

Hans mening om hendes krop

Portrætter af to køn i danske tekstkorpusser

Jann Scheuer
Københavns Universitet

Denne artikel, som er baseret på mit foredrag ved MUDS 16, handler om søgninger i danske tekstkorpusser efter *hendes* hhv. *hans* + nominal inden for en afstand af to ord. I artiklen bruges 'N' som symbol for nominalet.

Eksempler på søgeresultater er fx *hendes lejlighed*, *hans smil*, *hans mund*, *hendes værker*, *hans ideer*, *hans kolleger*, *hendes virksomhed*. Disse resultater kan have optrådt i hypotagmer som *hendes dejlige lejlighed*, *hans andet smil*, *hans tredje onkels altid vege mund*, *hendes syv værker*. Men nominalhypotagmernes beskrivere m.m. er blevet ignoreret. Et søgeresultat består altså af en bestemmer og en kerne, uanset hvad der oprindeligt måtte have optrådt indimellem (Jf. Hansen 1984; GDS:537).

Årsagen til begrænsningen af afstanden mellem pronomen og nominal er at antallet af søgeresultater ville eksplodere og blive u håndterligt hvis der anvendtes en større afstand.

Artiklens og foredragets afsæt er at søgeresultaterne viser hvordan kønnene bliver beskrevet i de anvendte korpusser. Hvis antallet af *hendes hår* er stort og *hans hår* er lille, fortolkes det som et udtryk for at kvinders hår omtales mere end mænds hår, og at kvinder i højere grad end mænd beskrives som nogle der har hår.

Når netop forbindelserne *hendes* og *hans* + nominal er valgt, skyldes det at der er tale om en relativt indiskutabel relation. Hvis der står *hendes hår* er der beskrevet en hun der har noget hår. Hvis man beskæftiger sig med sætninger, som fx *hun klippede hår hele dagen*, er relationen til diskussion (Vatvedt Fjeld 2015).

Nominalerne i søgeresultaterne refererer til forskellige størrelser: produkter af en eller anden form for virksomhed: *værker* (fysisk virksomhed), *ideer* (åndelig virksomhed); attributter: *ben*, *arm*, *hår* (kropslige), *søster*, *kolleger*, *kæreste* (relationelle); ejendom: *bil*, *hus*, *hjem*, *paraply*. Listen er ikke udtømmende. I artiklen sorteres paradigmerne i forskellige diskurser, mere om det senere.

Benyttede korpusser

Der er benyttet fire korpusser i undersøgelsen:

- Det Danske Sprog- og Litteraturselskabs database *KorpusDK* på 56 millioner ord og herunder *Korpus 90* og *Korpus 2000* på hver 28 millioner ord, som er dele heraf (ordnet.dk)
- Det Danske Sprog- og Litteraturselskabs *TiDK* på knap 400 millioner ord, som kan tilgås gennem søgemaskinen CoREST 2016 (korpus.dsl.dk). (Der er søgt i 231 millioner ord, jf. nedenfor.)
- Danmarks Grundforskningsfonds Center for Sociolingvistiske Sprogforandringsstudiers (DGCSS) database på 10 millioner ord transskriberet talesprog, indsamlet over en mere end 50-årig periode. Kan ikke tilgås gennem internettet.
- Nordisk Forskningsinstituts database med to millioner danske tweets fra 2011 til 2015 på 60 millioner ord. Kan ikke tilgås gennem internettet.

I artiklen er resultaterne af søgningerne vist i en række tabeller som er placeret i teksten. Hensigten er at spare læseren for at bladre frem og tilbage. Tekst og tabel kan ikke læses hver for sig.

Optælling af *hun* og *han* og *hendes* + N og *hans* + N

I tabel 1 vises indledningsvis antallet af forekomster af 3. persons pronomener i han- og hunkøn i nominativ og i forbindelserne *hendes* + N og *hans* + N.

	Antal ord	<i>hendes</i> + N	<i>hans</i> + N	<i>hun</i>	<i>han</i>
KorpusDK	56 millioner	36.698	74.270	265.369	542.351
TiDK udsnit	231 millioner	82.976	246.416	492.689	1.331.204
DGCSS	10 millioner	2.361	2.968	44.926	71.113
NFI tweets	60 millioner	3.410	6.183	548	1.017

Tabel 1. Antal forekomster af *hendes* + N, *hans* + N, *hun* og *han*

Generelt refereres der suverænt mest til hankønspersoner. *Han* optræder jævnt hen dobbelt så mange gange som *hun*. Teksten i danske tekstkorpusser handler overvejende om mænd. Hvis man læser en gennemsnitlig dansk tekst, hører man dobbelt så meget eller mere om mænd end om kvinder. *Hendes* + N og *hans* + N optræder med den samme fordeling, undtagen i DGCSS-korpusset hvor de er klædeligt anderledes fordelt: 2.361 *hendes* + N over for 2.968 *hans* + N.

I de to første korpusser er pronomenerne i nominativ langt hyppigere end *hendes* + N og *hans* + N i et nogenlunde ens forhold. I twitter-korpusset er de radikalt anderledes fordelt, idet *hun* og *han* optræder omkring 6 gange mindre end *hendes* + N og *hans* + N. I DGCSS-korpusset, helt omvendt, optræder *hun* og *han* 20 gange så mange gange som *hendes* + N og *hans* + N. Det er den mest radikale fordeling i modsat retning. Det er store og tankevækkende forskelle. Teksterne i de to sidstnævnte korpusser er genremæssigt anderledes end i de to øverste, som jævnt hen består i ”almindeligt” skrevet konsumsprog: romaner, avisartikler, tekster fra diverse nyhedsmedier. Sammenhængen mellem genrer og fordelingen af *hun* og *han* vs. *hendes* + N og *hans* + N diskuteres kortfattet og tentativt hen ad vejen.

