


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Ordbog over Dansk Talesprog (ODT) – værkstedspræsentation af en ordbogsportal	
Forfatter:	Carsten Hansen & Martin H. Hansen	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 225-235 Rapport från Konferens om lexicografi i Norden, Lund 24.-27. maj 2011	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for lexicografi

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Ordbog over Dansk Talesprog (ODT) – værkstedspræsentation af en ordbogsportal

Carsten Hansen & Martin H. Hansen

The article presents the initial investigating stages of the development of a dictionary of spoken Danish, based on a corpus consisting mainly of sociolinguistic interviews at the LANCHART Centre in Copenhagen. The data are stratified and represent different segments of Danish speakers with regard to geography, gender, age and social class, recorded over several decades.

The purpose of the dictionary is to present spoken Danish to the interested layperson as well as the research-oriented user, and the dictionary will be designed as an online portal with these recipients in mind.

The article presents several examples of work in progress: establishing the editing principles of linguistic fact boxes which will give a more detailed presentation of a specific topic; a sketch of the portal design; a pilot project on interjections with an, as yet, unfinished dictionary entry on the lemma *av* ('ouch'); and finally, an investigation of laughter as part of spoken language which also demonstrates how the tagging of a sociolinguistic corpus can be put to use in a dictionary.

Nøgleord: Leksikografi, Talesprogskorpus, Pragmatik, Konversationsanalyse

ODT og Sprogforandringscentret

På basis af et talesprogskorpus bestående af primært sociolingvistiske interviews optaget i perioden 1978-2010 og som i den ortografiske transskription tæller op mod 7 mio. løbende ord, arbejder vi på Danmarks Grundforskningsfonds Center for Sociolingvistiske Sprogforandringsstudier, i daglig tale Sprogforandringscentret, på en ordbogsportal kaldet *Ordbog over Dansk Talesprog* (ODT). Centrets baggrund og formål blev ved åbningen i 2005 formuleret sådan af centerleder Frans Gregersen:

Vi vil opbygge en ny sprogforandringsteori baseret på empirisk forskning i sprogforandring i virkelig tid inden for det danske sprogsamfund. Danmark er som sprogsamfund et meget velegnet laboratorium for en undersøgelse af hvordan sprog forandrer sig og

hvorfor. Landet er overskuelig stort, det er veldokumenteret både i sproglig, i historisk og sociologisk henseende, og der er gennemført en række undersøgelser af dansk tale-sprog i dets variationer, geografisk og socialt. Alle disse karakteristika gør det oplagt at svare på grundlæggende spørgsmål om hvordan sproglig variation og sprogforandring hænger sammen.

Vi ved at sproget bruges forskelligt af de forskellige aldersgrupper, af de to eller flere køn og af socialt og geografisk forskellige personer, men hvordan hænger det sammen med de langtrækkende tendenser i det danske sprogsamfunds forandring fra et dialekt-samfund med få store byer omkring århundredskiftet fra 1800 til 1900, og til et gennemurbaniseret informationssamfund med mange minoritetssprog omkring årtusindskiftet?

Vi har på centret gentaget tidligere undersøgelser med så vidt muligt samme informanter, vi har suppleret med informanter der har samme baggrund, og endelig har og vil vi fortsat afprøve en række hypoteser og fordomme på dette materiale. Data er derfor omhyggelige genindspilninger af de informanter som tidligere medvirkede i syv forskellige dialektologiske og sociolingvistiske undersøgelser fra forskellige dele af Danmark, nemlig



- Vinderup, Jylland: båndoptagelser udført i 1973, 1978 og igen i 2006-2008
- Vissenbjerg, Fyn: optagelser i 1982-83 og igen i 1998-99
- Odder, Jylland: optagelser i 1986-89 og igen i 2005-2008
- BySoc (Projekt Bysociolingvistik), København: optagelser i 1986-89 og igen i 2005-2009
- Næstved, Sjælland: optagelser i 1986-90, 1993 og igen i 2005-2007
- Køge, Sjælland: optagelser i 1989-98 og igen i 2006-2007
- DASVA (Dansk og Svensk i Akkomodation), København: optagelser i 2001-2002, endnu ikke gentaget.

