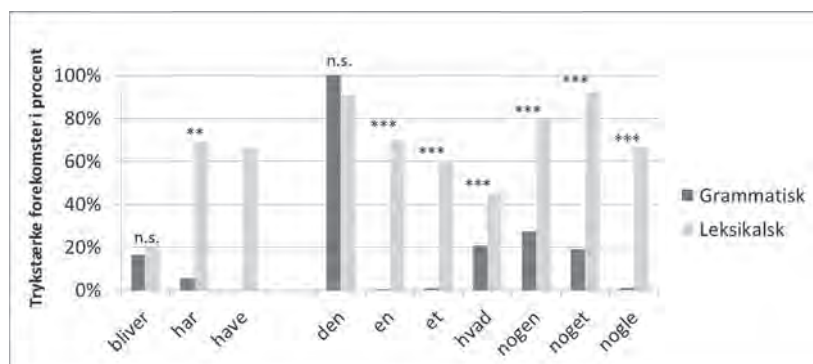


Figur 2 Det gennemsnitlige antal realiserede fonemer for grammatiske vs. leksikalske varianter af mål-ordformerne.

Den stiplede linje viser hvor mange lydlige enheder ordet kan have i den “forventede distinkte udtale”, fx vil ordet *har* udtalt maksimalt distinkt [hɑ:ʔ], i vores analyse have 4 lydlige enheder: [h], [ɑ], vokallængde og stød. De tre stjerner (***) over søjlerne for *har* og *en* viser at der realiseres signifikant færre enheder for de grammatiske forekomster for disse to mål-ordformer, og at p -værdien er under 0,001. En enkelt stjerne (*) over søjlerne for *et* signalerer en signifikant forskel hvor $p < 0,05$. For de øvrige mål-ordformer er der ingen signifikant forskel (“n.s.” = *not significant*).

Reduktion i udtalen af grammatiske vs. leksikalske ord

Figur 3 viser at de fleste mål-ordformer oftere realiseres med tryk på den leksikalske variant. Både verbale og pronomielle former er mere trykstærke i den leksikalske variant (verber: $z = 5,7$ $p < 0,001$; pronominer: $z = 9,9$ $p < 0,001$). *Have* forekommer aldrig trykstærkt i den grammatiske variant. *Har*, *en*, *et*, *hvad*, *nogen*, *noget*, *nogle* er oftere leksikalske end grammatiske når de realiseres med tryk (*har*: $z = 3,1$ $p < 0,01$; *en*: $z = 10,2$ $p < 0,001$; *et*: $z = 8,5$ $p < 0,001$; *hvad*: $z = 4$ $p < 0,001$; *nogen*: $z = 6,1$ $p < 0,001$; *noget*: $z = 9,1$ $p < 0,001$; *nogle*: $z = 6,5$ $p < 0,001$). Kun ordformerne *bliver* og *den* realiseres ikke mere trykstærkt i den leksikalske variant.



Figur 3 viser hvor mange procent af hhv. de grammatiske og de leksikalske ord der realiseres trykstærkt i korpuset. Evt. stjerner over søjlerne indikerer at der er signifikant forskel på om en ordform realiseres trykstærkt i den grammatiske vs. leksikalske variant (** betyder $p < 0,01$, *** betyder $p < 0,001$).

Posthoc undersøgte vi ved hjælp af en ny *mixed model* om der er forskelle i reduktion hvis man begrænser analysen til de trykssvage forekomster. Da den grammatiske version af *den* aldrig var trykssvag i korpuset, blev det udeladt af analysen. Posthoc-analysen viste ingen signifikant forskel på de grammatiske og leksikalske varianter af de pronomielle mål-ordformer. For de trykssvage verbalformer var der samlet set en signifikant forskel ($z = -2,6$ $p < 0,05$), dog ikke for analysen af de enkelte verbalformer, hvor kun *har* er grænsende til signifikant ($z = -1,8$ $p < 0,066$).

5. Diskussion

Korpusundersøgelsen viser overordnet at de grammatiske varianter af undersøgelsens ordformer er mere reducerede og mere trykssvage end de leksikalske – både når man ser samlet på gruppen af verbale former (*bliver*, *har*, *have*) og samlet på den pronomielle gruppe (*den*, *en*, *et*, *hvad*, *nogen*, *noget*, *nogle*). Analysen af individuelle mål-ordformer viser at de grammatiske varianter af 8 ud af 10 ordformer i undersøgelsen er mere trykssvage og/eller mere reducerede end de leksikalske varianter. Undtagelserne er ordene *den* og *bliver* hvor der ikke er forskel på den grammatiske og den leksikalske variant hvad angår tryk og reduktion.