Den Danske Ordbog 1995: *hans* og *hendes*

Den samme undersøgelse blev udført i 1995 på *Den Danske Ordbogs* tekstkorpus på 40 millioner ord. Resultatet opsummeres i tabel 2 herunder. Det blev præsenteret på MUDS 5 i 1995 (Scheuer 1995). Resultatet kan beskrives i få ord. Efter *hans* kommer med størst sandsynlighed (afspejles i høj t-score) ord med relation til den offentlige magtsfære: positioner i hierarkier, betydelige åndelige produkter o.l. Efter *hendes* kommer ord med relation til kroppens fysiske attributter. Det er magten og ånden over for kødet, symbolsk kapital over for fysiognomi.

For nogle af ordenes vedkommende er det måske ikke uden videre indlysende at de skal indregnes under magt og ånd. En konkret metode til at afgøre dette er at slå dem op i KorpusDK og se i hvilke betydninger de bruges. *Billeder* kan være et eksempel. I KorpusDK refererer *hendes billeder* og *hans billeder* i langt de fleste tilfælde til produkter af kunstnerisk virksomhed, malerier eller fotografier osv., og den slags er resultater af åndeligt præget virksomhed. *Død* kan være et andet eksempel. I KorpusDK optræder *hendes død* og *hans død* først og fremmest i biografisk sammenhæng, og det er ikke hvem som helst der skrives biografisk om. Det er hovedsageligt samfundsmæssigt betydelige personer. Så i tvivlstilfælde – ideelt i alle tilfælde, men så langt er livet ikke – er metoden at rette ind efter de undersøgte hypotagmers betydning i videst udbredte brug – og ikke intuition eller spekulation (Et resultat af en søgning efter kontekster for *ide* i KorpusDK er anbragt som et illustrerende bilag til artiklen). En teoretisk begrundelse kan desuden hentes i Habermas' offentlighedsmodel, hvor det der i artiklen henføres til ånd og magt er placeret i det offentlige område (Habermas 1975).

Hans mening om hendes krop

<i>Hans</i>	Ordklasse	T-score	<i>Hendes</i>	Ordklasse	T-score
3 <u>folk</u>	s	11.72	hår	s	-9.27
5 sønner	s	8.25	øjne	s	-8.86
6 brødre	s	7.61	mor	s	-8.05
7 <u>værker</u>	s	7.29	kinder	s	-7.55
8 <u>mænd</u>	s	7.29	krop	s	-7.36
9 <u>opfattelse</u>	s	6.81	stemme	s	-7.21
10 <u>politiske</u>	adj	6.77	forældre	s	-6.92
12 <u>død</u>	s	6.55	far	s	-6.91
13 <u>kolleger</u>	s	6.53	ansigt	s	-5.96
15 <u>billeder</u>	s	6.06	hud	s	-5.47
17 <u>parti</u>	s	6.04	skød	s	-5.38
18 <u>regering</u>	s	5.99	lår	s	-5.06
19 <u>musik</u>	s	5.99	smil	s	-4.42
21 <u>politik</u>	s	5.46	mund	s	-4.39

Tabel 2. Resultatet af *hans- hendes*-undersøgelse i 1995.

Referencer til krop er skrevet med fed; referencer til ånd og magt er understregede.

Korpus 2000: *hendes* + N og *hans* + N

Med tilkomsten af Korpus 2000 bliver det nærliggende at udføre undersøgelsen igen og få et diakront resultat. Det er nemt at gøre; man kan bruge funktionen naboordssøgning. Jeg vil benytte lejligheden til at gøre opmærksom på DSL's *KorpusDK*, som er en fremragende søgefacilitet. Den er nem at bruge og giver adgang til store mængder af anvendt dansk sprog.

Resultatet præsenteres i tabel 3. Det adskiller sig fra resultatet i 1995 ved at der er kommet flere kropsord ind i mandesiden. Kvinderne fører dog stadig; på *hendes*-listen er der 14 N'er der betegner krop og kød, på *hans*-listen er der 8. Lige som i 1995 er samtlige ord for magt og ånd – på nær *liv* – knyttet til *hans*. Hvad magt og ånd angår, er resultatet ikke væsentligt anderledes end resultatet fra 1995, og på det område giver resultatet ikke anledning til store tanker om en udvikling af kønsrollerne.

Det skal nævnes at t-score udelukkende kan forstås som en relativ værdi. Det betyder ikke noget i sig selv at en t-score er af en vis størrelse. Men det betyder noget at t-score for et givet N er større

end t-score for et andet N, nemlig at tilbøjeligheden til at det optræder efter *hendes* hhv. *hans*, er større (Jørg Asmuss, korpusgrandmaster i DSL, personlig kommunikation). Man kan ikke ”gå på tværs” af skemahalvdelene. *Far* har ikke en større tilbøjelighed til at optræde efter *hans* end *krop* efter *hendes*. *Øjne* har ikke en større tendens til at optræde efter *hans* end efter *hendes*. Resultaterne er udregnet i to separate søgninger.

Nr.	<i>Hans</i>	t-score	Nr.	<i>Hendes</i>	t-score
2	far	19.28	2	krop	17.93
3	<u>liv</u>	18.63	3	øjne	16.58
4	øjne	18.47	4	mor	16.33
5	mor	17.51	6	far	14.41
6	familie	16.16	7	hånd	14.39
7	ansigt	15.17	8	ansigt	13.57
8	navn	15.07	10	<u>liv</u>	13.24
10	hånd	13.49	11	hænder	11.45
11	<u>mening</u>	13.44	13	mund	11.09
12	<u>død</u>	13.34	14	hår	11.05
13	krop	13.14	15	stemme	11.04
14	søn	13.11	16	hoved	10.79
15	stemme	13.05	17	datter	10.55
16	<u>kollega</u>	12.90	18	familie	10.24
17	hænder	12.73	19	forældre	10.05
18	hoved	12.48	20	søn	10.05
19	<u>parti</u>	12.27	24	blik	9.56
20	<u>kolleger</u>	12.17	28	hjerte	8.34
22	forældre	11.87	29	børn	8.31
23	venner	11.82	31	navn	8.19
25	<u>karriere</u>	11.78	32	kæreste	8.12
26	<u>ord</u>	11.29	33	ryg	7.96
27	blik	11.13	37	fingre	7.68
28	<u>film</u>	10.94	38	bryst	7.52

Tabel 3. *hans-hendes*-undersøgelse i Korpus 2000

Flere ord optræder i begge skemahalvdele, fx *liv*, *familie* og *forældre*. Det skyldes at t-scoren for de ord er høj for både *hans* + N og *hendes* + N.

