

Narrativt talesprog – rå vs. redigeret

AF PHILIP DIDERICHSEN

I denne artikel undersøger jeg ideer om godt (medie)sprog ved at sammenligne to korpusser med forskellige gengivelser af autentisk talesprog: Dels præcise sprogvidenskabelige transskriptioner, dels citerede passager i journalistiske featureinterviews. De to korpusser sammenlignes statistisk for at afdække hvilke grammatiske konstruktioner der er karakteristiske for hver af de to sprogformer. Resultaterne viser at mens der er få konstruktioner der er signifikant overhyppige i aviscitaterne, er der en række forskellige konstruktioner der er signifikant overhyppige i de sprogvidenskabelige transskriptioner. Med andre ord: Der er en række grammatiske konstruktioner i naturligt talesprog der sjældent forekommer i talesprog redigeret til featureinterviews. De fleste af disse konstruktioner kan opfattes som ‘tomgang’ eller ‘støj’, som journalister eksplicit rådes til at holde sig fra. Der viser sig således en norm om at undgå sproglig ‘sjusk’ på konstruktionsniveau i redigeret talesprog.

1 Indledning

I denne artikel undersøger jeg ideer om godt (medie)sprog fra en ny vinkel. Vha. statistisk korpussammenligning afdækker jeg hvilke grammatiske forskelle der er på to forskellige gengivelser af talesprog: Dels den rå men præcise gengivelse i sprogvidenskabelige transskriptioner og dels den mere polerede, redigerede form i citerede passager i avisinterviews. Det kan, på et kvantitativt empirisk grundlag, sige noget om hvordan ideer og normer om hvad godt mediesprog er, kommer til udtryk i sproglige valg.

Der er inden for de seneste tiår vokset en rig litteratur frem i den gren af dansk sprogvidenskab der beskæftiger sig med talesproglige normer. SLICE-projektet er et europæisk netværksprojekt hvis overordnede formål er at undersøge hvordan sproglige ideologier påvirker sprogforandring (Kristiansen & Coupland 2011; Kristiansen & Grondelaers 2013; Thøgersen, Coupland & Mortensen 2016). SLICE står for *Standard Language Ideologies in Contemporary Europe*. Det grundlæggende spørgsmål er hvordan dialekter og andre lokale varieteter påvirkes af – og påvirker – udbredte og evt. officielt understøttede ideer om ‘det bedste sprog’ i forskellige lande og områder i Europa. Det helt naturlige fokus er således de forskel-

lige talesprog rundt om i Europa, dvs. navnlig dialekter, men også det talte mediesprog, der flere steder de facto udgør den talte standardnorm. Et hjørne af interessefeltet dækker 'sjusk', dvs. talesprogets mange (notorisk udskældte) fonetiske reduktioner og hvordan sproget kan fungere så glimrende med dem (Heegård & Henriksen 2012, 2013). I begge sammenhænge har der været et rigt udbytte hvad angår ny viden om hvorfor og hvordan især fonetisk variation forekommer og fungerer (se også Gregersen & Kristiansen 2015).

På det grammatiske felt har der også været grøde, og der er i de senere år blevet undersøgt talrige grammatiske fænomener i dansk(e) talesprog, herunder sætningsknuder og andre underordnings- og ordstillingsfænomener, realitets- og modalitetsmarkering, givethed, generiske og refleksive pronomener m.m. (Jensen (2003), Diderichsen (2008), Jensen (2005, 2007, 2009a,b, 2011), Boye et al. (2012), Christensen (2012), Mortensen (2012), Christensen & Jensen (2014), Christensen (2019); se desuden *Skrifter om Samtalegrammatik* på samtalegrammatik.dk). Hvor de fonetiske undersøgelser i høj grad har beskæftiget sig med hvilke former for variation der mere eller mindre bevidst ændres og/eller borttenses under påvirkning af diverse toneangivende talenormer (eller omvendt fremføres med stolthed og måske på trods), har man i de grammatiske undersøgelser i mindre grad interesseret sig for variationens eventuelle forbindelse til identitetsskabelse og -markering, eller mere generelt, normpåvirkning. Der findes dog relevante undersøgelser, fx le Fevre Jakobsen og Ejstrup (2011), som dokumenterer hvordan oplæsningsstilen i TV-avisen gradvist er blevet mere og mere talesproglig gennem tiderne – også på det grammatiske område. Se også Christensen og Jensen (2015) og Christensen og Jensen (under udgivelse).

Denne artikel bidrager til en forståelse af hvordan (skrift)sproglige normer og idealer påvirker repræsentationen af talesprog i en specifik sammenhæng, nemlig det journalistiske interview. Dette ved at identificere grammatiske forskelle på sociolingvistiske udskrifter og citeret tale i avisinterviews, altså to meget forskellige gengivelser af spontan tale, med korpusstatistiske metoder.

Talesprog er blevet undersøgt vha. korpusmetoder så længe der har været talesprogskorpusser til (Biber 1991; Allwood 1998; Leech, Rayson & Wilson 2001; Henriksen 2002, 2006; Henriksen & Allwood 2005). Ambitionen har bl.a. været at udfordre forestillingen om at talesprog skulle være skriftsproget underlegent i al sin kaotiske ustrukturerethed, herunder ideen om at talesprog skulle være en sjusket og mangelfuld manifestation i forhold til sprogets mere rene og ægte, skriftlige fremtrædelsesform. Korpusundersøgelserne har her været med til at dokumentere at talesproget fungerer inden for helt andre rammer end skriftsproget (på grund af

talens særlige produktionsomstændigheder). Det er fx blevet vist at dansk talesprog på væsentlige distributionelle parametre sågar ligner svensk talesprog mere end det ligner dansk skriftsprog (Henrichsen & Allwood 2005).

I denne artikel beskæftiger jeg mig med en skriftsproglig repræsentation af tale der på sin vis er netop det skriftsproget skulle være ifølge de gamle dogmer (altså en rensset, redigeret, idealiseret udgave af talesproget): den citerede tale i avisinterviews.

Hvorfor aviser? Først og fremmest fordi citater i avisinterviews er noget af det tætteste man kan komme på autentisk tale redigeret til den skriftlige modus – i hvert fald hvis materialet skal være tilgængeligt i elektronisk form i mængder som det giver mening at arbejde korpusstatistisk med. Der findes ganske vist andre typer materiale der kunne komme i betragtning. Dels er der referaterne af Folketingets forhandlinger (tilgængelige på folketingstidende.dk), som har gennemgået en redigeringsproces der bl.a. omfatter “tillempling af talesproget til et talesprogsnært, syntaktisk sammenhængende skriftsprog”, “en nænsom redigering af sproget, der sikrer, at talerens intention med det sagte fremgår” og “rettelse af faktuelle fejl og fortællelser” (redaktionssekretær ved Folketingstidende Michael Ejstrup, personlig korrespondance). Altså umiddelbart en form for bearbejdning der kunne være relevant at undersøge her. Dels er der talerne på portalen dansketaler.dk, der også udgør en anseelig mængde talt sprog i skriftlig, elektronisk form. Talerne er dog mindre oplagte end avisцитater da de sjældent gengiver spontant talesprog: Hvor avisinterviews i bund og grund har samme tilblivelseshistorie som sociolingvistiske interviews (der (sam)tales, og talen gengives herefter skriftligt), gennemgår taler den modsatte proces idet de bliver til som et skriftligt produkt og herefter fremføres mundtligt. Folketingsreferaterne er både gengivelser af spontant talesprog og tilgængelige i rigelige mængder, men de ligger til gengæld genremæssigt længere væk fra sociolingvistiske interviews end visse avisinterviews.

Avisinterviews kan være mange ting. Men da der er så meget avismateriale tilgængeligt, er dette egentlig bare en fordel. Det er nemlig dermed muligt at udtrække netop de citater der er mest sammenlignelige med det talesprog der findes udskrevet i talesprogskorpusserne. De avisinterviews der minder mest om sociolingvistiske interviews, findes sandsynligvis inden for den fortællende journalistik. I denne artikel tager jeg derfor så vidt muligt udgangspunkt i narrative passager i de to undersøgte delkorpusser.

Ud over det vigtige hensyn til sammenlignelighed er der flere fordele ved at tage udgangspunkt i narrativer. Der er praktiske fordele: For det første er den mundtlige

fortælling en af de talegenrer der nærmest pr. definition forekommer oftest i sociolingvistiske interviews (jf. fx Ag 2009); for det andet er narrativer på forhånd identificeret og opmærket i store dele af LANCHART-centrets store talesprogs-korpus; og sidst men ikke mindst er narrativer en velbeskrevet talegenre, og det er således muligt at operationalisere dens vigtigste definerende træk til fremsøgning af narrative passager i aviscitater.

Og ikke mindst fra en teoretisk synsvinkel er narrativgenren velvalgt. Fortællende journalistik bliver til i et interessant spændingsfelt mellem normer og fortælle-mæssige hensyn. Eksempelvis karakteriseres featurejournalistik som formidling af fakta med fiktionens virkemidler (Blom 2013: 120). Det afføder et krav om autenticitet, men også en forventning om liv og kolorit. Det interessante er at koloritten bl.a. netop kan komme fra mere eller mindre ordrette gengivelser af talesprogs-træk – som man imidlertid generelt rådes til at bruge med forsigtighed eller helt undgå, bl.a. for ikke at komme til at virke nedladende eller parodierende (Blom 2013: 108, Hennessy 2013: 77f.; Thurah 2013: 81ff.). Paradoksalt nok kan det at være loyal mod en kilde således også bestå i at undlade at citere for loyalt. Dertil kommer mere generelle formaninger om at skrive kort og ikke for kompliceret osv. (fx Thurah 2013: 55). Idealet om det gode sprog som en “blank, gennemsigtig rude” der lader budskabet passere uden at gøre opmærksom på sig selv (Kristiansen 1996: 11), kolliderer altså i princippet både med ønsket om kolorit og med kravet om autenticitet idet *for* tro gengivelser nemt kan få en ‘larmende’ eller på anden måde uønsket virkning.

I dette krydsfelt mellem en sober skriftnorm og en farverig genre skal journalisten træffe sine grammatiske valg. Spørgsmålet i denne artikel er: Hvordan adskiller journalistens redigerede gengivelse af det fortællende talesprog sig fra sociolingvistikens omhyggeligt registrerende ditto? Der findes efterhånden en solid viden om hvordan talesprog generelt adskiller sig fra skriftsprog, jf. kilderne ovenfor. Men hvad når skriftsproget skal repræsentere faktisk forekommet talesprog i dette kryds-pres mellem norm og genre?

Spørgsmålet er ikke behandlet for dansk før. Marianne Rathje har ganske vist beskæftiget sig med citater af andres talesprog – men *i* talesprog, ikke i skriftsprog (Rathje 2011). Og talesproglig stil i skriftlig fremstilling berøres fx i Galberg Jacobsen og Skyum-Nielsen (2007), men ikke med samme empiriske dybde som her.

I international sammenhæng må man konsultere litteraturforskningen for at finde lignende undersøgelser. I projektet Corpus Linguistics in Context ved Birmingham University (CliC-projektet) er der aktuelt en interesse for korpusbaseret

stilistisk analyse af ‘citeret’ tale, dvs. replikker/direkte tale i skønlitteratur. Her arbejdes der med hvad der gør replikker naturlige og taleagtige (“speech-like”) ud fra korpussammenligninger af replikkerne selv med ikke-replikker på den ene side og transskriberet talesprog på den anden. Den foretrukne metode i projektet er undersøgelser af forekomststatistik for “lexical clusters”, dvs. ordfølger (typisk 4-grammer) der forekommer signifikant hyppigere i én sprogform end i andre (Mahlberg & Wiegand 2018; Mahlberg et al. 2019). Selv om replikkerne er fiktive og altså ikke citerer af autentisk tale, har CliC-undersøgelserne været en vigtig inspirationskilde til den undersøgelse jeg præsenterer her.

Den nærværende undersøgelse er også en sammenlignende korpusundersøgelse af ordfølger. Udover rent leksikalske ordfølger som i CliC-projektet fokuserer jeg her især på delvist skematiske mønstre bestående af både konkrete ordformer og abstrakte grammatiske kategorier i et forsøg på at identificere grammatiske konstruktioner der er karakteristiske for de to sprogformer. Begrundelsen for dette er at jeg ved at abstrahere fra det rent leksikalske niveau dels eliminerer behovet for at holde emner og ordforråd i de sammenlignede delkorpusser konstant, og dels gør det muligt at identificere forskelle af mere grammatisk art. Med det for øje gør jeg i det følgende en del ud af at introducere konstruktionsgrammatikken såvel som dens udløber, konstruktikografien, en disciplin der beskæftiger sig med fremfindning af konstruktionskandidater til konstruktikoner (dvs. samlinger af konstruktioner, jf. Lyngfelt et al. (2018)).

2 Baggrund

I dette afsnit vil jeg komme ind på resultater fra flere forskellige forskningsfelter der har relevans for undersøgelsen af citeret overfor transskriberet tale.

Hvordan adskiller tale sig fra skrift? Det gives der bl.a. en oversigt over i grundbogen *Dansk Sprog* (Galberg Jacobsen & Skyum-Nielsen 2007). Her nævnes træk som ordrigdom, ordgentagelse og tomgang (dvs. tænkepauseudfyldende tale), forskellige typer disfluenser, (korte) helsætninger fremfor ledsætninger, konstruktioner som sætningsknode og sætningskløvning (og fravær af passiv), korte og/eller fordoblede forfelter m.m.¹.

¹ Bemærk at “tomgang” bruges lidt forskelligt i den relevante litteratur. I DR’s sprogpolitik (2018: 11) bruges termen fx om strategisk tomgang, dvs. talen uden om snarere end simpel tænke-tomgang, som er den intenderede betydning i denne artikel (jf. også Thurah 2013: 81).