Forskellen mellem de grammatiske og leksikalske varianter af de 8 øvrige ordformer går altid kun den ene vej: Det er aldrig den leksikalske version der er mere reduceret eller mere trykssvag. Ordets status som grammatisk eller leksikalsk afspejler sig derved i udtalen, i tråd med Boye & Harders (2012) teori om grammatik som konventionaliseret baggrundsinformation.

Som det fremgår af figur 2 og figur 3, har ikke alle ordformer samme reduktionsforskelle og samme trykmønstre. Disse individuelle forskelle understreger hvor vigtigt det er at man ikke slår alle grammatiske ord i hartkorn når man udtaler sig om deres reduktions- og trykmønstre. Ordformerne i undersøgelsen har forskellige reduktionspotentialer. Antallet af lydelige enheder i den maksimalt distinkte udtale varierer mellem 2 enheder (*et*) og 6 enheder (*bliver*). Nogle ord indeholder fonemer med stor reduktionstilbøjelighed, fx /ə/ i *have*, mens andre fonemer sjældent reduceres, fx /b/ i *bliver*. Endelig kan reduktionspotentialet tænkes at blive påvirket på *token*-niveau af ordets forudsigelighed i den givne kontekst og af hvilke naboord der forekommer før og efter mål-ordet. På denne måde kan nogle ord have et større reduktionspotential end andre.

Som vist i figur 1 er der forskel på hvor hyppigt de individuelle mål-ordformer forekommer i DanPASS-korpusset. Ordformerne *har*, *en* og *et* er de mest højfrekvente i undersøgelsen med flere hundrede forekomster hver. Den store variation i antal forekomster kan have haft indflydelse på den statistiske analyse. Hvis en ordform ikke

er tilstrækkelig frekvent, kan det være svært at finde signifikante forskelle for ordformen. De tre ordformer med flest forekomster (*har*, *en* og *et*) var netop de tre ordformer hvor vi så signifikante forskelle i reduktionsmønsteret for grammatiske og leksikalske varianter.

Det antages ikke sjældent at der er sammenhæng mellem et ords frekvens og tilbøjelighed til reduktion. Fx viser Pluymaekers et al. (2005) at visse affikser reduceres mere i højfrekvente ord end i lavfrekvente ord. Således inddrog Bell et al. (2009) i deres statistiske analyse af ords varighed både ordets frekvens i korpusset og om ordet var et indholdsord eller et funktionsord. De fandt at indholdsord havde kortere varighed hvis de var højfrekvente eller blev gentaget, hvorimod varigheden af funktionsords udtale ikke blev påvirket af frekvens. I vores undersøgelse er ordformsfrekvensen holdt konstant, idet vi sammenligner den grammatiske variant og den leksikalske variant af samme ordform. Dette er en forbedring ift. eksempelvis Bell et al (2009), for selvom de forholder sig til frekvens i analysen, er der så store frekvensforskelle mellem gruppen af funktionsord og gruppen af indholdsord at man kun på et teoretisk niveau kan skille frekvens fra ordkategori i den statistiske analyse.

Selvom frekvens og gentagelseeffekter ofte formodes at spille en rolle for reduktion, er der ikke tegn på det i vores korpusundersøgelse. I hvert fald ikke hvis man sammenligner den grammatiske variant med den leksikalske variant. I vores undersøgelse indgår både ordformer hvor den grammatiske variant er mere frekvent end den leksikalske (*en*, *et*, *nogle*), og ordformer hvor den leksikalske variant forekommer oftere end den grammatiske (*bliver*, *har*, *have*, *den*, *hvad*, *nogen*, *noget*). Når vi ser på reduktionsmønstre, er den grammatiske variant mere reduceret end den leksikalske, både for ordformerne *en* og *et* hvor den grammatiske variant er mere frekvent end den leksikalske og for ordformen *har* hvor den leksikalske variant er mest frekvent. På samme måde ses tryktab både når den grammatiske variant er mest frekvent (*en*, *et* og *nogle*) og mindst frekvent (*har*, *have*, *hvad*, *nogen*, *noget*). Det kan dog stadig tænkes at gentagelse af en given ordform inden for den samme optagelse kan påvirke udtalen i kombination med andre faktorer.