Resultatet giver anledning til en principiel overvejelse. Valget af de nominer der regnes for at tegne forskellige betydningsmæssige paradigmer, er ikke ukontroversielt. At *talje*, *hår*, *mund* osv. tegner et fysisk, kropsligt billede, mens *politik*, *ideer*, *kontor* osv. tegner et magtfuldt og åndeligt billede, er ikke uden videre givet. Men jeg vil påstå at man her er på nogenlunde sikker grund. Det ligger alt andet end lige for at påstå at *politik*, *ideer*, *kontor* skulle tegne kropsligt billede – og omvendt (jf. Habermas 1975). Mere kontroversielt er det nok at *liv* er kategoriseret som relateret til ånd og magt. Årsagen er den samme som for *død*, se ovenfor. *Hans liv* og *hendes liv* optræder altovervejende i biografisk materiale. Og når der skrives om nogens liv, har denne person spillet en magtfuld rolle i samfundslivet.

Resultatet er rensat for forbindelser der kun kan optræde i forbindelse med *hendes* hhv. *hans*. Det er *hendes mand*, *hans hustru*, *hendes kusse*, *hans pik* m.fl. Det er derfor der er spring i numrene for ordenes placering på t-score-listerne, fx fra *søn* til *blik* i *hendes*-siden. Udrenset er – til forskel fra resultatet fra 1995, hvor forfatteren tydeligvis går direkte efter det mest flamboyante resultatet – også forbindelser hvor én eller ganske få kilder har leveret korpusmaterialet. En sådan leverandør til Korpus 2000 er netstedet *fyldepennen.dk*, et publikationssite for vordende (eller nok ikke-vordende i mange tilfælde) forfattere. *Fyldepennen.dk* er leverandør af forbløffende mange af de *hendes* + N der overhovedet lader sig anvende i den pornografiske genre, således eksempelvis *hendes lår*, *hendes hofter* og sågar *hendes læber*, som derfor er luget ud af resultatet. *Hendes kusse*, *hans pik* og lignende er i øvrigt også langt overvejende leveret af *fyldepennen.dk*. Rent anekdotisk kan det nævnes at én enkelt forfatter på *fyldepennen.dk* har produceret flere af disse forbindelser end nogen som helst anden. Jeg taler her om Gert Christoffersen, den måske mest indflydelsesrige tekstproducent i nyere dansk tekstkorpushistorie.

Forandringer fra Korpus 90 til Korpus 2000

Den Danske Ordbogs korpus findes ikke som isoleret korpus mere, men det er en del af Korpus 90. En skematisk beskrivelse af udviklingen fra Korpus 90 til Korpus 2000 vises i tabel 4 og tabel 5 herunder. Tabellerne viser forandringer i de højst rangerende ord med krops- og ånd- og magt-betydning på t-score-listerne. Et N der har bevæget sig opad, er placeret i kolonnen *mere knyttet*; det er blevet mere sandsynligt som N i *hendes* + N. Kolonnen *mindre knyttet* rummer de N'er som har bevæget sig nedad i sandsynlighed. *Ud* er N'er der har forladt listen af de højst rangerende; *ny* er dem der er kommet ind. Familie- og relationsbetegnelser er også taget med, jf. de nederste rækker i skemaet. Hvis de inddrages, har man med enkelte undtagelser fat i alle de N'er der har bevæget sig, og som derfor er interessante i en diakron sammenhæng.

Mere knyttet	Uændret	Mindre knyttet	Ud	Ny
krop		øjne	kinder	fingre
hånd		ryg	hals	bryst
hænder		hår	hud	<u>mening</u>
mund		stemme	fødder	
		hoved		
		blik		
		hjerte		
		<u>liv</u>		
familie		mor	ven	kæreste
forældre		far		børn
		søn		
		datter		

Tabel 4. Ændring af *hendes* + N fra Korpus 90 til Korpus 2000

Det er tydeligt nok at der er flest forandringer i de kropslige udtryk. Men tabellen viser ikke et klokkeklart resultat. En meget forsigtig fortolkning kunne være at de kropslige udtryk er for nedadgående – der er flere af dem i kolonnerne *mindre knyttet* og *ud* end i *mere knyttet* og

Hans mening om hendes krop

ny. Det svarer til at de bliver mindre sandsynlige som N'er i *hendes* + N. Påstanden kunne være at opfattelsen af kvinder som kød går tilbage. Men som man kan se i tabel 3, er den stadig dominerende.

Tabel 5 herunder viser de tilsvarende ændringer for *hans* + N.

Større tilknytning	Uændret	Mindre tilknytning	Ud	Ind
<u>liv</u>	øjne	ansigt	hjerte	<u>parti</u>
<u>mening</u>	krop	hånd	mund	<u>kolleger</u>
<u>død</u>		stemme	hår	<u>karriere</u>
		hænder	<u>tanker</u>	<u>film</u>
		hoved	<u>kollega</u>	
		blik		
familie	far	søn	datter	ven
		mor		
		forældre		
		venner		

Tabel 5. Ændring af *hans* + N fra Korpus 90 til Korpus 2000

Hvad det kropslige angår, er resultatet for *hans* + N det samme, men det er mere entydigt. N'erne der er kropslige udtryk, falder uden undtagelse i kolonnerne *uændret*, *mindre tilknytning* og *ud*. Så sandsynligheden for at mænd beskrives som krop, er utvetydigt faldende. Hos mændene er der også bevægelse i de N'er der relaterer til ånd og magt. Man finder flest af dem i kolonnerne *større tilknytning* og *ind*.