Den korpusbaserede forskning bidrager med til dels overlappende indsigter: I talesprog genbruges de samme ord langt oftere end i skriftsprog, og de hyppigste ord er således hyppigere i talesprog end i skriftsprog, m.a.o. er den leksikalske diversitet større i skriftsprog. Både på dansk, svensk og engelsk er substantiver, præpositioner og adjektiver hyppigere i skriftsprog end i talesprog og pronomener, adverbier og interjektioner hyppigere i talesprog end i skriftsprog. Foruden fordelingen af ordklasser er også andre væsentlige statistiske parametre mere ens for dansk og svensk talesprog end for dansk (hhv. svensk) talesprog over for dansk (hhv. svensk) skriftsprog, nemlig både frekvensdistributionen af ordtyper og selve ordvalget. Disse særtræk kan til dels forklares af de særlige produktionsomstændigheder for tale og skrift, hvor tale fx er mere “impulsivt” og “reaktivt”, mens skrift fx pga. sin mere distancerede modtagerrelation er nødt til at være mere eksplicit og beskrivende (Allwood 1998; Henrichsen 2002; Henrichsen & Allwood 2005).

Fælles for disse kilder er at de ikke besvarer spørgsmålet om hvordan direkte efterligninger af talesprog i skriftlige genrer adskiller sig fra transskriberet tale. Til det formål er der inspiration at hente i international litteraturforskning. Det især Dickens-fokuserede CliC-projekt ved Birmingham University undersøger vha. et særligt korpussammenligningsværktøj hvordan litterære replikker adskiller sig fra ikke-replikker (Mahlberg & Wiegand 2018; Mahlberg et al. 2019). På den måde identificerer projektet mønstre der er særligt karakteristiske for replikker og dermed (forestillet) taleagtige (fx *never heard of such a thing*). Det kan udnyttes til at vurdere den litterære effekt af den slags vendinger *uden for* replikker, hvilket fx er relevant i diskussionen af såkaldt fri indirekte tale. Projektet finder også at udvalget af talesproglige konstruktioner er mere begrænset i fiktive replikker end i autentisk tale. Af gode grunde siger CliC-projektet dog ikke noget om hvordan tekstlige efterligninger af virkelig, autentisk tale tager sig ud.

Den undersøgelse der afrapporteres her, minder en del om CliC-projektet hvad metoder angår, og også med hensyn til undersøgelsesgenstand, nemlig enheder større end det enkelte ord, jf. nedenfor. undersøgelsen adskiller sig dog også på flere måder fra CliC-projektet, bl.a. i valget af materiale, der er endnu mere specifikt end CliC-projektets. For at gøre de to sprogformer der er i fokus her, så sammenlignelige som muligt, fokuserer jeg på en bestemt talegenre, nemlig den narrative.

Narrativer er blevet diskuteret grundigt inden for diverse forskningsfelter ud fra hvad deres funktion er (skabelse og fastholdelse af et konsistent selv(billede) over tid; fastholdelse af kollektiv, kulturel hukommelse; diverse makrotalehandlinger (fx underholde, overbevise osv.); som evolutionær motor bag grammatikken; osv.

(Linde 1993; Turner 1996; Nelson 2004; Stokoe & Edwards 2006). Genren er oprindelig især blevet studeret i myter, folkesagn og anden traditionel mundtlig overleveringskultur inden for etnologi og antropologi. Labov og Waletzky (1967) var de første der begyndte at interessere sig for narrativer i sociolingvistiske undersøgelser, dvs. som eliciteringsmål for sociolingvistiske interviews – især fordi de er både uformelle og allestedsnærværende og på den måde udgør en god ramme for undersøgelser af sociolingvistisk (fonetisk) variation. Narrativer defineres hos Labov og Waletzky med diverse obligatoriske og fakultative strukturelementer, som Erik Møller senere tog udgangspunkt i i sin definition af den mundtlige fortælling (Møller 1993). Motivationen for at analysere narrativer i den nærværende undersøgelse ligner Labov og Waletzky's. Det er ikke opbygningen af narrativerne eller de forskellige strukturelementers status som obligatoriske eller fakultative eller lignende der er i fokus. Genren er valgt af rent instrumentelle grunde. Den er hyppig og samtidig defineret så tilpas præcist at det er muligt at tilrettelægge materiale fra transskriptioner og aviscitater så det kan siges at være noget nær den samme type talesprog i to forskellige skriftlige sprogdragter. Især to definerende træk hentet fra Erik Møller er vigtige for den operationelle definition af narrativer der bruges til at fremsøge genren i avismaterialet. Det ene er at den begivenhed der berettes om i en narrativ, skal være selvoplevet, hvilket kan omsættes til det grammatiske træk 1. person singularis. Det andet er den såkaldte temporale lås, som er defineret som minimum to narrativsætninger (hovedsætning med handlingsverbum i datid eller historisk præsens) i kronologisk rækkefølge. Narrativer er således en del af denne undersøgelses metodiske rammeværk og ikke undersøgelsesgenstand i sig selv. Undersøgelsesgenstanden er talesproget som det kommer til udtryk i to sprogformer der hovedsageligt adskiller sig mht. skriftlig fremstillingsform – med særligt henblik på grammatiske enheder større end det enkelte ord og mindre end den komplette sætning.

Jeg afholder mig således i denne undersøgelse fra at sammenligne rent leksikalske forskelle. Dels er delkorporserne så små at forskelle i de emner der tales om, kan have skævvridende betydning. Desuden er emnevalg og lignende indholdsforskelle på de to korpusser ikke i fokus her; bestræbelsen er snarere at karakterisere de to forskellige repræsentationer af narrativt talesprog på et mere generelt plan. Endelig er det velbeskrevet hvordan visse funktionsord, interjektioner, udtryk for tøven osv. er overhyppige i talesprog i forhold til skriftsprog, jf. ovenfor. Derimod kunne det være interessant at finde grammatiske forskelle på de to sprogformer. Men abstrakte syntaktiske mønstre såsom bestemte sætningstyper eller lignende kan være svære at søge efter i korpusser fordi sådanne mønstre generelt vil være

sjældnere jo længere de er. Kortere syntaktiske mønstre kan også være svære at søge efter (jf. Asmussen & Ørsnes 2009); det er der dog til dels råd for, jf. nedenfor. Undersøgelsen her tager derfor en mellemvej mellem den leksikalske og den grammatiske yderpol. Med det for øje har jeg ladet mig inspirere af konstruktionsgrammatik og en af dens datterdiscipliner, konstruktikografien, som jeg præsenterer i det følgende.

Konstruktionsgrammatik opstod i kredsen omkring Charles J. Fillmore og George Lakoff ved UC Berkeley i senfirserne af et behov for at håndtere idiommer, ikke mindst idiommer med tompladser, som skabte problemer for traditionelle tilgange til grammatik, bl.a. fordi de ikke sjældent har deres helt egen syntaks. Begrebet konstruktion blev her indført i betydningen ‘konventionaliseret sproglig størrelse der kan være i det mindste delvis skematisk og kompleks’, i modsætning til fx enkeltord, der er konkrete og syntaktisk simple/minimale. Dette begreb blev så appliceret på både mindre og større sproglige størrelser i det såkaldte *syntax-lexicon continuum*, nu altså til og med konkrete enkeltord i den ene ende og op til maksimalt abstrakte skemaer i den anden. (Diderichsens sætningsskema (Diderichsen 1962) kunne således være et eksempel på en sådan maksimalt skematisk konstruktion). Efterhånden opstod der en komplet grammatisk teori hvor “The constructional tail has come to wag the syntactic dog” idet konstruktionsbegrebet gik fra at håndtere tilsyneladende perifere sproglige fænomener til at være helt centralt. Ideen blev videreført af bl.a. Adele Goldberg, Laura A. Michaelis, Knud Lambrecht og på sin vis Ronald Langacker, hvis Cognitive Grammar er kompatibel med konstruktionsgrammatik. Se (Croft 2001), som er et forsøg på at føre konstruktionsgrammatikken til sin logiske konklusion (ovenstående er refereret fra s. 15ff).

Konstruktioner kan altså være konventionaliserede form-indhold-størrelser på hvilket som helst grammatisk beskrivelsesniveau. Dog har der i konstruktionsgrammatikken altid været særligt fokus på konstruktioner der er i det mindste delvist skematiske, dvs. mønstre med i det mindste nogle tompladser. Et farverigt dansk eksempel kunne være udtrykket “NP *er ikke den* ADJ-*este* N PRÆP N”, fx “NP *er ikke den skarpeste kniv i skuffen*”, “NP *er ikke den hurtigste knallert på molen*” osv. med den fælles betydning ‘nogen er ikke alt for kvik’ (Trap-Jensen 2007, Farø & Lorentzen 2009). Et andet eksempel, der viser sig i de her analyserede data, er inkvitkonstruktionen “**være** .. sådan <cit>”² (Rathje 2011), der bl.a. er kendetegnet

² Når jeg enkelte gange i artiklen viser et lemma som repræsentant for alle dets former, er det markeret med **fed** skrift.

ved at være lettere grammatisk afvigende (det er ikke kanonisk grammatisk at en kopulakonstruktion har et citat som subjektsprædikat). Det er især sådanne delvist skematiske konstruktioner jeg vil beskæftige mig med i denne artikel med henblik på at afdække forskelle af mere strukturel art på de to delkorporer jeg undersøger – og på om der endda kan identificeres talesproglige konstruktioner der ikke er beskrevet før.

Gruppen af delvist skematiske konstruktioner, der lægger sig midt imellem leksikon og grammatik, har været genstand for videnskabelig interesse under forskellige navne. I international datalingvistik tales der om *multiword expressions* eller *MWEs* (Constant et al. 2017). I Danmark er forskningen hovedsageligt foregået i det fraseologiske område af leksikografien (jf. fx Farø & Lorentzen 2009), hvor faste flerordsforbindelser omtales som *frasemer* (i varianter med mange tompladser kaldet *fraseoskabeloner*). Andre taler simpelthen om faste udtryk. Uanset om feltet er konstruktionsgrammatik, datalingvistik eller fraseologi, karakteriseres delvist skematiske konstruktioner som minimum vha. de følgende to træk:

- Kollokation, dvs. overhyppig samforekomst af konstruktionens delelementer.
- Ikke-kompositionalitet, ikke-forudsigelighed, idiomacitet og/eller leksikalisering, dvs. lidt mere generelt den egenskab at en konstruktions fulde betydning ikke kan regnes ud alene ud fra delenes betydning og deres indbyrdes placering (dvs. den syntaktiske struktur).

Delvist skematiske konstruktioner voksede fra at være en sten i skoen til at agere startblok for konstruktionsgrammatikken. I moderne leksikografi er de et lignende irritationsmoment (Farø & Lorentzen 2009):

Fraseoskabelonerne volder problemer alle steder. Leksikografien har det meget bedre med manifeste størrelser end med tompladser.

Formentlig til dels motiveret af den slags frustrationer er der opstået et nyt felt der forsøger at understøtte både repræsentation og fremsøgning af konstruktioner. Dette felt kalder sig konstruktikografi (Lyngfelt et al. 2018).

Konstruktikografi ligger i grænselandet mellem grammatik og leksikografi og beskæftiger sig dels teoretisk med sproget ud fra en konstruktionsgrammatisk synsvinkel, dels med den praktiske opbygning af konstruktikoner. Et konstruktikon er en mellemtag mellem en grammatik og en ordbog. Et konstruktikon vil således typisk ikke opføre rent leksikalske størrelser såsom ord og (helt) faste udtryk og

heller ikke først og fremmest beskrive helt abstrakte grammatiske regler og mønstre, men derimod mestendels indeholde beskrivelser af sproglige størrelser der ligger imellem de to poler, altså delvist skematiske konstruktioner.

Konstruktikografien har bl.a. i Sverige fra begyndelsen været tæt knyttet til korpuslingvistikken og sprogteknologien. Sprogteknologiske korpusmetoder er oplagte værktøjer i fremfindingen af konstruktioner til indlemmelse i konstruktikoner. Således eksisterer der fx gennemprøvede metoder til at finde forbindelser af ord og ordklassekategorier der har en højere samforekomstfrekvens end man skulle forvente ud fra de enkelte elementers forekomstfrekvens (Bäckström et al. 2013; Forsberg et al. 2014).

Konstruktionsgrammatikken og konstruktikografien inspirerer denne artikel, på den ene side mht. de sproglige størrelser jeg vil fokusere på i min sammenligning, og på den anden side mht. metoder til fremfinding af dem.

3 Materiale

3.1 LANCHART-narrativer

LANCHART-korpusset er et p.t. knap 11 mio. løbende ord stort talesprogs-korpus indsamlet af Sprogforandringscentret ved Københavns Universitet. Dataene i korpusset er indsamlet fra forskellige egne af landet ad flere omgange (og senest også fra dansktalende enklaver på den anden side af Atlanten) for at kunne studere sprogforandring i virkelig tid (Gregersen & Kristiansen 2015). Korpusset er i verdensklasse i størrelse og kvalitet, og det er bl.a. værdifuldt ved at være et levende arkiv over de undersøgelser der er blevet udført i det, idet tidligere analysers kategoriseringer er bevaret i form af søgbar opmærkning.

I Gregersen, Nielsen & Thøgersen (2009) præsenteres et kodningsapparat, *Discourse Context Analysis* (DCA), der har til formål at karakterisere interaktionen i LANCHARTs store korpusmateriale på en sådan måde at centrets sociofonetiske sprogforandringsundersøgelser kan foretages på så sammenlignelige passager af materialet som muligt. Apparatet består af seks forskellige kodningsdimensioner, herunder dimensionen genre bestående af kategorierne narrativ, specifik redegørelse, generel redegørelse m.fl. (jf. også DGCSS 2011: 36). De seks dimensioner gør det muligt at udvælge passager til sammenligning der er så ens som muligt for at holde effekten af passeret tid maksimalt isoleret i studiet af sprogforandring. DCA-kodningen har imidlertid med det samme vist sig nyttig til andre formål også,

fx genreanalyser af materialet. Kodningen er også nyttig for nærværende undersøgelse idet genredimensionen gør det muligt automatisk at udtrække narrative passager.