Hvor tidligere undersøgelser har gjort brug af ad hoc-betegnelser som åbne og lukkede ordklasser eller indholdsord og funktionsord, viser vi med denne undersøgelse at grammatiske og leksikalske ord kan undersøges med et stærkere teoretisk afsæt. Vi brugte Kristensen & Boyes (2016) fire kriterier for leksikalske ord. Her er det ikke afgørende om et ord er fokuseret i den pågældende kontekst, men om ordet har et fokuseringspotentiale, og dermed er leksikalsk. Undersøgelsen viser at denne kategoriserings definition af grammatisk brug stemmer overens med reduktions- og eller tryktabmønstre for 8 ud af 10 ord. For *den* var det ofte vanskeligt at afgøre om ordformen var brugt demonstrativt eller ej. I forekomsten *den samme* kan *den* ikke tolkes demonstrativt, og det må derfor være brugt grammatisk, men i forekomster som *den firkant* gengivet ortografisk, er det mere uklart om *den* er brugt demonstrativt – selv med en længere kontekst. For ordformen *bliver* var der heller ingen forskel i udtalen mellem den grammatiske og den leksikalske variant. For *bliver* var der ikke på samme måde risiko for fejkategorisering, men der var til gengæld ret få forekomster, og det kan have haft indflydelse på mulighederne for at vise statistiske forskelle. Samtidig har *bliver* ikke en stor reduktionsmargin i materialet. Selvom *bliver* if. DanPASS' fonemskrift har 6 fonemer, bliver det i alle forekomster udtalt med mellem 2 og 4 lyde, med et gennemsnit på 3,2 og en standardafvigelse på kun 0,5.

Vores undersøgelse er ikke den første til at præsentere forskelle i udtale for grammatiske og leksikalske ord, men i modsætning til tidligere undersøgelser (fx Bell et al. 2009) kan vi med afsæt i Boye & Harder (2012) forklare hvorfor der er forskelle. Med en kategorisering baseret på Boye & Harders (2012) teori om grammatikken som baggrund viser vi at grammatiske ord der per konvention er diskursiv baggrundsinformation, også fonetisk markeres som baggrundsinformation – som gruppe set er de grammatiske ord mere reducerede og mere trykssvage. Denne observation, der er baseret på et studie af normale danske talere, stemmer også overens med tale fra agrammatiske afatikere. I talen fra personer med agrammatisme ser man ofte at grammatiske elementer helt udelades, mens leksikalske bevares i talen (Ishkhanyan et al. under udgivelse; Kristensen & Boye 2016).

Vores korpusundersøgelse viser at grammatikken er baggrundsinformation ikke blot i perception af sprog, men også i produktion. Tidligere læseforsøg har vist at læsere lægger mindre mærke til de grammatiske elementer end til de leksikalske, og vi viser nu at grammatiske varianter af den samme ordform også er markeret af taleren som baggrund.

I dette studie viser vi at en kategorisering med afsæt i Boye & Harder (2012) kan forklare forskelle i udtalen af verbale og pronomielle ordformer. Det er ukontroversielt at kategorisere hjælpeverber som grammatiske og fuldverber som leksikalske, mens det er mere kontroversielt at opdele pronomielle ordformer i grammatiske og leksikalske ord. Pronominelle udtryk anses i mange studier for grammatiske (fx i Bell et al. 2009), men vi viser her at der kan være god grund til at kategorisere visse pronominer som leksikalske.

Boye & Harders (2012) teori om grammatikkens baggrundsstatus er en af mange teorier om grammatik. Andre grammatiske teorier har andre kriterier for at skelne mellem grammatik og leksikon. Her skelner vi mellem grammatiske og leksikalske elementer, men som foreslået af grammatikaliseringsforskere som Hopper & Traugott (2003:6-7) kan man forestille sig et kontinuum mellem grammatisk og leksikalsk brug. I så fald kunne udtalen tænkes at variere med placeringen på dette kontinuum. Fremtidige studier kan vise om disse andre teories skelnen mellem grammatik og leksikon i samme grad kan forudsige reduktionsmønstre.

6. Konklusion

Vores korpusundersøgelse af *blive, få, have, den, det, en, et, nogen, noget, nogle* og *hvad* viser at der for de fleste ordformer er forskel i udtalen af de grammatiske og de leksikalske varianter. De ordformer der viser en forskel i udtale, har en mere reduceret og/eller mere trykssvag udtale i den grammatiske variant end i den leksikalske – i tråd med teorien om grammatiske ords baggrundsstatus.