Man skal være opmærksom på at tabel 4 og 5 beskriver bevægelser. Tabel 2 og 3 viser det overordnede resultat: der er kommet flere kropsord ind i mandsiden, men kvinderne fører stadig klart. Samtlige ord for magt og ånd – på nær *liv* – knytter sig til *hans*. Hvad magt og ånd angår, adskiller Korpus 2000 sig ikke væsentligt fra 1995, og på det område giver resultatet ikke anledning til store tanker om en udvikling af kønsrollerne.

De familiært og socialt relaterede N'er ligger højt både for *hendes* og *hans*. Der er ingen asymmetri og ingen forandring i opfattelsen af ham og hende som sociale skikkelser. Men ellers er forandringerne fuldkommen systematiske, og konklusionen er nem at drage.

I resten af artiklen bruges familierelationer og sociale relationer som et paradigme i sig selv. Årsagen er at man finder nogle gedigne forskelle mellem beskrivelser af kønnene her. En teoretisk begrundelse kan igen tage udgangspunkt i Harbermas' offentlighedsmodel. Familie og sociale relationer er her skilt ud som det private områdes intimsfære (Habermas 1975; Mortensen 1975).

Udviklingen fra Korpus 90 til korpus 2000 er at mænd beskrives mere som magt og ånd og mindre som krop. Kvinder beskrives, uforandret, som krop. Der sker alt i alt en radikaliserings af den kønnede dikotomi. Snak om øget ligestilling ser ud til at være vrøvl.

TiDK

Vender man sig mod korpusset TiDK, får man resultater som de nedenstående. Korpusset indeholder knap 400 millioner ord fra polinfo og en række aviser og dagblade. Det tilgås med søgeværktøjet CoREST 2016 (corest.dsl.dk) som er anvendt i beta-versionen fra 2016 (foredragets tidspunkt). TiDK's indhold udskiftes én gang om året; det indeholder altid tekst fra de sidste ti år – deraf navnet. I denne undersøgelse er der anvendt tekst fra to perioder, 2006-2008 og 2013-2015. Det er et udsnit på 231 millioner ord. To intervaller er valgt med henblik på at undersøge diakront.

CoREST udregner ikke t-score. Sandsynligheden af givne *hendes* + N og *hans* + N udregnes ikke. Metoden i de nærværende beregninger har været simpelthen at optælle og udregne hyppigheden af forekomster af et antal N'er fra de paradigmer der har vist sig mest udslagsgivende i de andre søgninger, dvs. ånd- og magtrelaterede udtryk, kropslige udtryk og familie- og sociale relationsudtryk. Fordi der ikke er anvendt t-score-lister, er der en risiko for at nogle prægnante *hendes* + N og *hans* + N ikke er blevet fundet. Men det er ikke sandsynligt. Der er foretaget optællinger af mere end 100 N'er for hvert køn, så de nævnte paradigmer er blevet undersøgt i en meget større gruppe af ord.

Hans mening om hendes krop

CoREST er udviklet til ordbogsarbejde, så man kan søge efter lemmer. Den mulighed er jævnt hen blevet brugt i søgningerne.

<i>Hendes</i>		ratio		ratio	<i>hans</i>		ratio		ratio
N	n	2006-8		2013-15	N	n	2006-8		2013-15
stemme	1000	5,7	↑	6,0					
hår	200+	3,5	→	3,6					
krop	600	2,1	↑	2,5					
ansigt	500+	2,1	→	1,9					
mund	130	2,1	→	1,9	hoved	400+	1,4	→	1,4
ben	200	1,8	→	1,6					
øje	800	1,7	↓	1,3					
hånd	500	1,3	→	1,2					
					<u>mening</u>	500+	1,0	↑	1,5
barn	1100+	4,4	↓	2,8	<u>folk</u>	400+	3,4	→	3,2
datter	1100+	2,5	↓	2,2	<u>karriere</u>	800	1,0	→	1,1
søn	1500+	1,7	→	1,9	<u>opfattelse</u>	300	0,9	↑	1,2
far	3300	2,2	↓	1,8	<u>parti</u>	800	1,0	→	1,2
mor	2900	2,4	→	2,4	<u>politik</u>	280	1,7	↓	1,1
familie	2000	1,6	↓	1,3	<u>død</u>	900	1,2	→	1,1
forældre	1100+	2,7	↓	2,4	<u>værk</u>	700	1,8	↓	1,2
søster	200+	2,9	↓	2,6	<u>bog</u>	1300	1,0	→	1,0
bror	500+	1,3	→	1,4					
					<u>formue</u>	90	1,0	↑	1,6
<u>hus</u>	380	1,4	→	1,2	<u>ejendom</u>	83	1,3	→	1,5
kæreste	1400+	2,7	↓	2,1	<u>kollega</u>	1600	0,9	→	1,1
veninde	300+	16,7	↓	14,0	ven	1400+	15,4	↓	11,0

Tabel 6. Resultater af søgninger efter *hendes* + N og *hans* + N i TiDK ved hjælp af CoREST 2016

Tabellen er delt i to, en for *hendes* og en for *hans*.

n angiver antallet af fund. *Hendes stemme* er fundet 1000 gange.

Hans parti er fundet 800 gange.