I LANCHART-korpusset er omkring en tredjedel af interviewene opmærket med genre. Det her anvendte materiale er et udtræk af samtlige 511 passager opmærket som narrativer fra 399 informanter fra København (Amager-, BySoc- og Clarin-projekterne), Næstved, Odder og Vinderup (jf. dgcss.hum.ku.dk/forskning). Udtrækket er begrænset til de nyeste optagelser i korpusset, dvs. optaget mellem 2005 og 2010, for at gøre dataene så sammenlignelige med avisdataene som muligt mht. interviewtidspunkt. Både passager fra enkeltinterviews og gruppesamtaler er medtaget. Antallet af løbende ord i LANCHART-materialet er på 105.244.

De udtrukne narrativer er opmærket med værktøjet “Lemmatiser (i løbende tekst, inkl. POS-tags)” på clarin.dk (opmærkning i SOLIDUS-format: token/POS/lemma). Herefter renses de for alle interjektioner. Interjektioner er allestedsnærværende i LANCHART-korpusset, og de fjernes for at tilfældige disfluenser (fx *øh*’er eller *ja*’er midt i en konstruktion) i LANCHART-materialet ikke skal skævvride forekomststatistikken for de mønstre jeg leder efter.

En metodemæssig bekymring ved opmærkningen af LANCHART-materialet er at sprogteknologiske værktøjer typisk er udviklet til at håndtere skriftsproglige data, jf. Henrichsen (2005: 377):

Automatic tagging is of course fast and convenient, but not without its problems. Our taggers were trained mainly on written text sources, so the quality of the DanSPO and SweSPO tagging cannot be expected to match that of the written corpora.

I analyserne i den nærværende undersøgelse bevares imidlertid mange af de hyppigste og (dermed) mest flertydige ordformer som ordformer (jf. nedenfor), hvilket må formodes at kompensere for denne fejlkilde i rimelig grad. Fordelene ved den automatiske opmærkning er at den alt andet lige vil være ensartet for de to sammenlignede materialer, samt at opmærkningen ikke har været for ressourcekrævende – og derfor i det hele taget har været mulig.

3.2 Aviscitater

I en sammenlignende undersøgelse som denne er det vigtigt at de udvalgte materialer er så sammenlignelige som muligt. Da der ikke her er tale om en generel sammenligning af talesprog med (avis)skriftsprog, er det ikke relevant blot at udtage

noget tilfældig tekst fra nogle aviser. Det der er brug for, er skriftlige repræsentationer af mundtlige narrativer. Derfor består avismaterialet af narrativ(agtig) citeret tale fra interviews.

Materialet er udtrukket af Dansk Sprognævnns aviskorpus (udelukkende artikler fra Politiken 2016). I (DGCSS 2011: 37) er narrativgenren operationaliseret, især på baggrund af Labov og Waletzky's/Erik Møllers definition. Denne operationalisering ligger igen til grund for avisudtrækket her. Et af de få træk der er operationaliseret i tilstrækkelig grad til at være søgbart (givet den rette opmærkning), er den temporale lås. Her nævnes således *handlingsverber i datid i hovedsætninger* (søgbare via syntaktisk opmærkning), hvor handlingsverber udtrykker fysiske eller mentale handlinger (fx *gik, tænkte, sagde, mødte, besluttede, blev brugt, havde spist* (søgbare som ordformer)), og desuden nogle eksempler der ikke er handlingsverber (*var, boede, hed, kendte, sad, lå, kunne* (kan bruges som filtre i søgningen)). Handlingsverber der indgår i en temporal lås kan passende kaldes narrativverber. Dertil kommer det selvoplevede, der kan operationaliseres som forekomst af pronomener i 1.pers.sing.

Udtræksproceduren er konstrueret til at fremfinde citater der besidder disse narrative træk, og den falder i to trin. I første trin udtrækkes et delkorpus bestående af hele avisartikler der indeholder mindst én temporal lås mellem to narrativverber. Det anvendte søgemønster for en temporal lås gengives i simplificeret form herunder, hvor "tokens*" betyder 0 eller flere ordformer, og hvor narrativverber operationaliseres som verber i præteritum aktiv der er opmærket syntaktisk som hovedverber samt ikke er en af ordformerne *var, havde, kunne, ville, skulle, måtte, vidste, troede, følte, syntes, mente, stod, sad, lå, hang, boede, skete, kendte, betød, forstod, virkede, hed* (kategoriseret som ikke-handlingsverber på basis af en gennemgang af præteritumformer med en hyppighed på 4+ i korpusset, hvilket jeg vurderede kun ville efterlade en ubetydelig mængde resterende ikke-handlingsverber):

« + tokens* + narr.verb. + tokens* + . + tokens* + narr.verb. + tokens* + »

I dette trin frafiltreres desuden tekster annoteret med Infomedia-emneordet POLITIK for at øge sandsynligheden for at ramme de mere featureprægede interviewartikler. Udeladelsen af politiske interviews har desuden den fordel at der dermed sandsynligvis forsvinder nogle af de mere 'sminkede' eller strategisk redigerede data fra avismaterialet. Det er velkendt at interviewpersoner ofte får lov til at sige god for, og i mange tilfælde selv tilrette og direkte omformulere, udtalelser i

interviews, og denne mulighed er antagelig ikke mindst udbredt i interviews hvor der er noget på spil politisk.

I andet trin udtrækkes fra dette delkorpus alle passager i citationstegn (ofte bestående af adskillige sætninger) der indeholder mindst ét verbum i datid og mindst ét pronomen i 1.pers.sing. (*jeg, mig, min, mine*³). Alle sådanne passager fra en given tekst behandles som en samlet narrativ passage i den videre analyse. Selv om alle dele af et givet interview der ikke er realiseret som citeret tale, således ignoreres (dvs. elementer såsom alvidende forfatterstemme, indirekte tale osv.), er de resulterende passager næsten kohærente fortællinger i sig selv, dog med en tydelig fornemmelse af manglende detaljer her og der. Se fx dette eksempel:

Vi forsøgte desperat at få lov at gifte os, men vi fik at vide, at det kunne vi godt glemme alt om

[..] «Jeg var rigtig heldig, at jeg blev udvalgt» [..] «Han var min første kæreste. Min eneste. Det var kærlighed ved første øjeblik» [..] «Han havde de her blå øjne. Blå, naive øjne, drillede jeg ham med. For han tror altid det bedste om folk. Og er så optimistisk. Det stråler ud af hans øjne. Det synes jeg var så smukt» [..] «Vi forsøgte desperat at få lov at gifte os, men vi fik at vide, at det kunne vi godt glemme alt om» [..] «Jeg vidste fra starten, at det aldrig ville kunne blive til noget. Jeg havde gamle forældre, som jeg havde pligt til at passe, og en onkel i partieliten. Han kunne få store problemer, hvis jeg giftede mig med en mand i et kapitalistisk land» [..] «Jeg fik kvalme. Det var tydeligt, de havde læst vores breve. Jeg endte med at skrive til Jan, at vi måtte slutte. At vi skulle gå videre med vores liv» [..] «Jeg kunne ikke tage tilbage til min uddannelse, jeg havde intet job, jeg havde ansvar for to pensionister og en nyfødt søn. Det var noget nær en katastrofe» [..] «Så stod jeg der. Enlig far med to børn» [..] «Jeg måtte finde ud af, hvad jeg ville. Hvad jeg syntes var godt og skidt. Hvilke kriterier jeg havde for et forhold. Der dukkede Alma op igen og igen i tankerne» [..] «Alma, det er mig, kan du huske mig? Hvordan har du det?». Jeg havde tænkt: Gud, han hedder det samme som min Jan, men jeg havde aldrig regnet med, det var ham» [..] «Jeg landede sent. Så kom Alma ind. Det kunne lige så godt have været dagen efter, vi skiltes. Det var fuldstændig den samme følelse» [..] «Årene var jo gået, men relationen var på en måde helt den samme. Det var virkelig mærkeligt. Jeg gik over og gav hende et knus. Det føltes bare fuldstændig

³ Ved en fejl kom *mit* ikke med her, men denne form er 10 gange mindre hyppig end *min*, så man må formode at det ikke skævvrider undersøgelsen alt for meget.

rigtigt» [...] «Det viste sig jo, at det, jeg havde tænkt var så godt dengang, faktisk stadig var godt - stadig er godt. Det var ikke noget, jeg havde bildt mig ind» [...]

Avismaterialet (der er opmærket med tagger/parseren DanGram (Bick 2000), som bl.a. muliggør søgning efter syntaktiske roller) genopmærkes af hensyn til sammenligneligheden ligesom LANCHART-materialet med lemmaer og ordklasser i SOLIDUS-format. Alle ordformer transformeres til små bogstaver. Desuden fjernes al interpunktion. Rationalet er her at interpunktionstegn som komma og punktum, der er langt hyppigere end de hyppigste ordformer i skriftsprog, ville puste det totale antal ordformer væsentligt op i avismaterialet, hvilket ville gå ud over den statistiske sammenlignelighed af de to materialer. Interjektioner fjernes ikke, da de er meget sjældne i avismaterialet (som det også vil fremgå nedenfor: Ikke engang i sammenligning med de interjektionsrensede LANCHART-passager dukker interjektionerne op blandt signifikant overhyppige mønstre i avismaterialet).

Antallet af løbende ord i avismaterialet er på 71.817, fordelt på 370 forskellige tekster.

Kriterierne for udtræk af narrative passager fra aviserne er rent formelle og simple end de kriterier der har været anvendt på LANCHART-materialet, der er manuelt opmærket, bl.a. på baggrund af fremanalyseringen af de fem narrative strukturelementer indleder, orientering, handlingssekvens, evaluering og afslutning (DGCSS 2011: 36ff). Det kan ikke udelukkes at dette kan have betydning for resultaterne, men der har ikke i dette projekt været ressourcer til en manuel opmærkning af avismaterialet, og da de udtrukne passager generelt fremstår personligt fortællende, jf. ovenstående eksempel, er det min vurdering at det vil være svært at frembringe et materiale der er bedre egnet til det aktuelle formål uden et endog meget stort manuelt arbejde.

4 Metode

LANCHART- og avismaterialet vil forventeligt være ret forskellige mht. emner og tematikker. LANCHART-materialet kommer fra sociolingvistiske interviews, der pr. definition ikke har noget specifikt, forudbestemt emne og som følge deraf ofte vil være præget af en relativt smal vifte af nære og personlige, forholdsvis generelle tematikker, mens avismaterialet kommer fra interviews der tværtimod ofte vil have et klart emnemæssigt fokus med en vis aktualitet. Af den grund afholder jeg mig i denne undersøgelse fra at undersøge forskelle i ordforrådet. I stedet koncentrerer jeg mig som omtalt ovenfor om konstruktioner.

I min undersøgelse har jeg ladet mig inspirere af to beslægtede statistiske metoder. Den ene er en klassisk metode til fremfinding af ord eller udtryk der er særligt karakteristiske (dvs. statistisk signifikant overhyppige) i et korpus i sammenligning med et andet (Rayson & Garside 2000). Den anden er en metode udviklet til at finde konstruktionskandidater mhp. indlemmelse i konstruktikoner (Forsberg et al. 2014).

Korpussammenligningsmetoden går ud på at udregne *Log Likelihood*-statistik for alle udtryk (som kan være sekvenser af et eller flere tokens, lemmaer, ordklasser eller andet, alt efter hvordan undersøgelsesgenstanden er defineret) i to korpuser. *Log Likelihood* (i det følgende LogLik) er et mål for hvor overhyppigt et givet udtryk er i det ene korpus i forhold til det andet sammenlignet med hvordan en ligelig fordeling ville se ud givet de to korpusers størrelse og udtrykkets samlede hyppighed i dem. Der beregnes en p-værdi som angiver om overhyppigheden er statistisk signifikant.

Jeg bruger LogLik-metoden på hybride 4-grammer (jf. nedenfor) og filtrerer som første trin i analysen alle 4-grammer fra der ikke er signifikant overhyppige i enten det ene eller det andet delkorpus ved et signifikansniveau (p-værdi) på 0,05.

Konstruktikonmetoden benytter sig af det statistiske mål *Specific Correlation* (SC), som er en generalisering af et gængs mål for overhyppig samforekomst, *Pointwise Mutual Information* (PMI). SC bruges som et statistisk mål for kollokation og minder lidt om LogLik: SC måler hvor overhyppig samforekomsten af elementerne i et n-gram er i et korpus i forhold til hvad man skulle forvente ud fra hvor hyppige elementerne er i korpuset hver for sig.

Jeg bruger SC-målet i analysens andet trin til at rangordne de 4-grammer der er signifikant overhyppige i det ene eller i det andet delkorpus, således at de mest kollokerende 4-grammer kommer først i de lister jeg tager udgangspunkt i i analysens næste trin. Tabel 1 nedenfor viser toppen af den SC-sorterede liste over 4-grammer der er signifikant overhyppige i LANCHART. Bemærk at SC-værdierne i tabellen er negative. Det betyder at de enkelte mønstrelementer er negativt korrelerede med hinanden. Sorteringen på SC-værdien placerer således de *mindst negativt* korrelerede mønstre øverst, men det er nok til at faste udtryk placerer sig i toppen, fint illustreret ved tilstedeværelsen af udtrykket *i hvert fald* i de fire øverste mønstre.

Rang	Mønster	Eksempel	LANCHART N	SC	LogLik	P-værdi
1	i hvert fald ikke	i hvert fald ikke	25	-30,7	15,10	0,000
2	men i hvert fald	men i hvert fald	7	-31,2	6,51	0,011
3	skulle/kunne/.. i hvert fald	kunne i hvert fald	7	-31,5	6,51	0,011
4	var i hvert fald	var i hvert fald	11	-31,9	6,14	0,013

Tabel 1. Top 4 på den SC-sorterede liste over 4-grammer der er signifikant overhyppige i LANCHART. Listen har de stærkest kollokerende (egentlig: mindst negativt korrelerede) mønstre øverst.