Figur 1. Undersøgelsesområder

For detaljer om de enkelte undersøgelser design, se oversigten på Sprogforandringscentrets hjemmeside: <http://dgcss.hum.ku.dk/undersogelsesomraader/>

Gentagelsen har givet os data om sprogforandringer *i virkelig tid*, dvs. de samme taleres forandrede – eller konstante – tale efter 30, 20 og 5 år, henholdsvis.

Normalt arbejder man med aldersfordelte informantsamlinger så man slutter fra forskellene mellem aldersgrupperne: Sådan som de unge taler, bliver sproget i fremtiden når de gamle er forsvundet. Det kalder man at studere sprogforandringer i forestillet tid (*apparent time*). Det vi gør, er altså at studere sprogforandring i virkelig tid (*real time*), heraf projektets engelske kælenavn: LANCHART, Language Change in Real Time.¹

Siden centrets åbning er nye undersøgelser kommet til, således at der nu som nævnt foreligger et transskriberet korpus med op mod 7 mio. løbende ord. Der ligger 1587 timer, 33 minutter og 25 sekunders lyd i Sprogforandringscentrets korpus, og der er brugt mere end 17.000 timer på transskription og opmærkning.

Ånden i ODT

ODT's virkefelt er dobbelt. For det for første bedriver vi i redaktionen egen *forskning*, dels i talesprog på linje med Sprogforandringscentrets øvrige aktiviteter og dels i det uopdyrkede felt talesprogsleksikografi. På den måde *producerer* redaktionen, tilknyttede specialeskrivere og ph.d.-studerende ny viden. Den anden side af vores virkefelt er *formidling*. Her ser vi det som vores opgave ikke blot at frugtbargøre og eksponere centrets og redaktionens egne forskningsaktiviteter, men opgaven består også i at formidle talesprogsforskning og viden om talesprog indvundet uden for centret. Imidlertid skal der ikke herske nogen tvivl om at den bærende ide i vores arbejde er etableringen af talesprogsportalen, dvs. praktiseret talesprogsleksikografi.

ODT's brugersegment er også dobbelt. *Primæradressaten* er den sprogligt/lingvistisk dannede alménbruger der interesserer sig for talesprog, *sekundæradressaten* er den forskningsorienterede bruger. Begge adressater skal opleve portalen som tilgængelig (dvs. med hurtig accesstid), segmentdifferentieret (dvs. relevant), servicerende, reklamefri og gratis. Det skal i det hele taget være en positiv oplevelse at frekventere portalen.

ODT er sig sin rolle som dataredaktør bevidst. Ud over de redigerede ordbogsartikler vil ordbogsportalen indeholde en række bokse med oversigts- hhv. dybdegående artikler om talesprogsfænomener dokumenteret i korpus og relateret til ordbogsartiklerne. Brugeren kan i den endelige version gå bag om de redigerede fremstillinger såvel i ordbogsartikler som i bokse ved at foretage direkte søgninger i korpus, men grundlæggende dokumenteres søgningerne i artiklens dokumentationsdel.

¹ Den oprindelige ordlyd er tilpasset nærværende tekst.

ODT insisterer på at demonstrere at en ordbog der konsekvent er baseret på talesprogsdata, på afgørende punkter adskiller sig fra den skriftsprogsbaserede ordbog. Det drejer sig for det første i hvert fald om lemmabestanden – også når den skriftsprogsbaserede ordbog medtager pseudotalesproglige kilder som fx fiktionsdialoger. Det drejer sig dernæst om morfologiske paradigmer, syntaktiske mønstre, betydningsvarianter og om de pragmatiske og diskursive niveauer.

Bokse

Sprogbokse skal formidle en (kortfattet) karakteristik af det talesproglige fænomen. Når ODT's interne redaktører og de eksterne boksforfattere skal redigere en boks, er de forpligtede på *boksmanualen*:

Fremstillingen skal

- karakterisere det talesproglige fænomen
- funderes i belæg fra korpus
- supplerende funderes kontrastivt i forhold til skriftsprog
- adresseres dobbelt.