Ratio betegner forholdet mellem hyppighederne af et givet *hendes* + N og *hans* + N og omvendt. Ratioen på 5,7 ud for *stemme* betyder at *hendes stemme* optræder 5,7 gange så hyppigt som *hans stemme*. Ratioen på 1,0 ud for *hans bog* betyder at *hans bog* og *hendes bog* optræder lige hyppigt.

I venstre side står alle ord som hyppigst optræder efter *hende*, i højre side ord som hyppigst optræder efter *hans* og desuden nogle ord som optræder tilnærmelsesvis lige hyppigt: *karriere*, *opfattelse*, *parti*, *død*, *bog* og *kollega*. De står der fordi de tilhører det samme paradigme som resten af ordene i det område, ord der betegner ånd og magt.

Tabellen viser at samtlige kropsligt relaterede udtryk optræder hyppigere, dvs. har en ratio større end 1 sammen med *hendes*, mens samtlige ånd- og magtrelaterede optræder hyppigst sammen med *hans* – med *hus*, som i parentes bemærket ikke er det allertydeligste eksempel på et magtrelateret N, som undtagelse og *kollega* lige på vippen. Dette gælder både i 2006-2008 og 2013-2015. Mænd er ånd og magt, og kvinder er krop, i begge målepunkter. Man kan notere sig at udtryk for familierelationer udelukkende optræder efter *hendes*. Kvinder beskrives som familie; de hører til i privatsfæren.

Pilene i tabellen angiver om ratioen for det pågældende ord er stigende, faldende eller uforandret fra 2006-2008 til 2013-2015. Opadgående pil viser at der er en stigning, nedadgående at der er et fald, vandret pil viser at der ikke er nogen bevægelse. En nedre grænse for hvornår der tegnes skrå pile er sat arbitrært ved bevægelser på 0,3.

Pilene viser fx at kvinder i faldende grad beskrives som mødre. N'et *barn* er det kraftigst faldende ord, men det relateres stadig i 2013-2015 mest til kvinder.

I det hele taget er ratioerne for udtryk for familierelationer efter *hendes* faldende. Men de er stadig gennemgående over 2. Kvinder betegnes som familiemedlemmer dobbelt så ofte som mænd og som forældre mere end dobbelt så ofte som mænd.

Interessant er det at ratioerne for magtudtryk ifm. *hans* er mindre end rationerne for krops- og familieudtryk ifm. *hendes*. Den ”kraft” med hvilken kvinder betegnes som krop og familiemedlemmer er stærkere end den ”kraft” med hvilken mænd betegnes som magtfulde og åndelige. Kvinder er krop og familie mere end mænd er ånd og magt.

De viste ord er udvalgt ud fra to kriterier. Det ene, og mest afgørende, er forekomstantallet, ’n’ i tabellen. 200 forekomster sat som den nedre grænse. Men der er også vist enkelte ord som er med kun for illustrationens skyld: *mund, formue* og *ejendom*.

Resultatet er i hovedtræk dette: for de fleste N’er fortsætter asymmetrien uforandret fra 2006-2008 til 2013-15. 17 pile går ligeud. For de næstfleste N’er er asymmetrien for nedadgående: 12 pile går nedad. Men generelt er asymmetrierne udtalte både i det tidlige og det sene målepunkt.

Både i TiDK i 2006-2008 og 2013-2015 beskrives mænd stadig som magt og ånd. Kvinder beskrives stadig som krop, og de er i høj grad også familie. Udviklingen fra 2006-2008 til 2013-2015 i TiDK er at de kønslige asymmetrier alt i alt bliver mindre, mest udtalt i beskrivelsen af kvinder som familiemæssigt relaterede. Men polariseringen i beskrivelsen af kønnene er stadig udtalt.

DGCSS

I DGCSS-databasen er der lavet en undersøgelse efter samme princip som ovenfor. Det er hyppigheds-ratioer der er fokuseret på. Ratioværdierne står i kolonnerne *Hans* og *Hendes*.

Resultaterne skal tages med det forbehold at nogle af ordene forekommer i meget lave antal (n-kolonnerne).

N	n	<i>Hendes</i>	<i>Hans</i>	N	n	<i>Hendes</i>	<i>Hans</i>
far	483		1,20	venner	65		3,18
mor	380	1,14		navn	65		2,90
søn	72		1,50	<u>arbejde</u>	42		2,55
bror	100		1,77	<u>lejlighed</u>	22		2,12
søster	82	1,87		<u>liv</u>	47		1,17
datter	62	1,26		<u>uddannelse</u>	27		1,17
børn	91	1,13		veninder	22	26,4	
familie	124		1,18	sprog	49	1,68	
forældre	268	1,26		stemme	30	2,93	
kusine	9	4,4		kæreste	61	1,00	
onkel	14		2,92				
søskende	18	1,01					
brødre	16		11,93				
lillebror	25		1,41				
lillesøster	17	1,12					
storebror	22		2,70				
storesøster	14	4,61					

Tabel 7. Resultater for *hendes* + N og *hans* + N i DGCSS's database med 10 millioner ord talt sprog

DGCSS-korpuset er det korpus blandt de fem anvendte som har den suverænt ligeligste fordeling af *hendes* + N og *hans* + N (se tabel 1). Men fordelingen af *hun* og *han* er den almindelige: 45.000 *hun* mod 70.000 *han*. Det betyder at kvinder også her omtales mindre end mænd med en solid forskel. Mænd omtales halvanden gang så meget.

Et spøjst resultat er her at ord som betegner familierelationer – hvoraf visse er uhyre talrige – bruges så der dannes kønsmæssigt homogene grupper. *Hans* er knyttet sammen med *far, søn, bror, onkel* og *brødre*. *Venner* scorer også højt. *Hendes* er knyttet sammen med *mor, søster, datter, kusine, storesøster* og *veninder*. *Søstre* er meget lavfrekvent i basen; *hendes søstre* er sjældent (4 forekomster). Så en mand og hans mandlige relaterede er en ofte omtalt gruppe, og ligeså er en kvinde og hendes kvindelige relaterede. Der er ingen forskel i om mænd og kvinder relateres til intimsfæren.