Eftersom det ikke (kun) er faste udtryk, dvs. rent leksikalske n-grammer, der er i fokus i denne undersøgelse, men derimod (især) delvist skematiske konstruktioner med mindst én tomplads, laver jeg en desuden en alternativ rangordning af 4-gramlisterne der er inspireret af Forsberg et al. De mønstre der ligger til grund for min analyse, er hybride 4-grammer som “<ngn/navn> har faktisk svært” (delvis skematisk), “har faktisk svært ved” (rent leksikalsk) eller “N ADV i Nen” (næsten helt skematisk) (mere om dem nedenfor). De hybride n-grammer kan således være temmelig abstrakte. Det gør dem mere generelle og alt andet lige mere hyppige end mere konkrete n-grammer, hvilket har betydning for den alternative rangordning. Korrelationsmål som PMI og SC er kendt for at give sjældne n-grammer forrang. Ligesom Forsberg et al. vægter jeg derfor i den alternative rangordning SC-målet efter hvor varieret et givet mønster er manifesteret, for at promovere de mere abstrakte konstruktionskandidater. Det sker ved at gange SC med en variationsfaktor, *Unique Instance Frequency* (UIF). UIF er antallet af unikke token-n-grams et mønster er manifesteret som i materialet. Eftersom helt konkrete n-grammer kun er manifesteret som én unik ordfølge (nemlig n-grammet selv), skubbes disse n-grammer nedad i sorteringen ved denne manøvre, mens mere skematiske n-grammer med en højere UIF-værdi skubbes opad.

Tabel 2 nedenfor viser toppen af den SC·UIF-sorterede liste for LANCHART-materialet. SC-værdierne er generelt negative, jf. ovenfor. SC·UIF-værdien bliver derfor en større negativ værdi når SC ganges med UIF, og tabellen er sorteret derefter.

Rang	Mønster	Eksempel	LAN CH. N	SC	UIF	SC·UI F	Log- Lik	P
1	og så VERBede <ngn/navn>	og så aftalte vi	89	-42,0	78	-3279,7	40,4	0,000
2	en/et N og så	en bil og så	38	-42,4	38	-1611,8	20,1	0,000
3	N og sådan noget	afdeling og sådan noget	35	-39,1	35	-1368,3	28,6	0,000
4	så VERBede <ngn/navn> <ngn/navn>	så foreslog han jeg	30	-44,9	28	-1255,9	7,8	0,005

Tabel 2. Top 4 på den SC·UIF-sorterede liste over 4-grammer der er signifikant overhyppige i LANCHART. SC·UIF-sorteringen sørger for at promovere mere abstrakte mønstre.

Grunden til at det netop er 4-grammer jeg arbejder med, er for det første at jeg forsøger at identificere mønstre der er nogenlunde lange og ikke alt for leksikaliserede for ikke at gøre analysen for afhængig af indhold. Jo mere abstrakte (læs: grammatiske) mønstrene er, des mere vil evt. forskelle have at gøre med sprogstruktur og ikke de emner der tilfældigvis omtales i materialet. For det andet formoder jeg at længere mønstre vil være for sjældne til at give ret mange resultater. Når jeg afholder mig fra at afsøge hele spektret af fx 2-6-grammer, er det dog også for at holde arbejdet overskueligt.

Mine hybride 4-grammer fremkommer gennem en række abstraktioner af de rå token-n-grammer i materialet:

1. Genusforskellen mellem en/et og den/det opløses.
2. Modalverber i hhv. præsens og præteritum slås sammen i hver sin abstrakte kategori.
3. Visse meget hyppige funktionsord bevares som ordformer (*og, men, at, der, som, eller, fordi, hvad, hvilken, hvilket, hvis, nogen, noget*).
4. En række af de hyppigste substantiver, verber, adjektiver, adverbier og præpositioner bevares som ordformer, bl.a. for at enkelte meget hyppige ordformer ikke skal skævvride statistikken for de mere generelle, ordklassebaserede mønstre. Disse ordformer er udvalgt ud fra et hyppighedskriterie kombineret med et skønmæssigt specificitetskriterie: De hyppigste ordformer fra hver ordklasse blev bevaret, og efterhånden som ordformerne i hyppighedslisten begyndte at blive mindre og mindre hyppige og mere og mere betydnings-specifikke/konkrete (fx er *lufthavnen* mere specifik end det lige så hyppige *kroner*), stoppede udvælgelsen. Dette ud fra en intention om at bevare de mest faste elementer i

potentielle konstruktioner der udgør den ramme hvori mere frie indholdselementer kan placeres, og abstrahere fra de mest betydningstunge indholdselementer.

5. De resterende ordformer erstattes med ordklassekategorier.
6. Personlige pronomener og navne slås sammen til én kategori.

Disse operationer er foretaget ud fra mine egne, subjektive vurderinger af hvilke kategorier der ville kunne slås sammen uden i for høj grad at slå forskellige konstruktioner sammen. Det skal siges at der generelt er abstraheret med ret let hånd, således at der snarere er en bias mod for lidt abstraktion (→ for mange 4-grammer der er for konkrete/specifikke) end for meget, ud fra en antagelse om at det vil være nemmere visuelt at identificere specifikke konstruktioner der skal slås sammen end abstrakte konstruktioner der skal skilles ad. Samtidig er abstraktionsproceduren i sin nuværende form ganske ad hoc og kan med fordel gøres mere stringent i fremtidigt arbejde.

For at gøre ordklassekategorierne en smule mere læsevenlige, er de i denne artikel desuden blevet transformeret som følger:

V_PAST/skulle ville kunne måtte gide	→	skulle/kunne/..
V_PRESENT/skulle ville kunne måtte gide	→	skal/kan/..
V_PAST	→	VERBede
V_PARTICIPLE_PAST	→	VERBet
V_PRESENT	→	VERBer
V_INF	→	VERBe
N_INDEF_SING	→	N
N_DEF_SING	→	Nen
N_INDEF_PLU	→	Ner
N_DEF_PLU	→	Nerne
PRON_PERSON EGEN	→	<ngn/navn>

Ordformer i den endelige notation repræsenterer generelt tekstord, ikke lemmer.

Vejen fra SOLIDUS-opmærkede ordformer til mine transformerede repræsentationer kan anskueliggøres med et før-/efter-billede – bemærk hvordan også tegnsetningen fjernes som beskrevet ovenfor:

Jeg/PRON_PERSON/jeg har/V_PRESENT/have faktisk/ADJ/faktisk svært/ADV/svær ved/PRÆP/ved at/UNIK/at forstå/V_INF/forstå de/PRON_PERSON/den mennesker/N_INDEF_PLU/menneske ,TEGN/, der/UNIK/der kan/V_PRESENT/kunne

finde/V_INF/finde tid/N_INDEF_SING/tid til/PRÆP/til en/PRON_UBST/en iron-
man/N_INDEF_SING/ironman oven/ADV/oven i/PRÆP/i arbej-
det/N_DEF_SING/arbejde ./TEGN/.

→

<ngn/navn> har faktisk svært ved at forstå PRON_DEMO Ner der skal/kan/.. finde
tid til en/et N ADV i Nen

Disse forskellige procedurer resulterer i fire datasæt i form af lister af hybride n-grammer, som illustreret ovenfor: en liste med mere abstrakte mønstre og en med mere konkrete mønstre for hhv. LANCHART- og avismaterialet. Mere specifikt udvælges de 100 mest abstrakte mønstre målt på SC·UIF-værdi og de 100 mest konkrete mønstre målt på SC-værdi fra hvert materiale. Disse lister tager jeg udgangspunkt i i en kvalitativ/hermeneutisk analyse af de grammatiske mønstre der kommer til syne i de to forskellige sprogformer.

5 Resultater

I dette afsnit præsenterer jeg resultaterne af min undersøgelse af hvilke konstruktioner der er signifikant overhyppige i hhv. LANCHART- og avisnarrativer. Målet er som nævnt at belyse hvilke forskelle af mere sprogestrukturert art man kan finde ad den vej.

Først gør vi dog et lille ophold ved fordelingen af ordklasser i de to materialer. Henrichsen & Allwood (2005) viser to rangordninger af ordklasser baseret på forekomstfrekvens for hhv. dansk talesprog (DanSPO) og dansk (avis)skriftsprog (DanWRI):

DanWRI: Noun - Verb - Prep - Pro - Adj - Adv - Conj

DanSPO: Pro - Verb - Adv - Noun - Conj - Prep - Adj

Vi kan se på de tilsvarende fordelinger for det materiale jeg undersøger her, for at se for det første om det her undersøgte, narrative talesprog ligner Henrichsen & Allwoods mere generelle talesprog, og for det andet i hvor høj grad aviscitaterne ligner LANCHART-narrativerne på dette punkt.

De ti hyppigste ordklasser i LANCHART-materialet fordeler sig procentvis sådan (summen er ikke 100 da opmærkningen indeholder nogle yderligere, små ordklasser. Ordklassen UNIK dækker infinitiv-*at*, *som* og *der*):

PRON	V	ADV	N	PRÆP	SKONJ	ADJ	UNIK	UKONJ	EGEN
22,3	21,3	18,9	9	7,2	6,5	5,4	4,1	2,7	1,1

Som det ses, ligner rækkefølgen DanSPO til forveksling. Og nu avismaterialet:

V	PRON	N	ADV	PRÆP	ADJ	SKONJ	UNIK	UKONJ	EGEN
22,2	19,8	14,4	10,2	9,9	7,4	4,3	3,9	3,3	3,2

Det er tydeligt at rækkefølgen ligner fordelingen for talesprog på flere punkter. Substantiverne, der ellers udgør klart den hyppigste ordklasse i skriftsprog og kan være over dobbelt så hyppige i skriftsprog som i talesprog (Biber 1991: 247ff, Henrichsen 2002: 10), er her rykket ned på en tredjeplads, omend de ikke er nær så sjældne som i LANCHART-passagerne. De to hyppigste ordklasser i avismaterialet er de samme som i LANCHART-materialet; adverbier er langt hyppigere end hos Henrichsen og Allwood og er endda hyppigere end præpositioner; de tre lavest rangerede ordklasser er de samme som i LANCHART-materialet. Som helhed er fordelingerne dog signifikant forskellige ($\text{Chi}^2 = 4285$, $\text{df} = 8$, $p < 0.01$), og hver enkelt ordklasse har også signifikant forskellig hyppighed i de to materialer (kun UNIK nærmer sig ikke-signifikans med $p = 0,039$).

Samtidig ligger de to sprogformer tættere på hinanden hvad leksikalsk diversitet angår end talesprog og skriftsprog gør hos Henrichsen (2002). Her rapporteres det at skriftsprog gennemsnitligt har under 13 forekomster pr. ordform mens talesprog gennemsnitligt har 38. For aviscitaterne er det tilsvarende tal knap 8 forekomster pr. ordform, og for LANCHART-materialet knap 14. Man skal her være opmærksom på at interjektioner er fjernet fra LANCHART-materialet og interpunktion er fjernet fra avismaterialet, hvilket formentlig er en del af forklaringen på at tallene er noget anderledes. Ikke desto mindre er det tydeligt at tallene i denne undersøgelse ligger noget tættere.

Med disse resultater in mente går vi nu videre til konstruktionerne i forvisning om at de to materialer er så sammenlignelige som praktisk muligt.

De mest sværtindkredselige og derfor interessante konstruktioner er dem med (evt. adskillige) tompladser. Derfor begynder vi med den abstrakte LANCHART-liste, dvs. den der er rangordnet efter SC·UIF. Hvad finder man blandt de stærkest kollokerende og samtidig mest varieret realiserede (og dermed mest skematiske) mønstre der er signifikant overhyppige i LANCHART-materialet?

5.1 Karakteristiske mønstre i LANCHART-materialet

5.1.1 Og så ...

Blandt de første 100 mønstre af den nævnte type er der ikke mindre end 27 der indeholder ordfølgen “og så”. Blandt de tilsvarende 100 avis-overhyppige mønstre er antallet 0. Et af de absolut mest overhyppige mønstre i LANCHART-materialet er således “og så VERBede <ngn/navn>” (fx realiseret som *og så aftalte vi*). Det er jo ikke i sig selv en komplet konstruktion, selvfølgelig bl.a. fordi der kun er plads til meget korte sætninger i et 4-gram. Mange andre mønstre med “og så” er endnu mindre konstruktionsagtige (dvs. både ukomplette og derudover også ‘skæve’ i forhold til den syntaktiske struktur, som fx “N og så VERBede” (fx *ballade og så smed*)). Ikke desto mindre er det bemærkelsesværdigt med så betydelig en overrepræsentation eftersom det narrative temporale *så* er så oplagt et udtryk for den kronologiske virkelighedsgengivelse der er definerende for narrativgenren (jf. Skytte (1999), der finder en kraftig overrepræsentation af *og så* i mundtlige vs. skriftlige genfortællinger af en mr. Bean-video). Man kunne derfor have forventet en i det mindste lidt hyppigere forekomst i avismaterialet. En nærmere undersøgelse af avismaterialet også udenfor de 100 statistisk set mest oplagte mønstre viser at “og så” trods alt forekommer, bare mindre hyppigt (1852/mio. i avismaterialet mod 13236/mio. i LANCHART-materialet, jf. Tabel 3).

Tabel 3. Overhyppigheden af “og så” i LANCHART-materialet (stærkt signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
og så	1393	13236	133	1852	786,01	0

Interessant nok kan man, hvis man går tilbage til det ubearbejdede korpus, se at ca. 35 % af avisforekomsterne har *Og så* efter punktum, et virkemiddel der er forbundet med en vis berøringsangst (bemærket af Dansk Sprognævn allerede i 1975 (Bjerrum 1975)). Punktummet virker som en ekstra tydelig markering af en given “og så”-sætning, og det bruges da også ofte i avismaterialet til at skabe en dramatisk, konkluderende eller anden effekt der passer godt til narrativgenren:

- (1) Min mand, vores børn og jeg flygtede ud i bushen. **Og så** begyndte sulten.
- (2) Jeg fik talt med min søns sagsbehandler, og vi fik lov til i første omgang at prøve den i tre måneder. **Og så** virkede det bare.

- (3) Det var som en salon, hvor de diskuterede alle mulige tanker. Men nogle gange ventede der nogle bøller uden for, som ville genere dem. **Og så** måtte jeg gå ud og rydde vejen.

