

# Køn i korpus

## Semantisk asymmetri i brugen af ordene *kvinde* og *mand*, *pige* og *dreng* i danske avistekster

IDA FLÖRKE OG ELISABETH ENGBERG-PEDERSEN

### ABSTRACT

På baggrund af de seneste ca. 30 års udvikling i forholdet mellem kvinder og mænd i Danmark har vi lavet tre diakrone undersøgelser af forskellige nedslag i omtalen af kvinder, mænd, piger og drenge i Det Danske Sprog- og Litteraturselskabs korpus af især avistekster. Den første undersøgelse viser en beskeden udligning i omtalen af kvinder i forhold til mænd bedømt ved antal forekomster af pronomenerne *hun* og *han* i *Berlingske (Tidende)* fra perioden 1990-1994 til 2018. Den anden undersøgelse viser at voksne kvinder omtales ved *pige* en del hyppigere end voksne mænd omtales ved *dreng*, med et lille fald fra 1990'erne til 2018. Den sidste undersøgelse ser på attributive adjektiver til *kvinde*, *mand*, *pige* og *dreng* i perioden 1930-2018 og kun i 2018 og viser klare forskelle i beskrivelsen af de to køn: Piger og kvinder karakteriseres især ved deres udseende, mænd ved deres styrke, handlekraft og moralske habitus. Piger omtales ofte som lyshårede, søde og artige, drenge som frække og slemme. Mænd karakteriseres især i forhold til magt og rigdom, som voldsudøvere, alkoholikere og kriminelle, kvinder og piger som ofre for vold og sygdomme. Den klareste forskel mellem 2018 og hele perioden 1930-2018 er at mænd i 2018 også beskrives ved deres attraktivitet og deres seksualitet.

**EMNEORD:** korpusundersøgelse, køn, avistekster, diakron semantik

### 1 KØN I TIDLIGERE KORPUSUNDERSØGELSER

Forekomsten af substantiver brugt om kvinder og mænd blev undersøgt i et dansk korpus så tidligt som 1912 (Jacobsen 1912). Men muligheden for at søge elektronisk i meget store korpusser og forbedrede muligheder for diakrone studier giver os i dag et langt mere omfattende billede af sprogbrugen på forskellige tidspunkter. Anledningen til de følgende analyser af sprogbrugen i forskellige korpusser var en umiddelbart oplevet skævhed i brug af *pige* og *dreng* om voksne: Hvor højt op i alder bruges *pige* om voksne af hunkøn, og hvor gammel kan man

være og stadig blive kaldt *dreng*? Det følgende chatindlæg fra 2012 viser at i hvert fald nogle sprogbrugere er usikre på brugen:

Hej tøser

Jeg har længe tænkt på noget...

Hvornår er de drenge? Hvornår er de fyre og hvornår er de mænd?

For min kæreste er nu 20... Og jeg ved ikke hvad jeg skal "kalde"

ham Han kalder mig for "min pige" men det lyder SÅ forkert at

sige min dreng... Og det lyder som et one-night stand med min fyr

(hvilket det ikke er, for vi har været sammen i snart 2½ år), og

næsten gammelt og gift med min mand...

Er jeg fjollet? (<https://forum.woman.dk/debatten/kaerlighed/dreng-fyr-eller-mand>)

Undersøgelsen af aldersspændet for brug af *pige* og *dreng* om henholdsvis kvinder og mænd blev suppleret med dels en undersøgelse af forekomsten af ordformerne *han* og *hun* i ikke-generisk brug i to perioder og dels en undersøgelse af hvilke adjektiver *kvinde*, *mand*, *pige* og *dreng* kædes sammen med i tekster.<sup>1</sup>

I undersøgelsen af omtalen af kvinder og mænd fokuserede Jacobsen (1912) på runeindskrifter og lovhåndskrifter i middelalderen og fandt at *Mand* blev brugt i betydningen 'menneske', *Kvinde* i betydningen 'gift hunkønsperson', og *Pige* i betydningen 'ung ugift hunkønsperson'. Mænd blev desuden omtalt ved deres funktion i samfundet, og kvinder ved deres status som bl.a. *Mø*, *Brud* og *Enke*. Det billede der tegnes her, er altså at mænds status hænger sammen med det offentlige liv, mens kvinders hænger sammen med deres status i familien og i forhold til mænd.