DGCSS-korpusset indeholder store mængder af tekst som består i narrativer fortalt af informanter (betegnelse i sociolingvistik) og meddelere (betegnelse i dialektologi) om deres egne liv. I personlige fortællinger om det egne liv refererer *han* og *hende* og *hans* og *hendes* til personer i det net af nære personer som fortælleren placerer sig i. Der tales ikke om distante skikkelser i et offentligt rum, men om proksimale personer i et mere nært, mere privat rum. Sådanne personer er prototypisk familie.

DGCSS-korpusset har af alle korpusser den relativt laveste forekomst af *hans* og *hendes* ift. *han* og *hende*. Det kan skyldes genren. Meget er narrativer om fortiden, og her kan *hendes* og *hans* indimellem undværes. Hvis nogen siger *Onklen var helt sindssyg*, ved man udmærket hvem der tales om hvis der fortælles om en bestemt person: denne persons onkel, naturligvis.

Man kan notere sig at resultaterne, med *stemme* som undtagelse, er fuldstændig barberet for N'er der refererer til det kropslige. Det kan skyldes at narrativerne for det meste er fortalt til fremmede, interviewere og dialektoptegnere o.l., som man ikke taler om den slags med.

Den lave forekomst af *hans* + N og *hendes* + N relaterer formentlig både til talesituationen og genren.

I Danmarks Grundforskningsfonds Center for Sociolingvistiske Sprogforandringsstudiers (DGCSS) database på 10 millioner ord transskriberet talesprog fortælles der meget mere om mænd end om kvinder. Ingen af kønnene beskrives ved hjælp af kropslige attributter; begge køn beskrives som medlemmer af kønsligt homogene grupper, oftest af familiekarakter.

NFI's korpus med 2 millioner danske tweets

I twitter-korpusset er fordelingen af *hendes* + N og *hans* + N som i tabellen herunder. Der vises nu to oplysninger, frekvens og PMI.

PMI er en forkortelse for *pointwise mutual information*, et mål for sandsynligheden for at to ord optræder sammen. PMI er ligesom t-score. Værdierne betyder ikke noget i sig selv, men i forhold til hinanden. Jo højere (dvs. nærmere på 0) PMI desto højere sandsynlighed for forekomsten af en given kollokation.

Værdierne er udregnet i én søgning, så her er værdierne for *hendes* + N og *hans* + N direkte sammenlignelige. Når *hendes veninde* har en PMI på -2,37342 og *hans venner* har en PMI på -3,44208 er sandsynligheden for at *veninde* optræder sammen med *hende* større end sandsynligheden for at *venner* optræder sammen *hans*. *Hans navn* har en PMI på -2,96024, mens *hendes navn* har en PMI på -2,69187, og det betyder at *navn* optræder med størst sandsynlighed sammen med *hendes*.

Mand og *kone* er ikke taget ud her fordi det giver en anden type oplysning end i t-score- og ratioberegninger. Af tabel 8 fremgår at *mand* er mindre sandsynlig som N efter *hendes* end fx *stemme*. Der skrives altså med større sandsynlighed *hendes stemme* end *hendes mand*. Det er ikke redundant viden.

Hans mening om hendes krop

<i>Hendes</i>	PMI	n	<i>Hans</i>	PMI	n
kæreste	-1,82687	95 ^f	navn	-2,96024	102
mor	-2,02404	78 ^f	stemme	-3,29476	73
stemme	-2,30321	59 ^f	kæreste	-3,29476	73
veninde	-2,37342	55 ^f	mor	-3,32254	71
far	-2,37342	55 ^f	far	-3,35111	69
mand	-2,46873	50 ^f	venner	-3,44208	63
navn	-2,69187	40 ^f	<u>musik</u>	-3,57788	55
fødselsdag	-2,71719	39	kone	-3,65339	51
forældre	-2,71719	39	ven	-3,67319	50
familie	-2,76983	37	familie	-3,67319	50
hår	-2,82540	35	hår	-3,71401	48
<u>musik</u>	-2,94676	31	ansigt	-3,80103	44
søn	-3,01345	29	<u>film</u>	-3,84755	42
ansigt	-3,04855	28	søn	-3,92165	39
skyld	-3,12265	26	skyld	-3,92165	39
datter	-3,12265	26	<u>sange</u>	-3,92165	39
veninder	-3,43631	19	forældre	-3,92165	39
tøj	-3,43631	19	<u>liv</u>	-4,00170	36
børn	-3,43631	19	<u>karriere</u>	-4,02987	35
<u>arbejde</u>	-3,20270	24	nummer	-4,08871	33
<u>liv</u>	-3,24526	23	øjne	-4,15123	31
side	-3,49038	18	<u>tid</u>	-4,15123	31
<u>bog</u>	-3,49038	18	side	-4,21792	29
bryster	-3,60816	16	bil	-4,25301	28
værelse	-3,67270	15	<u>bog</u>	-4,28938	27
vegne	-3,67270	15	<u>hold</u>	-4,32712	26
søster	-3,67270	15	seng	-4,36634	25

Tabel 8. Pointwise mutual information for *hendes* + N og *hans* + N i NFI's twitter-korpus

De 12 mest sandsynlige forbindelser for hvert køn er skrevet på grå baggrund. Der er ikke umiddelbart den store betydningsmæssige forskel på hvilke N'er der knyttes til *hendes* og *hans*. Der er hverken mere kød eller magt i nogen af siderne. *Hans musik* er på top ti, men *hendes musik* er på 12.-pladsen i *hendes*-kolonnen. Der er for både *hendes* og *hans* en del benævnelser for sociale relationer, familie og andet. Kæreste ser ud til at være en vigtig relation, *hendes kæreste* står øverst hos hende og tredjeøverst hos ham.