Fremstillingen kan

- indeholde supplerende karakteristik på baggrund af andet end korpus. Det angives eksplicit når karakteristikken baseres på andet end centrets korpus
- inddrage forskning i fænomenet.

Det talesproglige fænomen kan være stort set alle aspekter af talesprog. Det kan være et leksem: *ah, hov, jamen, nå, undskyld*, det kan være en funktionsgruppe: *begejstring*, eller det kan være noget andet talesprogsrelevant: *latter, pauser* el. *-agtig*. I redaktionen er vi løbende på jagt efter boksemner og trækker gerne på kræfter uden for redaktionen. Blandt Sprogforandringscentrets forskere har vi sikret os bokse om følgende emner i tilknytning til deres igangværende forskning: *Epistemiske adverbier* (Tanya Karoli Christensen), *at-bortfald* (Kasper Boye), *Genstandsfald* (Jeffrey Parrott), *Generiske pronominer* (Torben Juel Jensen). Vi låner også gerne specialviden fra den relevante faglitteratur, fx *plus at* (Nørgård-Sørensen 2001), *"ja men nej"- og "nej men ja"-svar* (Thøgersen & Nielsen 2009) eller *her'ere og der'ere* (Therkelsen 2002).

ODT som ordbogsportal

ODT er designet som en netbaseret ordbogsportal med mulighed for kontrastive parallelsøgninger i et skriftsprogskorpus og paralleloplag i skriftsprogsbaserede ordbogsressourcer, dvs. at ODT er et "sted på internettet som fungerer som indgang til søgning efter bestemte oplysninger, idet det indeholder en stor mængde links til forskellige hjemmesider som er opdelt efter emne eller vedrø-

rer et bestemt emne” (Den Danske Ordbog). Dette design afspejler sig i nedenstående udkast til portalens velkomstsider:

Velkommen til Ordbog over Dansk Talesprog (ODT)

Her kan du få mere at vide om ODT

Her kan du få mere at vide om DGCSS' korpus

Her kan du få mere at vide om undersøgelsesområderne der ligger til grund for DGCSS' korpus

Her kan du få mere at vide om DGCSS

Her kan du få mere at vide om andre danske talesprogskorpora

Her kan du få mere at vide om engelske talesprogskorpora

Her kan du få mere at vide om norske talesprogskorpora

Her kan du få mere at vide om svenske talesprogskorpora

Her kan du få mere at vide om [xxx] talesprogskorpora

Kontakt ODT

Du er besøgende nr. 1236

Indtast søgeord:

ha

Boks: Kort og godt om *ha* og *latter*

Links:

- Her kan du få mere at vide om *latter* i andre ordbøger, korpora og fremstillinger
- *ha*-konkordans
- *ha*-statistik

Pilotprojekt om interjektioner

ODT-projektet befinder sig i en indledende eksplorativ fase, både teoretisk og metodisk. I loyalitet med vores talesprogskorpus og det sociolingvistiske interview har vi valgt at gennemføre et pilotprojekt med de entydigt samtalepecifikke tokens der i korpus er annoteret som interjektioner. For at tage højde for eventuelle fejl i den automatiske PoS-opmærkning supplerer vi centrets egen inventarliste over interjektioner med Den Danske Ordbogs interjektionsliste. Vi går i denne eksplorative fase noget lemfældigt til værks mht. den ellers nødvendige kategoriale skelnen mellem interjektioner, onomatopoieta og modalpartikler. Schwitalla (2003) og Fiehler (2005) åbner op for en funktionel sammenfatning af disse grupper i termen *Gesprächspartikeln*.