5.1.2 Og sådan noget

En anden ordfølge der går igen mange gange i LANCHART-materialet, er “og/eller sådan noget”. Den indgår i 4 af de 100 mønstre på den skematiske LANCHART-liste. I LANCHART-korpusset forekommer den oftest efter substantiv i ubest. sing./plur., og i øvrigt desuden i varianten “og så noget”. I avismaterialet indgår den ikke i nogen signifikante mønstre overhovedet, heller ikke uden for de 100 mønstre på den skematiske avisliste. I LANCHART-delkorpusset er frekvensen af frasen 1777/mio., i avismaterialet 70/mio. Frasen er et eksempel på det der af Christensen og Jensen (2018) er beskrevet under betegnelsen generelle udvidere.

Her er der i højere grad tale om noget der ligner en konstruktion, omend en meget lidt skematisk en af slagsen. Grunden til at den dukker op blandt de statistisk set mest skematiske, er at dens tre elementer forekommer sammen med diverse andet materiale på den resterende plads, hvilket muliggør variation i realiseringen af ‘værts’-4-grammerne. Frasen er intuitivt udpræget talesproglig, men dette træk bruges altså sjældent i repræsentationen af tale i aviskorpusset. Her dog adskillige gange inden for samme citat (hvormed den samlede pulje af forekomster allerede er næsten opbrugt):

- (4) Jeg tror ikke, de forventede noget særligt af mig. Ikke andet, end at jeg behandlede hende godt. De er ikke religiøse **og sådan noget**. Moderen er gammel kommunist. Men der er selvfølgelig mange i Cuba, der tror på spøgelse **og sådan noget**. Det er hun massiv modstander af, moderen. Hun griner ad det, når de slagter en høne og smider med perlekæder **og sådan noget**.

5.1.3 Den her ...

En anden struktur der både er (over)hyppig og varieret nok i LANCHART-dataene til at optræde blandt de ovennævnte 100 mønstre, er demonstrativen “den her N”, der dukker op i mønsteret “den/det her N og”. Strukturen har flere forskellige funktioner, herunder en mere idiomatisk med ekstra betydning ud over den almindelige syntaktisk-kombinatoriske (jf. også Therkelsen 2002 og 2003). Mere herom straks.

Undersøger man antallet af forekomster i LANCHART-materialet og avismaterialet af en let udvidet variant af strukturen med optionelt adjektiv (altså “den/det

her (ADJ N”)), viser det sig at den faktisk ikke er signifikant overhyppig i nogen af de to delkorporer, jf. Tabel 4.

Tabel 4. Overhyppigheden af “det/den her (ADJ) N” i LANCHART-materialet (ikke-signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
det/den her (ADJ) N	47	447	25	348	1,02	0,313

En nærmere undersøgelse viser dog at der er stor forskel på fordelingen af forekomster på forskellige funktioner i de to delkorporer.

I begge delkorporer bruges strukturen på tre forskellige måder (jf. også Von Heusinger (2011)):

1. **Anafor.** Tilbagereference til noget tidligere nævnt der ikke er til stede i de fysiske omgivelser. Eksempel: *ja så var de ringforlovet og skulle giftes sidste år her i november i Chicago og så forærede vi så min mor en flybillet til **det her bryllup i Chicago** ikke.*
2. **Demonstrativ.** Deiktisk reference til noget i talesituationen – eller (oftest) noget taleren er/har været involveret i. Eksempel: *Vi har fundet en bedre balance og genopfundet kærligheden til **det her sted**, også inden vi fik tre stjerner i Michelin-guiden.*
3. **Specifik indefinit.** Den mest idiomatiske funktion fra et traditionelt grammatisk synspunkt. Her er tale om reference til noget som er bekendt for taleren og samtidig forudsættes ubekendt for lytteren – svarende til den specifikke læsning af en indefinit nominalfrase (en funktion der if. Hansen og Heltoft (2011, s. 178f, 472f) ikke er grammatikaliseret i dansk; se endvidere Von Heusinger (2011) om tilsvarende konstruktioner på tysk med henvisninger til velkendte beskrivelser af den tilsvarende indefinitte *this*-konstruktion på engelsk). Eksempel: *ej det var lidt anstrengende men **øh det værste det var nok den her senile dame** som man også skulle dele toiletet med.*

De tre funktioner fordeler sig som vist i Tabel 5 nedenfor.

Tabel 5. Fordelingen på grammatiske funktioner af “det/den her (ADJ) N” i LANCHART-materialet og avismaterialet.

	LANCHART	%	AVISER	%
Anafor	30	67	2	8
Demonstrativ	4	9	19	76
Specifik indefinit	11	24	4	16
Total	45	100	25	100

Der er klar overvægt til den anafore funktion i LANCHART-materialet. Strukturen “det/den her N” bruges her ofte til at genoptage referencer der har været i spil i samtalen for kortere eller længere tid siden, og kun sjældent demonstrativt. I avismaterialet bruges den derimod langt overvejende deiktisk/demonstrativt om noget taleren er involveret i (foruden et par deiktiske tidsangivelser: to forekomster af *den her uge*). En nærliggende forklaring på overvægten af anaforer i LANCHART er at den anafore funktion i samtaler udgør en vital, diskursstrukturerende mekanisme der ikke i samme omfang er nødvendig i redigeret tekst – en genorienterende funktion der kan modvirke at tilhøreren bliver hægtet af på diverse sidespor i fortællingen. Overvægten af den ‘involveringsdemonstrative’ funktion i avistekst skyldes formentlig at noget en bestemt person er involveret i (fx en Michelinrestaurant), ofte vil være selve anledningen til at de bliver interviewet, hvorimod sociolingvистiske interviews pr. definition ikke har et specifikt, forudbestemt emne (Labov 1972: 113ff.; Gregersen 2015: 224f.).

Fordelingen i LANCHART er altså forskellig fra fordelingen i avismaterialet, og forskellen er statistisk signifikant ($\text{Chi}^2 = 34,7$; $\text{df} = 2$; $p\text{-værdi} < 0,0001$).

Interessant nok bidrager fordelingen af specifikke indefinitter ikke til denne skævhed. Der er procentuelt lidt flere forekomster i LANCHART-materialet, men denne overvægt er ikke statistisk signifikant. Netop denne konstruktion er altså ikke skævfordelt mellem de to delkorporer. Således er der altså her en konstruktion som de to repræsentationer af narrativt talesprog så at sige er enige om – frasen den væsentlige detalje at mønsteret “den/det her (ADJ) N” generelt forekommer noget sjældnere i avismaterialet. I lyset af hvor lidt mønsteret i øvrigt forekommer i avistekst, er det tydeligt at det i høj grad er markør for talesproglig stil: Leder man efter “den/det her (ADJ) N” uden for citeret talesprog i aviserne, leder man næsten forgæves. Der findes enkelte eksempler, men de forekommer næsten uden undtagelse i klummer eller essayistiske beretninger med jeg-fortæller (jf. fx “Derfor starter *den her historie* med den lidt irriterende situation, der fandt sted, da jeg tog i døren til

et af de nyopførte offentlige toiletter på Nørreport [...]”). (Det samme gælder i øvrigt også den mere skriftsprogligt-formelle variant “denne/dette her (ADJ) N”). Det eneste ikke-talesproglige eksempel jeg har været i stand til at grave frem i hele POLITIKEN2016-korpusset, er denne anafor i en filmomtale:

- (5) Solen hænger lavt over forstæderne den efterårs morgen, hvor Rachel ser noget mystisk fra sin faste plads i pendlertoget mod New York. [...] Men **den her morgen** er noget forkert. I det korte øjeblik, toget venter på signal, ser Rachel noget, hun ikke skulle have set. [...]

Ved første øjekast og som rent syntaktisk struktur betragtet ser “den/det her (ADJ) N” altså ud til at være lige karakteristisk for narrativt talesprog i de sociolingvistiske interviews og i avisartiklerne. Som konstruktioner betragtet (dvs. inklusive strukturens forskellige funktioner) er der imidlertid en klar opdeling af hvad der er karakteristisk for hvilken sprogform – bortset fra specifik indefinit-konstruktionen, der faktisk forekommer i citeret narrativ tale noget nær lige så hyppigt som i ikke-redigerede, udskrevne narrativer. Det står i skarp kontrast til de andre mønstre der er omtalt ovenfor, hvor der er op til flere størrelsesordeners forskel på hyppigheden i de to delkorpusser. En del af grunden til denne jævnbyrdighed er formentlig at specifik indefinit-konstruktionen i modsætning til de to andre ikke opfylder en funktion der er dikteret af genre, medie, kommunikationsmodus osv., men derimod en funktion der er uafhængig af disse faktorer, nemlig en grammatisk: at skelne specifik fra ikke-specifik indefinit reference.

5.1.4 Sådan en ...

En anden demonstrativ indefinit konstruktion er det indefinitte nominalsyntaxme med prædeterminativt *sådan* (“sådan en/et (ADJ) N (der)”) (Jensen 2013; Pedersen 2014). Også den er særligt karakteristisk for LANCHART-materialet. Den er således knap seks gange hyppigere i LANCHART-materialet end i avismaterialet (hhv. 1378/mio. og 237/mio.). Konstruktionen er altså ikke helt fraværende i avismaterialet, men forekommer dog fx aldrig med det efterhængte adverbelle *der*, der, når det optræder, er med til at gøre konstruktionen ekstra talesproglig.

Når jeg bruger betegnelsen “demonstrativ indefinit”, skyldes det Von Heusinger (2011), der peger på slægtskabet på tysk mellem “dies N” (jf. “den her N”) og “so’n N” (jf. “sådan en N”), både i funktionen som specifikke indefinitte demonstrativer (med taler-bekendt og lytter-ubekendt reference) og i andre funktioner (med både taler- og lytter-bekendt reference). Von Heusinger opstiller den hypotese at sådanne

demonstrativer i de tilfælde hvor der ikke findes en for lytteren tilgængelig referent, har som funktion at få lytteren til at være ekstra opmærksom på denne referent – ikke som den allerede fremgår af konteksten (for det gør den ikke), men som den kommer til at fremgå af det umiddelbart efterfølgende diskursforløb. Hypotesen binder specifikke indefinitte demonstrativer sammen med mere traditionelle demonstrativer via en fælles opmærksomhedsfokuserende/-orienterende funktion og giver samtidig en grammatikaliseringsbane de kan have fulgt fra deiktisk til indefinit funktion. Der er desværre ikke plads til at udfolde disse referentielle slægtskabsrelationer yderligere på dette sted (men se Skafte Jensen 2013, hvis betydningsbeskrivelse af *sådan* (inkl. i den demonstrative indefinitte konstruktion) også tager udgangspunkt i det henvisende).

5.1.5 Andre observationer i delvis skematiske mønstre i LANCHART

Man kan gøre forskellige andre observationer med udgangspunkt i de 100 bedst kollokerende mønstre med mindst én tomplads. Jeg vil her bare give et kort gennemgang uden at gå ind i detaljerede analyser.

Lige som “den her ...” er også “den der ...” stærkt repræsenteret. (Betydeligt stærkere end “den her ...”, faktisk). Ved at prøve sig lidt frem med forskellige søgemønstre kan man komme frem til en specifik konstruktion der er meget signifikant overrepræsenteret i LANCHART-materialet: “den/det der (ADJ) N (der)”. Konstruktionen er over 20 gange hyppigere i LANCHART-materialet end i avis-materialet, jf. Tabel 6 nedenfor. Eksempler med det efterhængte *der* forekommer slet ikke i avismaterialet.

Tabel 6. Overhyppigheden af “den/det der (ADJ) N (der)” i LANCHART-materialet (stærkt signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
den/det der (ADJ) N (der)	94	893	3	42	76,04	0

Man kan også spore den overhyppighed af forfeltsfordobling, der af Galberg Jacobsen og Skyum-Nielsen (2007) nævnes som et typisk talesprogstræk. Trækket dukker op i mønsteret “og PRON_POS N <ngn/navn>”, som peger på ordfølger såsom (i hyppighedsrækkefølge) “min mor hun”, “min far han”, “min mand han” osv. (“min kone hun” forekommer pudsigt nok slet ikke) – altså korte possessive nominalsyntagmer og deres genoptagende subjektspronomen.

Endelig spøger gambitter i form af “kontaktpåkaldelser”, der også nævnes som typisk talesproglige af Galberg Jacobsen og Skyum-Nielsen (2007), også i LANCHART-materialet. Der er 13 mønstre blandt de nævnte 100 der indeholder *ikke*, og lige over halvdelen af dem har *ikke* i en position hvor det ikke kan fungere som sætningsadverbium og altså snarere fungerer som kontaktpåkaldelse. (Og dermed egentlig burde have været opmærket som interjektion, hvilket, måske ikke helt overraskende, har været hinsides opmærkningsalgoritmens formåen).

5.1.6 Konkrete mønstre i LANCHART

Efter at have analyseret de mere skematiske mønstre i LANCHART ser vi nu på de mere konkrete – dvs. de 100 mønstre der står øverst når de sorteres efter SC-værdien uden at vægte efter UIF.

Det absolut mest iøjefaldende konkrete mønster i LANCHART er adverbialt “i hvert fald”. Det forekommer i adskillige forskellige mønstre i toppen af den SC-sorterede liste, og det er knap 8 gange hyppigere i LANCHART-materialet end i avismaterialet, jf. Tabel 7.

Tabel 7. Overhyppigheden af “i hvert fald” i LANCHART-materialet (stærkt signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
i hvert fald	89	846	8	111	51,56	0

De generelle udvidere fylder også meget i denne sortering (22/100 mønstre, hvoraf mange dog overlapper og/eller er det samme mønster, bare med forskellige ‘ekstra’ elementer i 4-grammet, fx “faren og sådan noget”, “og sådan noget og”, “sådan noget og så”; “eller et eller andet”, “eller noget som helst”, “og alt det der”), det samme gælder *så*-sætninger (20/100 mønstre), og demonstrativ indefinit reference med “sådan en/et” fylder også noget (6/100 mønstre). Mønstre med “kan (..) (ikke) huske” forekommer også i forskellige variationer (6/100 mønstre), et fænomen der nok dels afspejler den situation det sociolingvistiske interview forsøger at skabe, nemlig en intim stemning med så vidt muligt personlig samtale, og dels det faktum at det jo netop er selvoplevede, fortidige begivenheder der tales om qua materialets narrative genre.