Det er betydningsfuldt at sandsynligheden for *hendes kæreste* er større end sandsynligheden for en hvilken som helst *hans* + N. Det er mere sandsynligt at der skrives *hendes kæreste* end at der skrives *hans* + noget som helst i tweets. De 12 første N'er i *hendes* + N optræder med større sandsynlighed end noget som helst N i *hans* + N. Det er mere sandsynligt at kvinder beskrives med 12 bestemte N'er end at mænd beskrives med et hvilket som helst N i *hendes* + N- og *hans* + N-forbindelserne i tweets.

Blandt disse 12 ord i *hendes* + N er en del familie- og relationsbetegnelser (*kæreste*, *mor*, *veninde*, *far*, *forældre*, *familie*). Det er igen intimsfæren der dominerer i beskrivelsen af kvinder. Sandsynligheden for at kvinder omtales som socialt relaterede, er større end sandsynligheden for at mænd omtales på nogen som helst måde.

Dirk Hovy, korpusgrandmaster for NFI's tweet-korpus, har testet to hypoteser. En hypotese om at *hendes* efterfølges af en relationsbetegnelse, familiebetegnelse eller anden understøttes statistisk signifikant. Der er dokumentation for denne hypotese. En hypotese om at *hendes* efterfølges af en betegnelse for noget kropsligt understøttes, men ikke signifikant. *Hendes* + noget kropsligt scorer højt på sandsynlighedslisten, men ikke så systematisk at det er statistisk signifikant.

Det er påfaldende at *han* og *hun* i twitter-korpusset er langt mindre frekvente end *hans* og *hendes* (jf. tabel 1). Deltagerne ved foredraget på MUDS mente at have gjort den observation at tweets handler meget om afsenderen selv. Pronomen i 1. person (dvs. *jeg*) har-

monerer med den strategi, hvorimod pronomener i 3. person nominativ ligefrem kan stille sig i vejen. Måske er sagen at *han* og *hun* blokerer muligheden for *jeg* som subjekt mere end *hans* og *hendes* gør. Det er en grammatisk årsag, og der kan være en parallel pragmatisk årsag. Måske er det fordi *hans* og *hendes* ikke nødvendigvis flytter fokus væk fra afsenderen. *Jeg elsker hendes smil* eller bare *elsker hendes smil* handler om afsenderens mening.

I NFI's korpus med 2 millioner danske tweets omtales kvinder som personer i sociale relationer.

Hvem skriver de 12 mest hyppige ord om kvinder?

En egenskab ved NFI's twitter-korpus er at skribenternes køn kendes med stor sikkerhed. Man kan regne ud om det er mænd eller kvinder der med størst sandsynlighed har produceret de forskellige *hendes*- og *hans*-forbindelser. Kønnen er afgjort ved opslag i en fornavns-ordbog genereret ud fra et stort trustpilot-korpus, hvor brugere har oplyst deres køn binært, dvs. som mand eller kvinde. Brugerne af trustpilot oplyser deres køn binært: han- eller hunkøn.

Tabellen herunder viser i første kolonne de 12 ord som med størst sandsynlighed er skrevet af ét køn. *Kæreste* er det ord der med størst sandsynlighed er skrevet af ét køn. Sandsynligheden er udregnet efter samme princip som PMI; målet kan bruges til sammenligninger, ikke andet.

At *navn* optræder to gange, er ikke en fejl. Det skyldes at sandsynligheden for at en kvinde skriver *navn* og at en mand skriver *navn* begge er blandt de 12 mest sandsynlige udfald.

N	Skrevet af	PMI	n	N'er i <i>Hendes</i> PMI-rangering
kæreste	F	-2,54649	105	1
stemme	F	-2,73455	87	3
mor	F	-2,79374	82	2
navn	F	-3,17510	56	7
far	F	-3,21147	54	5
ansigt	F	-3,30863	49	14
familie	F	-3,35031	47	10
navn	M	-3,35731	67	7
venner	F	-3,41627	44	9
veninde	F	-3,61694	36	4
hår	F	-3,61694	36	11
<u>musik</u>	F	-3,64511	35	12

Tabel 9. N'er i NFI's twitter-korpus som med størst sandsynlighed er produceret af et bestemt køn

Tabellen viser at de 11 ord med størst PMI er skrevet af en kvinde. Faktisk skal man ned på 21.-pladsen i listen over de til køn som afsender mest knyttede ord før der kommer et N der med den største sandsynlighed er skrevet af en mand; det er ordet *film*.

Tabellen viser også, i den sidste kolonne, hvilken plads ordet har på sandsynlighedslisten over de N'er der efterfølger *hendes* (tabel 8). Her optræder nummer 1-5, 7 og 9-12. Nummer 6 og 8 ligger uden for den viste liste, men ord nummer 6 og 8, *mand* og *fødselsdag*, er også med størst sandsynlighed skrevet af en kvinde. Det kan man ikke se i tabel 9 men længere nede på PMI-listen. Samtlige de N'er i tabel 8 der med størst sandsynlighed kommer efter *hendes*, er med størst sandsynlighed skrevet af kvinder.

Så tabellerne over twitter-data viser at den monotone og konservative beskrivelse af kvinder sandsynligvis udføres af kvinder!

Blandt de resultater der optræder længere nede på PMI-listen over afsenderkøn end hvad der er vist i tabellerne, er to iøjnefaldende: 1) Samtlige sociale relationsbetegnelser: familie, venner osv. inden for de øverste 90(!) ord er mest sandsynligt skrevet af kvinder 2) Samtlige

betegnelser for fysiognomi inden for de samme 90(!) ord er mest sandsynligt skrevet af kvinder.

Det sidstnævnte er et godt resultat. Kvinderne blandt twitterbrugerne fokuserer sanseligt på fysik mere end mændene gør. Det er ikke et konservativt resultat.