Vi arbejder med to forskellige procedurer der supplerer hinanden. I den første procedure gennemgår vi de to interjektionslister semasiologisk og funktionsbestemmer hver kandidat: Hvilken slags job udfører kandidaten i samtalen? Fra Adolphs (2008) har vi hentet termen *funktionsprofil*, nemlig det sæt af forskel-

lige men supplerende funktioner der i deres hele udgør leksemets funktionsprofil. Fra Schwitalla (2003: 157) og Fiehler m.fl. (2004: 204ff) har vi hentet inspiration til en egen funktionsgruppetytologi, vel vidende at Schwitallas og Fiehler m.fl.s lister er orienteret mod sekventiel kategorisering, det vil sige mod turtagning i konversationsanalytisk forstand, i modsætning til vores der på nuværende stadium er mere indholdsorienterede.

I den anden procedure supplerer vi de forskellige funktionsgrupper med kandidater fra det semasiologiske analysearbejde. Det tjekkes om funktionsprofilerne kan bidrage med én eller flere funktioner til det samlede funktionsgruppelinventar. Tentative eksempler på vores funktionsgrupper er: *Beklagelse, bekræftelse, bekymring, erkendelse, hilsen, overraskelse (glædelig), overraskelse (forskrækket), selvrettelse, tøven, skepsis, væmmelse*.

Nedenfor er gengivet første version af en artikel over interjektionen *av*, skrevet af Mette Skovmand Juul, en af ODT's studenterredaktører. Artiklen er anonymiseret idet problematikken om informanternes anonymitet i forhold til ODT-portalen endnu ikke er afklaret.

Lemma: *av*

*av*¹ (overraskelse)

Antal belæg: 40

Betydninger:

1. *av*: bruges for at udtrykke overraskelse eller anden (ofte ubehagelig) følelse: *Av det dur ikke* (Vinderup 2006)
2. *av* for satan/søren/pokker/dælen, *av* min arm: bruges som mildt kraftudtryk for at udtrykke ærgrelse eller overraskelse: *Av for satan en historie altså* (BySoc 2007)

Funktionsgruppe: overraskelse/forskrækkelse, beklagelse

*av*² (minimalrespons)

Antal belæg: 27

Betydninger:

1. *av*: ord man siger mens man lytter til en andens tale: *Ja... ja... av... ... laver I andet* (Odder 1988)

Funktionsgruppe: minimalrespons

*av*³ (medgiven)

Antal belæg: 1

Betydninger:

1. *av ja*: give nogen ret i noget, medgive, indrømme: *Det er det jo det... av ja det er varmt* (Næstved 2006)

Funktionsgruppe: bekræftelse

Værkstedsrapportering om latter

I vores interjektionsliste finder man også strengen *ha*. I det transskriberede og annoterede korpus er alle lydsekvenser der høres som latter, opmærket på samme måde, nemlig som *ha*. Uanset om der bliver leet højt og længe, om der bliver leet kort og forsigtigt, eller om der faktisk bliver sagt 'ha', står der *ha* i korpus. Det har den fordel at vi hurtigt kan identificere al latter i korpuset, men det har til gengæld den ulempe at det er vanskeligt at skelne mellem forskellige former for latter. Der er i alt cirka 60.000 forekomster af *ha*. Redaktionens opgave består nu i en teoretisk og metodisk afklaring af hvordan vi håndterer latter.

Det teoretiske afsæt

En traditionel opfattelse består i at opfatte latter som en kausal reaktion på en stimulus, typisk humor. Både Carroll (2001) og Vøge & Wagner (2010) peger på et generelt problem i denne opfattelse, nemlig vanskeligheden ved at forklare hvorfor en bestemt stimulus nogle gange udløser latter og andre gange ikke. Den konversationsanalytiske forskning (CA) antyder en løsning på dette problem idet latter faktisk følger et forudsigeligt mønster når vi betragter den som interaktion. Desuden tyder CA-resultater på at latter alligevel er en slags sprog: Ganske vist udfører latter ikke en semantisk reference til samtaleens indhold, men den udfører en pragmatisk funktion i samtalsituationen. Latter er med andre ord en systematisk produceret, socialt organiseret aktivitet der tilpasser sig eller påvirker den løbende samtale – den er interaktionelt funderet (Jefferson m.fl. 1987).