Et mønster der er ekstremt overhyppigt i LANCHART-materialet, er den udvidede konjunktion “men altså”, jf. Tabel 8.

Tabel 8. Overhyppigheden af “men altså” i LANCHART-materialet (stærkt signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
men altså	111	1055	1	14	105,21	0

Og konstruktionens enlige forekomst i avismaterialet forekommer interessant nok igen efter punktum, og tilmed med kolon efter:

- (6) Vi var den nye generation, vi var ikke bange, og vi rædede igennem på arrogancens høje bølge. **Men altså:** En selvbevidsthed om, at man kan noget, er ikke forbudt, synes jeg.

Det kan undre at et så karakteristisk talesprogstræk, der på den ene side ikke fylder meget, men på den anden side formentlig vil kunne gøre en stor forskel for oplevelsen af talesproglighed, ikke får lov til at fylde mere i de citerede passager (og når det endelig sker, så kun med en vis interpunktionel pomp).

To mønstre indfanger fraserne “hvad hedder det” og “hvad er det nu det/den hedder” (jf. Clausen & Pedersen 2017). Ingen af dem forekommer nogensinde i avismaterialet, og de er da også rene tomgangskonstruktioner, dvs. i virkeligheden bare fyldte pauser der helt bogstaveligt er udtryk for at den talende leder efter ord. Med pronomenet *den* (en enkelt forekomst) i “hvad er det nu den hedder”, er vendingen dog ikke helt semantisk tom, for så er der nødt til at være en konkret referent som taleren ikke kan huske navnet på eller ordet for (i det relevante belæg en by: *og så har vi været i Avignon . og . Garpe . og . hvad er det nu den hedder . en eller anden anden by*), hvilket ikke er nødvendigt når pronomenet er *det*, jf. fx *jeg havde jo heller ikke været i militæret hvad er det nu det hedder øh men jeg søgte faktisk frivilligt ind*. I samme kategori er i øvrigt også “hvad skal man sige”, der heller ikke forekommer i avismaterialet. Se Tabel 9.

Tabel 9. Overhyppigheden af tomgangsfraserne “hvad er det nu det/den/de hedder”, “hvad hedder det” og “hvad skal man sige” i LANCHART-materialet (alle signifikante).

Mønster	LANCH.	ppm	AVISER	ppm	Log-Lik	P-værdi
hvad er det nu det/den/de hedder	11	105	0	0	10,59	0,001
hvad hedder det	33	314	0	0	33,26	0,000
hvad skal man sige	7	67	0	0	6,51	0,011

Endelig er der mønstrene “den/det var sådan lidt” og “den/det er bare sådan”, der rummer en talesprogs karakteristisk konstruktion. Udover den kanoniske grammatiske brug, som fx i *det var sådan lidt skægt* og *det er bare sådan en lille detalje*, har de to mønstre også en funktion som en slags inkvit (‘quotative’, jf. Rathje 2011). Således viser de to udtryk i de nedenstående eksempler at det følgende er et citat, der desuden ofte fremføres med en eller anden form for særmarkeret udtale (her markeret med anførelstegn, der ikke hører til den oprindelige transskription):

- (7) det var sådan lidt .. “nå okay”
 (8) det er bare sådan .. “hvem er det”

Kernen i konstruktionen er “**være** .. sådan <cit>”, med ret vide muligheder for variation, således med *jeg* eller *det* som subjekt, fleksibilitet ift. tempus, med eller uden *bare* eller andre adverbier, med eller uden *lidt*. (Den nyere variant med efterhængt *der* er der ingen forekomster af i materialet, fx *jeg var sån der “jeg er feminist”* (Jensen 2019, Hansen 2020)). En søgning på mønsteret “det **være** (0-2 ord) sådan ..” giver, efter frafiltrering af resultater hvor *sådan* er en del af en nominalhelhed, i alt 107 forekomster i de to korpusser, med hhv. 15 og 92 forekomster i aviserne og LANCHART. I LANCHART er 13 % af forekomsterne inkvitkonstruktioner. I aviserne er antallet 0, omend formuleringen *Det var meget sådan værsgo-at-tage-stilling-til-det-agtigt uden nogen form for hjælp* kommer tæt på at tælle med. Uanset hvad er forskellen signifikant. I aviserne fungerer *sådan* som rene mådesadverbialer i 60 % af forekomsterne, med betydningen ‘på den måde’ som i *det var sådan, det hang sammen*, mens det i LANCHART-materialet kun er 10 %. I LANCHART-materialet er en indforstået (“anamnestisk”, jf. Skafte Jensen 2013) og attributiv-modificerende funktion som i *det var sådan lidt skægt* meget mere

fremherskende, så funktionen af *sådan* er i det hele taget væsentligt forskellig i de to materialer.

5.2 Karakteristiske mønstre i avismaterialet

Efter at have identificeret et antal konstruktioner og andre mønstre der er signifikant overhyppige i LANCHART-materialet, skifter vi nu perspektiv og ser på mønstre der er signifikant overhyppige i avismaterialet. Det følgende er altså baseret på de to top 100-liste af mønstre fra avismaterialet, dvs. top 100 på den mere abstrakte/skematiske liste baseret på rangordning efter SC·UIF og top 100 på den mere konkrete liste baseret på rangordning efter SC alene (dvs. to forskellige rangordninger af den samme, overordnede liste over alle mønstre der er signifikant overhyppige i avismaterialet).

Det mest iøjefaldende ved de mest skematiske mønstre er forekomsten af verbale komplement, dvs. ledsætninger og infinitiver. 43 af de første 100 mønstre indikerer forekomsten af et sådant led, fx “at <ngn/navn> skulle/kunne/.. VERBe”, “en/et ADJ N <ngn/navn>” (fx *en træls oplevelse vi [..]*), “<ngn/navn> ikke skulle/kunne/.. VERBe”, “<ngn/navn> til at VERBe” osv. osv. Det bekræfter Galberg Jacobsen & Skyum-Nielsens (2007) observation af at talesprog foretrækker (korte) helsætninger i stedet for ledsætninger, jf. også Jensen (2003), der i sit sætningsmateriale finder flest simple sætninger og desuden næsten ingen komplekse sætninger med mere end en ledsætning, i hvert fald så længe domænet er en enkelt intonationskontur. Samtidig er det dog vigtigt at bemærke hvor tæt de to sprogformer der er undersøgt her, faktisk ser ud til at ligge på hinanden hvad angår brugen af verbale komplement. Forekomsten af underordningskonjunktioner er således kun en smule højere i avismaterialet end i LANCHART-materialet, jf. ordklassefordelingen ovenfor. Og forekomsten af ordklassen UNIK, der dækker infinitiv-*at*, *som* og *der*, er næsten ens.

Det slører dog billedet lidt at både *som* og *der* har forskellige funktioner og ikke altid er underordnende; den underordnende funktion af *som* og *der* er formentlig reelt en smule hyppigere i avismaterialet end i LANCHART-materialet. Desuden er underordningskonjunktioner jo ikke altid obligatoriske i ledsætninger.

Disse usikkerheder fortjener at blive belyst lidt bedre. Jeg tog derfor en stikprøve af 50 helsætninger ud af hvert korpus og undersøgte ledsætningerne i dem. Tabel 10 nedenfor viser fordelingen af helsætninger med og uden ledsætninger, og Tabel 11 forskellen på antallet af verbale komplementer (ledsætninger og infinitiver) i de to stikprøver (hhv. totalt og pr. kompleks helsætning, dvs. med et eller flere verbale komplement). I ingen af fordelingerne er forskellen signifikant ($\text{Chi}^2 = 0,16$, $\text{df} =$

1, $p = 0,69$; $t = 0,20$, $df = 46,97$, $p = 0,84$) – omend forskellene i begge tilfælde peger i retning af lidt flere verbale komplementer i aviserne.

Tabel 10. Andelen af helsætninger med og uden ledsætninger (kun ledsætninger, ikke infinitiver) i LANCHART- og avismaterialet. Forskellen på LANCHART- og avismaterialet er ikke signifikant.

	LANCHART	%	Aviser	%
Helsætninger uden ledsætning(er) i stikprøve	30	60	27	54
Helsætninger med ledsætning(er) i stikprøve	20	40	23	46

Tabel 11. Andelen af verbale komplementer (ledsætninger og infinitiver) i LANCHART- og avismaterialet, herunder pr. kompleks helsætning. Forskellen på LANCHART- og avismaterialet er ikke signifikant.

	LANCHART	Aviser
Verbale komplementer i stikprøve	32	38
Procent verbale komplementer i stikprøve	64	76
Verbale komplementer pr. kompleks helsætning	1,41	1,45

Samtidig peger forskellen på fordelingen af ledsætninger med og uden konjunkional (som dog ikke er signifikant: $\text{Chi}^2 = 3,02$, $df = 1$, $p = 0,08$, jf. Tabel 12) i retning af flere eksplicitte konjunktioner i aviserne, hvilket kan være med til at give indtryk af en lidt større forskel i syntaktisk kompleksitet end der faktisk er belæg for. I parentes bemærket er der også i undergruppen af substantiviske ledsætninger flest eksplicitte konjunktioner (i denne stikprøve udelukkende *at*'er) i avismaterialet (58 % mod 47 % i LANCHART-materialet), omend forskellen i stikprøven er mindre end hos Boye et al. (2012) (og ikke-signifikant: $\text{Chi}^2 = 0,05$, $df = 1$, $p = 0,83$).

Tabel 12. Andelen af ledsætninger med og uden konjunkional i LANCHART- og avismaterialet. Forskellen på LANCHART- og avismaterialet er ikke-signifikant på trods af at den kan synes markant.

	LANCHART	%	Aviser	%
Ledsætninger uden konj. i stikprøve	7	35	2	9
Ledsætninger med konj. i stikprøve	13	65	21	91

Det er altså til syvende og sidst en ret lille forskel der kommer til udtryk i de mange overhyppige mønstre med verbale komplementer.

En anden ting der springer i øjnene, er forekomsten af adjektiver. Adskillige af de mest overhyppige mønstre indeholder adjektiver. Det er ikke helt overraskende da adjektiver generelt er lidt hyppigere i avismaterialet, jf. ovenfor. Overhyppigheden af mønstre med adjektiver er dog ikke udtryk for forekomsten af egentlige konstruktioner der er særegne for aviscitaterne, men snarere for fænomener af mere stilistisk art inden for de almindelige grammatiske rammer. Dels er der sideordningen af adjektiver i mønsteret “ADJ og ADJ N”, fx *åbent og dejligt menneske, bæredygtig og permanent månelandsby* osv., som er et af de kraftigst overrepræsenterede mønstre i avismaterialet, og dels sideordnede adjektiver uden konjunktion, der også er overhyppigt forekommende i avismaterialet: “en/et ADJ ADJ N”, fx *en varm broderlig interesse* osv. Nominalhelhederne i avismaterialet kan således karakteriseres som en smule mere syntetiske, dvs. præget af forvægt, hvilket især i sine ekstreme fremtrædelsesformer forbindes med formelt skriftsprog (jf. Galberg Jacobsen & Skyum-Nielsen 2007: 138).

Desuden er der visse forskelle på brugen af konstruktioner med *være*, som der desværre ikke er plads til at undersøge til bunds her. Dog er det et faktum at visse *være*-konstruktioner er signifikant overhyppige i avismaterialet, mens andre er signifikant overhyppige i LANCHART-materialet. I avismaterialet således fx *være* + indefinit subjektspredikativ i det udvidede mønster “være (ADV) en/et (ADJ+) N” (fx *Danmark var også et rigt samfund; den er et kubistisk virvar*; osv.) og *være* + adjektiv (som i *åh det ville være dejligt; man var velkommen*; osv.). I LANCHART-materialet fx spatiale konstruktioner med *være* + “(der)inde/ude/nede/oppe/ovre/hen-ne” og eksistentielle *der*-konstruktioner.

Blandt de konkrete mønstre er der et lille antal kollokerende mønstre der er overhyppige i avismaterialet, men flere af dem viser sig at være artefakter af andre effekter. Det komplekse verbum “være sammen med” er fx signifikant overhyppigt – men kun i infinitiv. De finite former er derimod overhyppige i LANCHART-materialet. Det viser sig altså at det der i virkeligheden er på spil her, snarere er forskellen på forekomsten af verbale komplementer såsom infinitiver. Noget lignende gør sig gældende for mønsteret “klar over at” – det er overhyppigt i avismaterialet, men kun med *at*’et. Forbindelsen “klar over” alene forekommer nogenlunde lige hyppigt i de to delkorporer. På lignende vis er det komplement-indikerende mønster “jeg kan huske at” overhyppigt i avismaterialet, mens lemmaet *huske* er klart overhyppigt i LANCHART, jf. Tabel 13.

Tabel 13. Overhyppigheden af lemmaet **huske** i LANCHART-materialet (stærkt signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
huske	183	1739	52	724	35,77	0

Blandt de relativt konkrete, avisoverhyppige mønstre findes også konstruktionen “for NUM år siden”, jf. Tabel 14.

Tabel 14. Overhyppigheden af “for NUM år siden” i avismaterialet (signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	LogLik	P-værdi
for NUM år siden	6	57	14	195	-7,05	0,008

En nærmere undersøgelse i korpusset viser at forskellen bliver noget mindre (omend stadig signifikant) når kvantorerne *mange*, *nogle* og *flere* tages med i betragtning, jf. tabel 15 nedenfor. Der viser sig således en præference i avismaterialet for præcise frem for vage tidsangivelser, hvilket formentlig afspejler de to interviewtypers natur, fx mht. hvor specifikke emner der tales om.

Tabel 15. Overhyppigheden af “for NUM/mange/nogle/flere år siden” i avismaterialet (signifikant).