Litteratur

- Bell, Alan, Jason M. Brenier, Michelle Gregory, Cynthia Girand & Dan Jurafsky (2009) Predictability effects on durations of content and function words in conversational English. *Journal of Memory & Language* 60:92-111.
- Boye, Kasper & Peter Harder (2012) A usage-based theory of grammatical status and grammaticalization. *Language* 88:1-54.
- Christensen, Marie Herget (2015) Ændringsblindhed i sætningsprocessering i skriftlig dansk – opmærksomhedsforskelle på homonyme hjælpe- og fuldverber. 15. Møde om Udforskningen af Dansk Sprog, 9.-10. oktober 2014:83-97.
- Friederici, A. D. (1982) Syntactic and semantic processes in aphasic deficits: The availability of prepositions. *Brain and Language* 15:249-258.
- Garrett, M. F. (1980) Levels of processing in sentence production. *Language production*. Gordon H. Bower (udg.). London: Academic Press:177-220.
- Grønnum, Nina (2005) *Fonetik og Fonologi – Almen og dansk*. 3. udg. København: Akademisk Forlag.
- Grønnum, Nina (2009) DanPASS – A Danish Phonetically Annotated Spontaneous Speech Corpus. *Speech Communication* 51:594-603.
- Hansen, Erik & Jørn Lund (1983) *Sæt tryk på, syntaktisk tryk i dansk*. København: Danmarks Lærerhøjskole.
- Harley, Heidi (2006) *English words: A linguistic introduction*. Cambridge, MA: Blackwell.
- Hopper, Paul. J. & Elizabeth C. Traugott (2003) *Grammaticalization*. Cambridge: Cambridge University Press.
- Hothorn, Torsten, Frank Bretz & Peter Westfall (2008) Simultaneous Inference in General Parametric Models. *Biometrical Journal* 50:346-363.
- Ishkhanyan, Byurakn, Halima Sahraoui, Peter Harder, Jesper Mogensen & Kasper Boye (under udg.). Grammatical and lexical pronoun dissociation in French speakers with agrammatic aphasia: a usage-based account and REF-based hypothesis. *Journal of Neurolinguistics*.
- Jensen, Eva Skaftø (2011) Antagelser om fokus – skitse til et forskningsprojekt. *Ny Forskning i Grammatik* 18:105-122.
- Kallay, Jeffrey & Melissa A. Redford (2014) Longitudinal data on the production of content versus function words in children's spontaneous speech. *The Journal of the Acoustical Society of America* 136:2263-2263.
- Kristensen, Line Burholt (2013) Context, you need: Experimental approaches to information structure processing. Ph.d.-afhandling, Københavns Universitet.

- Kristensen, Line Burholt & Kasper Boye (2016) Svært ved grammatikken – del 3. *Logos* 77:27-31.
- Lehmann, Christian (2002) *Thoughts on grammaticalization*. Second, revised edition. Erfurt: Seminar für Sprachwissenschaft der Universität. https://www.researchgate.net/profile/Christian_Lehmann10/publication/242515466_Thoughts_on_grammaticalization/links/553a3ada0cf245bdd7627268.pdf (tilgået oktober 2016).
- Pawley, Andrew (2006) Where have all the verbs gone? Remarks on the organisation of languages with small, closed verb classes. *11th Binnennial Rice University Linguistics Symposium*:16-18.
- Pluymackers, Mark, Mirjam Ernestus & R. Harald Baayen (2005) Lexical frequency and acoustic reduction in spoken Dutch. *Journal of the Acoustical Society of America* 118:2561-2569.
- R Core Team (2014) *R: A language and environment for statistical computing*. Wien, Østrig: R Foundation for Statistical Computing.
- Rischel, Jørgen (1983) On unit accentuation in Danish – and the distinction between deep and surface phonology. *Folia Linguistica* 17:51-98.
- Rosenberg, Beth, Edgar Zurif, Hiram Brownell, Merrill Garrett & Dianne Bradley (1985) Grammatical class effects in relation to normal and aphasic sentence processing. *Brain & Language* 26:287-303.
- Schachtenhaufen, Ruben (2013) *Fonetisk reduktion i dansk*. Ph.d.-afhandling, Copenhagen Business School.
- Scheuer, Jann (1995) *Tryk på Danske Verber* (= *RASK*. Supplement 4). Odense: Odense Universitetsforlag.
- Segalowitz, Sidney J. & Korri C. Lane (2000) Lexical access of function versus content words. *Brain and language* 75:376-389.
- Trousdale, Graeme (2014) On the relationship between grammaticalization and constructionalization. *Folia Linguistica* 48:557-578.
- Vinther, Nicoline Munck, Kasper Boye & Line Burholt Kristensen (2015) Grammatikken i baggrunden: Opmærksomhed under læsning. *NyS* 47:99-139.