Fra kvalitativ til kvantitativ analyse

CA har anvendt en kvalitativ metodisk tilgang til at analysere latter: Hvem ler først? Hvilken funktion har det at nogen ler først? Ler man med hinanden eller ad hinanden? Fortæller man noget sjovt eller alvorligt mens man ler – og hvilken effekt har dette for hvordan samtaleren bør reagere? Et meget stort antal isolerede analyser har givet stof til udvikling af en latterens typologi. Men man har derimod ikke noget kvantitativt overblik over de forskellige lattertypers udbredelse i forskellige kontekster, og dette manglende overblik indebærer også

at man ikke har mulighed for at vide om man har alle væsentlige latter typer og deres eventuelle blandingsformer med i typologien.

Sprogforandringscentrets talesprogs korpus kan bidrage til løsningen af disse spørgsmål. Korpusset er som nævnt designet med henblik på studiet af indflydelsen på talesprogspraksis fra de traditionelle sociolingvistiske variable køn, alder, klasse, men især tid og sted. En refleksion og inddragelse af disse træk hører netop til ODT's konstituerende træk. Konkret: Bliver der leet mere af kvinder end af mænd, af unge end af ældre, af arbejdere end af middelklasse, i 70'erne end i 00'erne, i Vinderup end i København?

Det store problem ved en kvantitativ analyse af sprogets pragmatiske funktioner er at der ikke er et simpelt én-til-én-forhold mellem form og funktion. Man må derfor i hvert enkelt tilfælde finde og systematisere de forskellige typer af form-funktion-relationer der gør sig gældende for det konkrete, sproglige udtryk der skal undersøges (i dette tilfælde latter), og derefter beskrive hvilke funktioner udtrykket typisk udfylder. Her har vi adopteret konceptet *funktionsprofil* fra den engelske korpuspragmatiker Svenja Adolphs:

The development of functional profiles [...] tries to shift the focus from inference procedures to form/function composites. The relationship between form and function is here described in terms of recurrent associations in language-in-use. These associations become part of the functional profile [...]. (Adolphs 2008: 29-30)

Et vigtigt aspekt af ODT's fortsatte arbejde med latter, og med multifunktionelle talesprogsudtryk i det hele taget, bliver at udarbejde sådanne funktionsprofiler.

Præsentations- og redigeringsværktøjer

Vi har Ole-Opfinderagtigt udviklet en række værktøjer mens vi venter på at der via eksterne midler kan knyttes leksikografisk it-kompetence til ordbogen. Nogle af værktøjerne vil vi integrere i portalen som præsentationsværktøjer, mens andre tjener som redigeringsværktøjer. Desuden har vi værktøjer af mere eksperimentel karakter, værktøjer som vi endnu ikke er sikre på hvordan vi vil bruge, hvis overhovedet. Vi befinder os som sagt i en eksplorativ fase.

På ordbogsportalens velkomstsider ovenfor så vi to menu punkter: *Konkordans* og *Statistik*. En reduceret konkordans med *ha*-forekomster ser således ud:

Fund	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10
1.på	nogen	måder		nej		mm		ja	okay	ha		ja		ja		ja		okay		nej	
2	ha		ja	okay	ja	doktor	hjaelpeløs	ja	ha	okay	okay	ja		ja		ja		okay		ja	
3	er	da	noget	forkalket	ham	den	ene	er	han	ikke	ha	hvad	nej	ja	vi	ville	sgu	ikke	med		
4			mm	ha	ha	ha	mm		ha	det	er	okay			nogle	gange	fordi	man	før		
5	ja	nåh	jamen	det	kunne	også	være	meget	sjovt	ha	men	der	har	jeg	nemlig	sådan	en	her	øh	hå	
6	nej	ha	vi	ville	sgu	ikke	med	ind	ha		Lady	er	det	ikke	sådan	den	hedder	ha			
7	nå	ej	hvor	irriterende	så	ja	ikke	med	ind	ha	ja	ja	og	det	doktor	hjaelpeløs	ja				
8	betændt	skal	vi	ikke	snakke	om	noget	andet	ha	ja	og	det	er	meget	frst	det	er			meget	
9	ham	den	ene	er	han	ikke	ha	hvad	nej	ha	vi	ville	sgu	ikke	med	ind	ha				
10	ikke	det	er	lidt	tåbeligt	jo	ej	dog	ikke	ha		dog	ikke	nu	der	bryder	jeg				
11	hedder	ha		ha	ha	ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg				
12	sådan	den	hedder	ha	ha	ha	ha	ha	ha	ha	men	skriver	du	selv	sangtekst	da	for	eksempel	ja		
13	ha		mm		mm		mm		ja	uh	ha	jamen	nu	har	han	jo	igennem	mange	mange	år	
14	Thor	Timler	han	kan	jo	en	masse			uh	ha	jamen	nu	har	han	jo	igennem	mange	mange	år	
15	ha		ha		ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	var	på			
16	lektioner	som	vi	har	mm	mm	mm	mm	timer	ha	ha	mm	mm	mm	mm	mm	mm	sådan	ja		
17	og	hvad	med	computer	bruger	du	det	meget	den	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	
18	ha	Lady	er	det	ikke	sådan	det	den	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	
19	på	en	eller	anden	dårlig	måde	og	ja	ha	der	har	også	de	ha	har	også	lavet	en			
20	det	ikke	sådan	den	hedder	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ligesom

Figur 2. Reduceret ha-konkordans

Som anført ovenfor er al latter i alle lydsekvenser der høres som latter, opmærket på samme måde, nemlig som *ha*. De tomme felter er pauser, men pauser skal i denne sammenhæng forstås som intervaller hvor nogen andre taler. De enkelte konkordanslinjer viser kun den enkelte talers tale – konkordansen består kun af tale, ikke af *samtale*. Et af de redskaber vi p.t. mangler i vores værktøjskasse, er således en let generérbar samtalekonkordans for at kunne klassificere *ha*-sekvenserne funktionelt. Her er der ikke bare tale om et internt redigeringsbehov men også om et brugerfladebehov: Brugeren skal kunne finde eksempler på de forskellige latterfunktioner og skal kunne vælge at høre lyden og/eller læse dialogen. Dels som belæg i forbindelse med den redigerede funktionsbestemmelse og funktionsprofilering men også som suveræn brugerkontrol af de redigerede fremstillinger i ordbogsartikler og bøger.

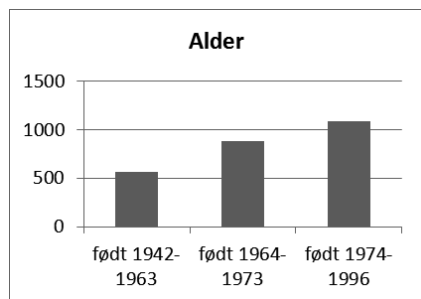
Konkordansen i Figur 2 er imidlertid også utilstrækkelig i en anden henseende. Hvis korpus virkelig skal ydes retfærdighed for sine sociolingvistiske variable skal det – ikke bare som redaktør men også som bruger – være muligt at generere en udvidet konkordans med de sociolingvistiske variable køn, alder, klasse, tid og sted samt andre opmærkninger:

	-2	-1	0	1	2	3	4	5	6	7	8	9	10	filename	project	recording	recording	participant	gender	birthyear	socialclass	xmin	
ja	okay	ha		ja	ja	ja	ja	ja	okay	nej	byso2ny:BySoct	ny	2006	XMM	MALE	0	1743	1886					
ja	ja	ha		okay	okay	ja	ja	ja	okay	ja	byso2ny:No name	ny	2006	XMM	MALE	0	2070	182					
han	ikke	ha	hvad	nej	ha	vi	ville	sgu	ikke	med	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2024	84				
meget	sjovt	ha	det	er	okay	nogle	gange	fordi	man	får	amager3	Amager3	gl	2009	FWA	FEMALE	0	8,75					
ind	ha	ha	men	der	har	jeg	nemlig	sådan	en	her	øh	vinderup2	No name	ny	2007	XMF	FEMALE	0	3705	108			
ja	ha	ha		Lady	er	det	ikke	sådan	den	hedder	ha	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2030	031			
ja	and	ha	ja	og	det	er	doktor	hjaelpeløs	ja	ha	byso2ny:No name	ny	2006	XMM	MALE	0	2651	403					
nej	ha	vi	ville	sgu	ikke	med	ind	ha		ha	byso2ny:BySoct	ny	2006	TEL	MALE	1949	WORKING	6033	532				
dog	ikke	ha		vi	ville	sgu	ikke	med	ind	ha	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2027	655				
ja	ha	ha		dog	ikke	nu	der	der	bryder	jeg	byso2gl:1	BySoct	gl	1987	MIP	MALE	1970	MIDDLE_C	3506	155			
ja	ha	ha	ha	hh	det	er	ligesom	min	kusines	hest	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2073	075				
ja	ha	ha	ha	ha	hh	det	er	ligesom	min	kusines	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2073	997				
ja	ha	ha	men	skriver	du	selv	sangtekst	da	for	eksempel	ja	byso2gl:1	No name	gl	2010	XMA	FEMALE	0	1493	07			
ha	uh	ha	jamen	nu	har	han	jo	igennem	mange	mange	år	byso2gl:1	BySoct	gl	1987	ATR	MALE	1932	WORKING	4469	42		
ha	ha	hh	det	er	ligesom	min	kusines	hest	jeg	var	på	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2087	021			
ja	timer	ha	ja	ja	mm	mm	mm	mm	mm	sådan	det	amager3	Amager3	gl	2009	FWA	FEMALE	0	21	3355			
ja	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	ja	byso2gl:1	No name	gl	2010	XMA	FEMALE	0	1419	033			
ja	ha	ha	der	har	også	de	ha	ha	ha	ha	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2058	455				
ja	ha	ha	der	har	også	de	ha	ha	ha	ha	lavet	en	odder3-05	Odder3	2006	SMP	FEMALE	1989	WORKING	2475	715		
ja	ha	ha	ha	ha	ha	ha	ha	ha	ha	ha	er	ligesom	vinderup2	Vinderup	gl	1978	AFH	FEMALE	1964	2069	748		

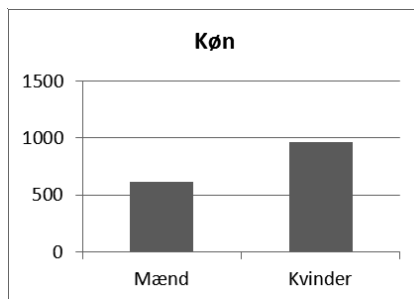
Figur 3. Udvidet ha-konkordans

De sociolingvistiske variable fortjener imidlertid at blive statistisk raffineret:

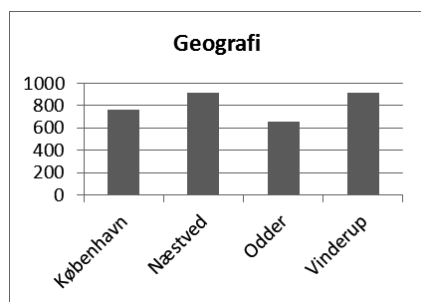
Figur 4–7: Relative frekvenser pr. 100.000 tokens for søgestrengen *ha*



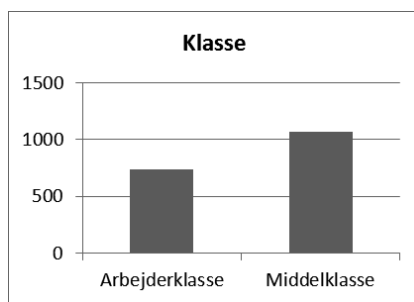
Figur 4



Figur 5



Figur 6



Figur 7

I Figur 4–7 ser vi i formidlet form de relative frekvenser for søgestrengen *ha* fordelt på nogle af de sociolingvistiske variable. Nu kan vi se hvem det er der ler: Unge ler mere end gamle, kvinder ler mere end mænd, middelklassen ler mere end arbejderklassen, og endelig bliver der tilsyneladende leet mere i Næstved og Vinderup end i København og Odder.