Mønster	LANCHART	ppm	AVISER	ppm	Log-Lik	P-værdi
for NUM/mange/nogle/flere år siden	13	124	20	278	- 5,37	0,02

Den sidste ting i denne del af materialet der kan vrides en lille, interessant forskel ud af, er infinitiven “at gå i N”. Den er der flest forekomster af i avismaterialet, og det viser sig at være fordi den figurative brug af dette mønster er ikke-eksisterende i LANCHART-materialet. Udtrykkene “at gå i dialog”, “at gå i opløsning” og “at gå i panik” forekommer således kun i avismaterialet, jf. Tabel 16.

Tabel 16. Forekomsten af “at gå i N” i LANCHART- og avismaterialet. I figurativ betydning forekommer mønsteret kun i avismaterialet.

Ordfølge	LANCHART	Aviser
at gå i klasse	1	0
at gå i sele	1	0
at gå i skole	1	4
at gå i synagoge	1	0
at gå i dialog	0	1
at gå i kirke	0	1
at gå i opløsning	0	1
at gå i panik	0	1
at gå i søndagsskole	0	1

Den konkrete “gå i N”-konstruktion som helhed (dvs. med en styrelse med konkret betydning af *i* og med alle former af lemmaet *gå* og dermed ikke nødvendigvis med det indledende infinitiv-*at*) er dog præcis lige hyppig i de to delkorporer. Foruden en potentielt større brug af figurativ betydning i avismaterialet, synes det altså igen især at være forekomsten af verbale komplementer (såsom infinitiver) der spiller en rolle.

Alt i alt er der altså ikke lige så mange iøjefaldende, karakteristiske mønstre i avismaterialet som i LANCHART-materialet. Der er ikke rigtig nogen konstruktioner der er særligt karakteristiske for avisartiklerne – men dog en tendens til en smule mere forvægt i nominalhelhederne pga. sideordnede adjektiver. Derudover er og bliver det særlige den signifikante, men bemærkelsesværdigt lille overhyppighed af verbale komplementer.

5.3 Ikke-signifikante mønstre

Til sidst kaster vi et blik på de mønstre der ikke er signifikant overhyppige i nogen af de to delkorporer. Det er altså mønstre der fordeler sig mere eller mindre ens i de to delkorporer. Er der nogen konstruktioner i dette materiale der således potentielt karakteriserer nedskrevet talesprog mere generelt?

Til dette formål sorterede jeg mønstre med en signifikansværdi på 0,1 eller derunder fra. Samtidig frasorterede jeg mønstre fra der havde under 6 forekomster i LANCHART og under 5 i avisartiklerne (forskellen kompenserer for at der er færre løbende ord i avismaterialet), og satte desuden som kriterie at UIF i avismaterialet skulle være over 1. Det efterlod 89 mønstre, som blev sorteret efter SC·UIF.

Blandt disse mønstre er det mest slående hvor høj forekomsten af personlige pronomener er. Blandt de 89 mønstre er der ikke mindre end 70 der indeholder personlige pronomener (79 %), hvilket dog ikke er helt overraskende givet ordklassefordelingen ovenfor.

Derudover er der et antal mønstre der springer i øjnene som mere eller mindre konstruktionsagtige. De øverste pladser indtages af “<ngn/navn> ved ikke om” og “<ngn/navn> synes den/det er” – fragmenter af hhv. epistemisk og deontisk markerede ytringer. En topplacering indtager også “<ngn/navn> var NUM år”, et mønster der giver god mening i den personligt fortællende genre. Den grænsende til idiomatiske vending “lov til at VERBe” figurerer også højt, hvilket især skyldes den særligt stærke kollokation (målt på SC-værdi) mellem *lov*, *til* og *at*.

En konstruktion der nævnes som talesproglig af Galberg Jacobsen og Skyum-Nielsen (2007), er sætningskløvningen. Den er i brug i begge sprogformer, naturligvis med pronominelts subjekt, hvorved “den/det var den/det <ngn/navn>” dukker op på listen over stærkest kollokerende, ikke-signifikante mønstre. Det indledende “den/det” er aldrig *den*, kun *det*, hvorimod det andet “den/det” repræsenterer forekomster af begge.

Endelig er mønsteret “den/det var så ADJ” (altid med *det* initialt, og med diverse forskellige fortsættelser) ganske interessant. Det forstærkende *så* giver konstruktionen et tydeligt talesprogligt præg. Samtidig er det som om forekomsterne i LANCHART er en smule mere hverdagsagtige end forekomsterne i avismaterialet, jf. Tabel 17 (NB: “stærk” fortsætter med “en oplevelse”; “varm” burde have været “varmt” i LANCHART-transskriptionen).

Tabel 17. Forekomsten af “den/det var så ADJ” i LANCHART- og avismaterialet.

Ordfølge	LANCHART	Aviser
det var så åndssvagt	1	0
det var så cool	1	0
det var så lækkert	1	0
det var så langt	1	0
det var så uvirkeligt	1	0
det var så varm	1	0
det var så anderledes	0	1
det var så rart	0	1
det var så stærk	0	1
det var så underligt	0	1
det var så usexet	0	1
det var så absurd	0	1
det var så surrealistisk	0	1
det var så ukompliceret	0	1

Bemærk at fund som disse, hvor mønstrene ikke er hyppigere i den ene sprogform end i den anden, jo ikke nødvendigvis betyder at de er karakteristiske for narrativt talesprog uafhængigt af repræsentationsform; kun at de potentielt er det. Kun ved at sammenligne aviscitaterne og LANCHART-narrativerne på den ene side med mere generel avistekst på den anden vil man kunne påvise dette – en oplagt fortsættelse af undersøgelsen her.

6 Opsamling og diskussion

Jeg har i denne artikel ved hjælp af korpusstatistik identificeret mønstre og grammatiske konstruktioner der er signifikant overhyppige i to forskellige gengivelser af narrativt talesprog – hhv. udskrifter af sociolingvistiske interviews og citerede narrative passager i avisinterviews.

Overordnet ser aviscitaterne ud til at ligne transskriberede narrativer mere end avistekst generelt gør. Således ligner ordklassefordelingen i aviscitaterne på mange måder ordklassefordelingen i LANCHART-narrativerne, selv om der dog stadig er en vis overvægt af substantiver, præpositioner og adjektiver i aviscitaterne. Den leksikalske diversitet i aviscitaterne er større end i LANCHART-narrativerne, men ikke så stor som man ville have forventet af avissprog generelt. Ud fra en forestil-

ling om at featureinterviews eliciterer nogenlunde samme sprogbrug som sociolingvistiske interviews (selv om hele pointen med det sociolingvistiske interview jo netop er at det skal være særligt ikke-invasivt (jf. traditionen efter Labov & Waletzky 1967)), kan man spørge sig hvorfor der egentlig er forskel mht. leksikalsk diversitet. Er det fordi de interviewede har et større ordforråd, eller har det snarere med den journalistiske redigeringsproces at gøre? Det kan ikke udelukkes på basis af dette materiale at avisinterviewpersoner benytter sig af flere forskellige ord end sociolingvistiske informanter. Alene det at blive interviewet til avisen kan sikkert få mange til at tale 'pænere' og bruge flere 'fine' ord end de ellers ville have gjort, og den noget bredere emnekreds der antagelig afstedkommes af avisernes naturlige præference for det uhørte, vil nok også bidrage til større leksikalsk diversitet. Men meget af forskellen skyldes givetvis at avisinterviews redigeres således at udtrykket bliver mere koncist og mindre præget af gentagelser – "cut out the wordiness", som Brendan Hennessy skriver (2013: 277), hvilket vil give plads til flere indholdsord på den forhåndenværende plads. I den sammenhæng er det også interessant at den gennemsnitlige ordlængde for unikke ordformer (dvs. for det samlede anvendte ordforråd) kun er en smule større i aviscitaterne, nemlig 7,6 bogstaver mod 7,1 bogstaver i LANCHART-materialet. Målt på denne måde er ordforrådet i aviscitaterne altså ikke præget af væsentligt mere komplicerede ord end ordforrådet i LANCHART-materialet.

Selv om LANCHART-narrativerne og aviscitaterne altså på flere måder ser ud til at minde ikke så lidt om hinanden, er der en hel del talesproglige konstruktioner og andre træk der er signifikant mindre hyppige i avismaterialet – træk der virker velegnede til at tilføre talesproglig autenticitet og kolorit. Her er en opsummering i listeform:

- Sætningsindlederen "og så ...". Dette træk er ekstremt overhyppigt i LANCHART-narrativerne. De steder hvor det forekommer i aviscitaterne, er det ofte stilistisk markeret ved at stå efter punktum.
- Generelle udvidere såsom "og sådan noget".
- Indefinit demonstrativ reference med "sådan en ..".
- Definit reference med "den der .. der".
- Forfelftsfordobling.
- "Ikke" som gambit (kontaktpåkaldelse).
- Adverbialt "i hvert fald".
- Konjunktionalet "men altså".

- ‘Tomgangsfraser’ som “hvad hedder det” m.fl.
- Inkvitkonstruktionen “**være** .. sådan <cit>” (samt den anamnesticke funktion af *sådan*).
- Spatial *være*-konstruktion (“være (der)henne/ovre” osv.).
- Eksistentiel/præsentativ der-konstruktion (“der var en der ...”).

Hvorfor ‘mangler’ nogle af de talesproglige mønstre og konstruktioner i aviscita-terne? Hennessy (2013: 277) sammenfatter den journalistiske balanceakt i at stramme og rense uden at kvæle al kolorit:

However well controlled an interview is, talk is more rambling and repetitive than written-up material, a quality more conspicuous to the eye than the ear. So cut out the wordiness, but not to the extent of being left with gnomic utterances. Get across your interviewee’s way of speaking – ‘you know’, ‘indeed it is’, ‘not me matey’. Indicate a person’s accent at the start if it is interesting or distinctive, but don’t produce slabs of incomprehensible dialect. Correct faulty grammar that leads to lack of clarity but leave it in if it’s a trivial matter that adds colour.

Det der i hvert fald i rå udskrifter kan se ud som en rablende og selvgentagende stil, bør altså ifølge fagkundskaben redigeres om til noget mere renskuret, dog uden at kaste hele interviewpersonen ud med badevandet – en afvejning det er svært at være uenig i. I en samtale med mangeårig Politiken-journalist Lotte Thorsen bekræfter hun at det er en balancegang der arbejdes aktivt med i feature- og portræjtjournalistikken. Hun pointerer at en persons talestil, ofte i høj grad karakteriseret af bestemte ‘tomgangsfraser’ og lignende, er noget man som udgangspunkt gerne vil formidle i en interviewartikel, men som må gives med stor forsigtighed for ikke at blive for dominerende. Der er to meget reelle hensyn på spil: Dels risikerer man at skræmme en læser væk med for mange forstyrrende falbelader, og dels risikerer man at få sin kilde til at føle sig karikeret og til grin. I Thorsens arbejde vil hun således i nogle interviewartikler give et bestemt, idiosynkratisk taletræk en enkelt gang for at slå tonen an, men i mange andre helt udelade kraftige idiosynkrasier fordi allerede en enkelt forekomst kan have for stor effekt.

Hensyn som disse er formentlig en væsentlig del af forklaringen på de statistiske forskelle der ses i resultaterne. Mønstrene og konstruktionerne på den ovenstående liste kan for de flestes vedkommende opfattes som indeholdende sproglig tomgang og/eller ‘støj’ (fx er en konstruktion som “**være** .. sådan <cit>” i avis-korpussets 2016 stadig ret ny og kan derfor virke idiosynkratisk og ‘larmende’). En stor del af

de forskelle der dukker op her, skyldes således efter alt at dømme bortredigering af hvad der kan opfattes som forstyrrende elementer og overflødig fylde.

Der er dog også visse grammatiske præferencer der ikke dækkes af denne forklaring, fx de systematiske forskelle i brugen af konstruktioner med *være*, som jeg desværre ikke har haft plads til at undersøge nærmere her, samt distributionen af mønsteret “den her N” og den indefinitte demonstrativ “sådan en N”.

De indefinitte demonstrative konstruktioner, herunder “den her N” i specifik indefinit betydning, kunne godt fortjene at blive undersøgt nærmere mht. funktion og distribution i dansk talesprog. “Den her N” i specifik indefinit betydning er i øvrigt et eksempel på en talesprogs-karakteristisk konstruktion der er omtrent ligeligt fordelt mellem aviscitater og sociolingvistiske udskrifter.

Hvad aviscitaterne angår, er det især forekomsten af verbale komplementer (ledsætninger og infinitiver) der er karakteristisk. Men det er vigtigt at bemærke at selv om forekomsten er signifikant overhyppig i avismaterialet, ligger de to sprogformer alligevel relativt tæt på hinanden, også hvad dette træk angår. Et andet mønster der er signifikant overhyppigt i avismaterialet, er en præference for flere adjektiver efter hinanden i nominalhelheder, både med og uden sideordningskonjunktion – et stilistisk fænomen snarere end et grammatisk. En analyse af mønsteret “for NUM/man-ge/nogle/flere år siden” tyder på en præference for præcise tidsangivelser i aviscitaterne. Og endelig ser der ud til at være en præference for den klassiske kopulafunktion, mens brugen af konstruktioner med *være* i LANCHART er mere varieret, jf. ovenfor.

Hvorfor er det småt med karakteristiske konstruktioner i avismaterialet? En del af grunden er muligvis at den leksikalske diversitet er større i avismaterialet – den større diversitet kan bevirke at de karakteristiske konstruktioner aldrig bliver hyppige nok til at blive indfanget af de statistiske procedurer.

Blandt mønstre der ikke er signifikant overhyppige i nogen af delkorpuserne bemærker man især den hyppige forekomst af personlige pronomener. Der er også enkelte konstruktionsagtige forbindelser, fx “lov til at V”, sætningskløvning og “det var så ADJ”.

Sammenligningen af de to små delkorporer af mundtlige narrativer i to forskellige sprogdragter har vist sig at kaste en hel del af sig, især hvad angår mønstre og konstruktioner der påviseligt forekommer i narrativt talesprog, men tilsyneladende i ganske høj grad forbigås i fortællende interviews. Det lader desuden til at de redigerede citater er mere skriftsproglige mht. forekomsten af verbale komplementer (dvs. ledsætninger og infinitiver) – uden dog at være på niveau med det generelle

avissprog. Visse klassiske forskelle på tale og skrift er altså stadig at finde, selv når det skriftlige materiale er en decideret repræsentation af tale.