I NFI's korpus med 2 millioner danske tweets beskrives kvinder monotont og konservativt som personer i sociale relationer, og ikke mindst som kærester. Beskrivelsen udføres af kvinder.

Twitter-korpusset indeholder tweets fra fem lande: Danmark, Holland, Frankrig, England og Tyskland. I artiklen her er kun det danske materiale anvendt, men der vil let kunne laves søgninger på tværs af lande.

Konklusioner

Resultaterne af søgningerne hvad angår køn, er skrevet med fed skrift for hvert korpus ovenfor. En hurtig opsummering, som altså er resultatet af søgninger i omkring 400 millioner ords tekst i forskellige genrer og produktionsformater, skrevne, talte, tweetede, kan være denne:

- Kvinder behandles ikke som lige så vigtige og interessante som mænd. Tekster og tale handler mest om mænd, uanset genre.
- Kvinder beskrives som kropslige. Mænd beskrives som åndelige og magtfulde.
- Kvinder beskrives som relationelle. De er placeret i net af familierelationer og andre sociale relationer i højere grad end mænd er.
- Med hensyn til køn er resultatet konservativt og asymmetrisk uanset hvilket korpus man søger i.
- Snak om øget ligestilling ser ud til at være vrøvl.
- Genre betyder meget – endog rigtig meget – for de resultater der kommer ud af korpus-søgninger.
- Selv meget store korpusser kan være sårbare over for at visse genrer er over- eller underrepræsenterede.
- Man er tvunget til at læse søgningsresultater i deres kontekst hvis man skal være nogenlunde sikker på resultaternes art og repræsentativitet.

En bemærkning om køn

At køn er en social konstruktion, og at der er flere end to, er ikke noget problem i den nærværende sammenhæng. At der ude i virkeligheden findes mennesker der ikke føler sig og ikke ønsker at blive betragtet simpelt som mand eller kvinde, og som ikke definerer sig binært ud fra det sæt af kropslige og magtfulde og åndelige karakteristika som beskrives af de N'er som følger efter *hendes* og *hans* i korpusserne, er ikke et problem.

Tekstkorpusser afbilder helt overvejende nemlig ikke virkeligheden på så kompleks en måde. Når der bruges *hendes* og *hans*, kan man være helt tryk ved at der i mere end 99 % af tilfældene refereres til hvad almindelige mennesker – og hermed menes ikke andet end de fleste – tænker om køn, og det er som et binært fænomen. Det er almindelige mennesker – og hermed menes stadig ikke andet end de fleste – der har produceret mere end 99 % af teksterne.

Det er ikke *min* påstand at *hendes* refererer til ét af to køn og at eksempelvis *hendes talje* skal læses som en beskrivelse af kvinder. Det er derimod min påstand at det er den verdensopfattelse *tekstproducenterne* har. Og derfor er jeg helt tryk ved at betegne det der refereres til med *hendes* og *hans*, som *mænd* og *kvinder*.

Litteratur

GDS = Erik Hansen & Lars Heltoft (2011) *Grammatik over det Danske Sprog*. Odense: Syddansk Universitetsforlag.

Habermas, Jürgen (1975) *Borgerlig offentlighed - dens framvekst og forfall, henimot en teori om det borgerlige samfund*. København: Fremad.

Hansen, Erik (1977) *Dæmonernes Port*. København: Hans Reitzel.

KorpusDK = Det Danske Sprog- og Litteraturselskabs database KorpusDK på 56 millioner ord og herunder Korpus 90 og Korpus 2000 på hver 28 millioner ord, kan tilgås gennem ordnet.dk.

Mortensen, Frands (1976) *Ytringsfrihed og offentlighed. De hemmelige dokumenter fra Danmarks Radios afskedigelse af Erik Thygesen*. Århus: Modtryk.

Hans mening om hendes krop

PMI = https://en.wikipedia.org/wiki/Pointwise_mutual_information (tilgået oktober 2016).

Scheuer, Jann (1995) Hans hustru og hendes bryster. 5. *Møde om Udforskningen af Dansk Sprog*, 13.-14. oktober 1994:246-256.

TiDK og CoREST 2016 = Det Danske Sprog- og Litteraturselskabs TiDK på knap 400 millioner ord, som kan tilgås gennem søgemaskinen CoREST 2016, kan findes på corest.dsl.dk.

Vatvedt Fjeld, Ruth (2015) Om ordbokseksempler og stereotypisering av kjønn i noen nordiske ordbøker. *Perspektiv på lexikografi, grammatik och språkpolitik i Norden*. Caroline Sandström, Ilse Cantell, Eija-Riitta Grönros, Pirko Nuolijärvi & Eivor Sommerdahl (udg.). Helsinki: Institutet för de inhemska språken:35-65.

Bilag: kontekster for ordet *idé*

Kontekster for de første 5 søgeresultater for ordet *idé* i KorpusDK

Det fælles ide og udviklingsforum omkring informationsteknologi i arbejderbevægelsen

Statsminister Poul Schlüters ide om at genforhandle EFs landbrugsreform efter en samlet GATT-løsning bliver modtaget med interesse i Frankrig.

Det er fornuftigt at regeringen at tage fat på arbejdsgiverafgifterne, som netop er blevet forhøjet af et flertal i Folketinget og som forøger de indirekte lønomkostninger. Det er en god ide at sænke disse, understreger Hans Skov Christensen.

De to såkaldte EF-bemyndigelseslove skrives sammen til en, og alle er enige om, det er en god ide.

Maskinmester Henning Houkjær, Orehoved, har nu udsendt en projektskitse for oprettelse af en lystbådehavn i Orehoved. Hensigten er at få tilbagemeldinger fra interesserede bådejere og derved få en ide om, hvor mange bådpladser, der vil blive brug for.