Nogle af de her viste forskelle skal nok forklares med at korpus er sammensat af geografiske og tidsmæssige punktnedslag og dermed i streng forstand ikke er repræsentativt for dansk talesprog. En forklaring på de geografiske forskelle kunne være at man har gennemført forskellige typer af sociolingvistiske undersøgelser på de forskellige lokaliteter. Men vi kan faktisk, i hvert fald i store træk, tegne en profil af hvilke sprogbrugere der ler mest, ligesom vi kan generere tilsvarende præsentationer for alle andre søgeord ved hjælp af samme skabelon. Den dag hvor vi får opmærket de forskellige latterfunktioner i korpus, kan vi tilsvarende generere individuelle præsentationer for hver latterfunktion.

Afsluttende bemærkninger

Afslutningsvis vil vi pege på to udfordringer. Den første er af økonomisk art. Vi har her i projektets indledende eksplorative fase trukket på både centerinterne og centereksterne ressourcer. På længere sigt er det imidlertid en forudsætning at vi får tilført eksterne midler til et så ambitiøst projekt som ODT, både til det redaktionelle arbejde og til it-løsninger.

En anden udfordring i arbejdet med ODT består i en række teoretiske og metodiske afklaringer: Kan det forsvars udelukkende at udarbejde ordbogen på basis af en ortografisk transskription, eller skal lyd inddrages både i skrivning af artikler og i dokumentationen? Er det rimeligt at overtage det literale lemmabegreb, eller er det nødvendigt at etablere et oralt? Endelig er vi overbeviste om at ODT på længere sigt må inddrage multimodale dimensioner.

LITTERATUR

- Adolphs, Svenja, 2008: *Corpus and Context. Investigating pragmatic functions in spoken discourse*. Amsterdam.
- Carroll, Noël, 2001: *Horror and Humor. I: Beyond aesthetics*. Cambridge. S. 235-254.
- Fiehler, Reinhard, 2005: *Die Gesprächspartikel. I: Duden Band 4, 7. Auflage*. S. 601-606.
- Fiehler, Reinhard (m.fl.), 2004: *Eigenschaften gesprochener Sprache*. Tübingen.
- Jefferson, Gail (m.fl.), 1987: *Notes on laughter in the pursuit of intimacy. I: G. Button and J.R.E. Lee (red.): Talk and Social Organization*. Clevedon. S. 152-205.
- Nørgård-Sørensen, Jens, 2001: *Plus at – en ny konjunktion i dansk*. *Danske Studier* 2001. S. 65-84.
- Schwitalla, Johannes, 2003: *Gesprochenes Deutsch. Eine Einführung*. Berlin.
- Therkelsen, Rita, 2002: *Hvad laver de der her'ere og der'ere egentlig? I: 9. Møde om Udforskningen af Dansk Sprog*.
- Thøgersen, Jakob & Nielsen, Søren Beck, 2009: *Ja men nej og nej men ja*. Når man hverken be- eller afkræfter ja/nej-spørgsmålets antagelser. Eller: Det fuldt udfoldede 'nja'. I: Hovmark, H., Stampe Sletten, I., & Gudiksen, A. (red.): *I mund og bog: 25 artikler om sprog tilegnet Inge Lise Pedersen på 70-årsdagen d. 5. juni 2009*. København. S. 337-350.
- Vøge, Monika & Wagner, Johannes, 2010: *Social Achievements and Sequential Organization of Laughter. I: Journal of Pragmatics* 42. S.1469-1473.

Internet

Den Danske Ordbog: <http://ordnet.dk/ddo>

Sprogforandringscentret: <http://dgcss.hum.ku.dk>

Carsten Hansen

Sprogforandringscentret DGCSS, Københavns Universitet
carhan@hum.ku.dk

Martin H. Hansen

Sprogforandringscentret DGCSS, Københavns Universitet
martin_h_hansen@yahoo.dk