Den vigtigste operative norm der kommer til syne, er at 'sproglig tomgang' bør begrænses på bekostning af noget talesproglig kolorit, i en afvejning mellem en sober skriftnorm og en autentisk gengivelse af det vildtvoksende talesprog. Det mest fremtrædende i de her fremlagte resultater er altså ikke så meget en udlugning af træk fra dialekter, etnolekter, kronolekter osv., men snarere en soignering af folks idiolekter, dvs. en bortredigering af 'sjusk' (her på konstruktionsniveau) og sproglige idiosynkrasier.

Litteratur

- Ag, Astrid 2009: Talegenrer: Kvantitative og kvalitative forskelle på redegørelser og narrativer i det sociolingvistiske interview. Inger Schoonderbeek Hansen & Peter Widell (eds.) *12. Møde om Udforskningen af Dansk Sprog*. Århus: Århus Universitet. 23–34.
- Allwood, Jens 1998: Some frequency based differences between spoken and written swedish. *Proceedings of the 16th Scandinavian Conference of Linguistics, Turku University, Department of Linguistics*. 18–29.
- Asmussen, Jørg & Bjarne Ørsnes 2009: Konstruktioner og korpora - et case-studie i konstruktioner af typen „keine macht dem stress“. Ken Farø, Alexandra Holsting, Niels-Erik Larsen, Jens Erik Mogensen & Thora Vinther (eds.) *Sprogvidenskab i glimt*. Odense: Syddansk Universitetsforlag. 208–218.
- Bäckström, Linnéa, Lars Borin, Markus Forsberg, Benjamin Lyngfelt, Julia Prentice & Emma Sköldbog 2013: Automatic identification of construction candidates for a swedish constructicon. *Proceedings of the workshop on lexical semantic resources for NLP at nodalida 2013. NEALT proceedings series 19*. Linköping University Electronic Press. 2–11.
- Biber, Douglas 1991: *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Bick, Eckhard 2000: En Constraint Grammar parser for dansk. Peter Widell & Mette Kunøe (eds.) *8. Møde om Udforskningen af Dansk Sprog*. Århus: Århus Universitet. 40–50.
- Bjerrum, Marie 1975: Og? – efter punktum?? *Nyt fra Sprognævnet* 15. 1–4.
- Blom, Jonas Nygaard 2013. Radio- og tv-sprog. Thomas Thurah & Martin Kristiansen (eds.) *Sproget i de elektroniske medier*. København: Hans Reitzels Forlag. 105–161.

- Boye, Kasper, Mads Poulsen, Hannah Bruun Pedersen, Marie Herget Christensen, Line Dalberg & Nicoline Munck Vinther 2012: At eller ikke at i tale og skrift. *NyS* 42. 41–61.
- Christensen, Tanya Karoli 2012: Epistemiske sætningsadverbier i dansk talesprog. *Ny forskning i grammatik* 19. 419–483.
- Christensen, Marie Herget 2019: Sætningskløvnings i moderne talt dansk. *Danske Talesprog* 19. 37–76.
- Christensen, Tanya Karoli & Torben Juel Jensen 2014: Sætningsadverbialer i led-sætninger som forgrundssignal. *Ny forskning i grammatik* 21. 37–54.
- Christensen, Tanya Karoli & Torben Juel Jensen 2015: Grammatisk variation og forandring i moderne dansk. Gregersen & Kristiansen 2015 (eds.). 203–219.
- Christensen, Tanya Karoli & Torben Juel Jensen 2018: *Og sådan noget i den stil: Generelle udvidere i moderne dansk talesprog*. Tanya Karoli Christensen, Christina Fogtmann Fosgerau, Torben Juel Jensen, Martha Sif Karrebæk, Marie Mægaard, Nicolai Pharao & Pia Quist (eds.) *Dansk til det 21. århundrede – sprog og samfund*. 73–91.
- Christensen, Tanya Karoli & Torben Juel Jensen (eds.): *Explanations in sociosyntax: Dialogue across paradigms*. Cambridge: Cambridge University Press. Under udgivelse.
- Clausen, Olja Hedelund & Nikoline Pedersen 2017: En samtaleanalytisk undersøgelse af respons i forbindelse med formatet *hva hedder det*. *Skrifter om Samtalegrammatik* 4:3.
- Constant, Mathieu, Gülşen Eryiğit, Johanna Monti, Lonneke Van Der Plas, Carlos Ramisch, Michael Rosner & Amalia Todirascu 2017: Multiword expression processing: A survey. *Computational Linguistics* 43(4). 837–892.
- Croft, William 2001: *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- DGCSS 2011: Kodningsmanual til diskurskontekstanalyse (aka. IIV-analyse). dgcss.hum.ku.dk/aarsberetninger/rapporter/IIV-kodningsmanual_opdateret_januar_2011__2_.pdf. Besøgt 5.3.2020.
- Diderichsen, Paul 1962: *Elementær dansk grammatik*. København: Gyldendal.
- Diderichsen, Philip 2008: *One of a kind: The processing of indefinite one-anaphora in spoken Danish*. Ph.d.-afhandling fra Lund University Cognitive Science.
- DR 2018: Sprogpolitik for DR. www.dr.dk/static/documents/2018/11/20/2018_dr_sprogpolitik_101_e106a5b0.pdf. Besøgt 14.7.2020.

- Farø, Ken & Henrik Lorentzen 2009: De oversete og mishandlede ordforbindelser – hvilke, hvor og hvorfor? *LexicoNordica* 16. 75–102.
- Forsberg, Markus, Richard Johansson, Linnéa Bäckström, Lars Borin, Benjamin Lyngfelt, Joel Olofsson & Julia Prentice 2014: From construction candidates to construction entries: An experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames* 6(1). 114–135.
- Galberg Jacobsen, Henrik & Peder Skyum-Nielsen 2007: *Dansk sprog: En grundbog*. København: Schønbergs Forlag.
- Gregersen, Frans 2015: Talens forandringer – sproglig variation hos den talende inden for interviewsituationer og i samme situation over tid. Gregersen & Kristiansen 2015 (eds.). 221–242.
- Gregersen, Frans, Søren Beck Nielsen & Jacob Thøgersen 2009: Stepping into the same river twice: On the discourse context analysis in the LANCHART project. *Acta Linguistica Hafniensia* 41:1. 30–63.
- Gregersen, Frans & Tore Kristiansen 2015 (eds.): *Hvad ved vi nu – om danske talesprog?* København: Sprogforandringscentret, Københavns Universitet.
- Hansen, Josefine Straube 2020: Jeg var sån fuck mit liv. *Mål og Mæle* 41.
- Hansen, Erik & Lars Heltoft 2011: *Grammatik over det danske sprog*. København: Det Danske Sprog- og Litteraturselskab.
- Heegård, Jan & Peter Juel Henriksen (eds.) 2012: *Speech in action: Proceedings of the 1st SJUSK conference on contemporary speech habits*. Frederiksberg: Samfundslitteratur.
- Hennessy, Brendan 2013: *Writing feature articles*. New York: Taylor & Francis.
- Henriksen, Peter Juel 2002: Some frequency based differences between spoken and written Danish. *Gothenburg papers in theoretical linguistics* 88. 1–19.
- Henriksen, Peter Juel 2006: Det talte sprog – det tålte sprog? Anna Braasch (ed.) *Sprogteknologi i dansk perspektiv: En samling artikler om sprogforskning og automatisk sprogbehandling*. København: C.A. Reitzels Forlag. 178–201.
- Henriksen, Peter Juel & Jens Allwood 2005: Swedish and Danish, spoken and written language: A statistical comparison. *International journal of corpus linguistics* 10:3. 367–399.
- Jensen, K. Anne 2003: *Clause linkage in spoken Danish*. Ph.d.-afhandling fra Københavns Universitet.
- Jensen, Torben Juel 2005: *Irrealitetsmarkører i dansk talesprog: En korpuslingvistisk undersøgelse af hvis-konstruktioner og modalverber i medarbejdersamtaler og sociolingvistiske interview*. Ph.d.-afhandling fra Københavns Universitet.

- Jensen, Torben Juel 2007: Generisk du – en igangværende sprog(brugs)forandring. *Ny forskning i grammatik* 14. 163–184.
- Jensen, Torben Juel 2009a: Non-realitetsmarkører i dansk talesprog: Hvis man kunne sige det på den måde. *Studier i Nordisk* 2006. 75–101.
- Jensen, Torben Juel 2009b: Refleksivt anvendte pronominer i moderne dansk. *Ny forskning i grammatik* 16. 131–151.
- Jensen, Torben Juel 2011: Ordstilling i ledsætninger i moderne dansk talesprog. *Ny forskning i grammatik* 18. 123–150.
- Jensen, Mathias Juul 2019: *Sån der* – et talesprogsfænomen. *lingoblog.dk/saan-der-et-talesprogsfaenomen*. Besøgt 23.8.2020.
- Kristiansen, Tore 1996: Det gode sprogsamfund: Det norske eksempel. *NyS* 21. 9–22.
- Kristiansen, Tore & Nikolas John Robert Coupland (eds.) 2011: *Standard languages and language standards in a changing Europe*. Oslo: Novus Press.
- Kristiansen, Tore & Stefan A. Grondelaers (eds.) 2013: *Language (de)standardisation in Late Modern Europe: Experimental studies*. Oslo: Novus Press.
- Labov, William 1972: Some principles of linguistic methodology. *Language in society* 1:1. 97–120.
- Labov, William & Joshua Waletzky 1967: Narrative analysis: Oral versions of personal experience. June Helm (ed.) *Essays on the verbal and visual arts*. New York: American Ethnological Society. 12–44.
- Leech, Geoffrey, Paul Rayson & Andrew Wilson 2001: *Word frequencies in written and spoken English: Based on the British National Corpus*. London: Longman.
- le Fevre Jakobsen, Bjarne & Michael Ejstrup 2011: Takt og tale og tv-avis gennem 40 År. *Forandringens journalistik*. Århus: Ajour. 41–72.
- Linde, Charlotte 1993: *Life stories: The creation of coherence*. Oxford: Oxford University Press.
- Lyngfelt, Benjamin, Lars Borin, Kyoko Ohara & Tiago Timponi Torrent (eds.) 2018: *Constructicography: Constructicon development across languages*. Amsterdam: John Benjamins Publishing Company.
- Mahlberg, Michaela & Viola Wiegand 2018: Corpus stylistics, norms and comparisons: Studying speech in Great Expectations. *Rethinking language, text and context*. London: Routledge. 139–160.
- Mahlberg, Michaela, Viola Wiegand, Peter Stockwell & Anthony Hennessey 2019: Speech-bundles in the 19th-century english novel. *Language and Literature* 28:4. 326–353.
- Mortensen, Janus 2012: Epistemisk positionering i dansk talesprog. *NyS* 42. 62–91.

- Møller, Erik 1993: *Mundtlig fortælling – fortællingens struktur og funktion i uformel tale*. Institut for Dansk Dialektforsknings publikationer, serie A, nr. 31. København: C.A. Reitzels Forlag.
- Nelson, Katherine 2004: Construction of the cultural self in early narratives. Colette Daiute & Cynthia Lightfoot (eds.) *Narrative analysis: Studying the development of individuals in society*. Thousand Oaks: Sage. 87–109.
- Pedersen, Henriette Folkmann 2014: Om konstruktionen *sån en N* i danske samtaler. *Skrifter Om Samtalegrammatik* 1:1. 1–23.
- Rathje, Marianne 2011: Quotations and quotatives in the speech of three Danish generations. Frans Gregersen Jeffrey K. Parrot & Pia Quist (eds.) *Language variation. European perspectives III. Selected papers from the 5th international conference on language variation in Europe*. Amsterdam: John Benjamins. 71–82.
- Rayson, Paul & Roger Garside 2000: Comparing corpora using frequency profiling. *Proceedings of the workshop on comparing corpora-volume 9*. Association for Computational Linguistics. 1–6.
- Skaftø Jensen, Eva 2013: Ordklasseproblemer, tilfældet *sådan*. *LexicoNordica* 20. 55–74.
- Skytte, Gunvor 1999: Julekrybben. Diskursmarkering og konnexion. Gunvor Skytte, Iørn Korzen, Paola Polito & Erling Strudsholm (eds.) *Tekststrukturering på italiensk og dansk: Resultater af en komparativ undersøgelse*. København: Museum Tusulanum. 419–483.
- Stokoe, Elizabeth & Derek Edwards 2006: Story formulations in talk-in-interaction. *Narrative Inquiry* 16:1. 56–65.
- Therkelsen, Rita 2002: Talesprogets og skriftsprogets bestemthedsparadigmer. *Ny forskning i grammatik* 10. 189–208.
- Therkelsen, Rita 2003: Hvad laver de der her'ere og der'ere egentlig? Peter Widell, & Mette Kunøe (eds.) *9. Møde om Udforskningen af Dansk Sprog*. Århus: Århus Universitet. 273–284.
- Thurah, Thomas 2013: Det journalistiske sprog. Thomas Thurah & Martin Kristiansen (eds.) *Sproget i de elektroniske medier*. København: Hans Reitzels Forlag. 55–102.
- Thøgersen, Jacob, Nikolas Coupland & Janus Mortensen (eds.) 2016: *Style, media and language ideologies*. Oslo: Novus Press.
- Trap-Jensen, Lars 2007: Den skarpeste kniv i skuffen. sprognet.dk/raad-og-regler/artikler-mv/sprogligt-politikens-sprogklumme/10-februar-2007. Besøgt 5.3.2020.

Turner, Mark 1996: *The literary mind: The origins of thought and language*. Oxford: Oxford University Press.

Von Heusinger, Klaus 2011: Specificity, referentiality and discourse prominence: German indefinite demonstratives. Ingo Reich, Eva Horch & Dennis Pauly (eds.) *Proceedings of Sinn und Bedeutung 15*. 9–30.

Philip Diderichsen
Specialkonsulent, ph.d.
Institut for Nordiske Studier og Sprogvidenskab
Københavns Universitet
cpd@hum.ku.dk