I amerikansk engelsk brugte Bebout (1984, 1995) spørgeskemaer i to omgange til at undersøge hvor unge eller gamle mennesker kan være når de omtales som enten *girl/boy* eller *woman/man*. Hun fandt bl.a. at minimumsalderen for hvornår respondenterne mente de kunne omtale

1 I lighed med tidligere undersøgelser taler vi i denne artikel om to køn. Det betyder ikke at vi tager stilling til debatten om kønsvariationer, kønsidentitet og kønsforståelser (se fx Kerrn-Jespersen 2016), kun at vi følger den mest udbredte sprogbrug i den periode vi undersøger, hvor individer der benævnes ved *pige* eller *kvinde*, også typisk omtales som *hun*, og individer der benævnes ved *dreng* og *mand*, også typisk omtales som *han*.

en person som *woman*, faldt fra 20 år i 1980'erne til 16 år i 1990'erne. Maksimumsalderen for at omtale en person som *girl* i 80'erne var for ca. halvdelen af de adspurgtes vedkommende 23 år eller mere, mens ingen mente at man kunne bruge *boy* om et individ der var ældre end 16 år. Her spiller det muligvis ind at *boy* er blevet brugt i amerikansk engelsk i nedsættende om- og tiltale af afroamerikanske mænd (*Merriam-Webster*).

Sigley og Holmes (2002) undersøgte brugen af *girl* og *boy* i bl.a. to korpusser med britisk engelsk fra 1961 og 1991, hver bestående af 500 tekster af 2.000 ord valgt inden for en række skriftsprogsgenrer. De fandt at 20,8 % af anvendelserne af *boy* i 1961 handlede om voksne mænd og 16,9 % i 1991, hvor *adult* er defineret ved om teksten refererede til "individuals in sexual relationships" (Sigley & Holmes 2002: 144). Til sammenligning handlede 62,6 % af anvendelserne af *girl* om voksne kvinder i 1961 og 51,6 % i 1991. Der var altså meget større sandsynlighed for at en voksen kvinde blev benævnt ved *girl*, end at en mand blev benævnt ved *boy* både i 1961 og i 1991, selv om der var tale om et fald fra 1961 til 1991.

Baker (2014) brugte de samme to korpusser til undersøgelser af frekvensen af en række kønsmarkerede ord. Han udvidede undersøgelsen både 30 år bagud til 1931 og 15 år frem med et webbaseret korpus fra 2006 og fandt at brugen af *girl* og *boy* om voksne, som han ikke definerer, var steget kraftigt i 2006 idet *girl* dog stadig blev brugt oftere om voksne kvinder end *boy* om voksne mænd. *Girl* om voksne kvinder var ofte forbundet med sex og prostitution, *boy* om voksne mænd med lavt betalte job, farlig tiltrækning og umodenhed. Baker (2014) fandt desuden en nedgang i brugen af hankønspronomen (*he, his, him*) fra især 1961 til 2006. Brugen af hunkønspronomen (*she, her, hers*) steg især fra 1931 til 1961 og derefter kun ganske lidt. Den absolutte frekvens af de to typer pronomen i 2006 er næsten 1,75 gange så mange hunkønspronomen som hunkønspronomen.

Sigley og Holmes (2002) så også på kollokationer med *girl* og *boy* og fandt at i 1961 fokuserede modifikatorerne til *girl* (fx *pretty, attractive*) på udseende syv gange oftere end modifikatorerne til *boy*. Denne tendens var reduceret til dobbelt så ofte i 1991. Ved at se på de hyppigste adjektiver ved *woman* og *man* fandt Baker (2014) ikke den store forskel

mellem de to køn i fire nedslagspunkter mellem 1931 og 2005, bortset fra at der blev brugt langt flere forskellige adjektiver om mænd end om kvinder. Grunden hertil er sandsynligvis at mænd blev omtalt langt hyppigere end kvinder: *Man/men* blev brugt tilsammen ca. fire gange så ofte som *woman/women* i 1931 og godt dobbelt så ofte i 2006. Baker så også på hvilke adjektiver der kun forekom sammen med enten *woman* eller *man*. Både kvinders og mænds udseende beskrives ovevejende ved negative, men altså forskellige adjektiver i 1931 og overvejende ved positive – og forskellige – adjektiver i 2006. I alle årene beskrives mænd, og ikke kvinder, som magtfulde og fysisk stærke med adjektiver som *celebrated*, *distinguished* og *energetic*, mens kvinder, og ikke mænd, beskrives som attraktive med adjektiver som *beautiful*, *desirable* og *sensuous*.

Pearce (2008) så også på kollokationer med *woman/women* og *man/men*, denne gang i *British National Corpus* på 100 mio. ord, heraf 10 % talesprog fra forskellige genrer. Hovedparten af både de skrevne tekster og talesprogsteksterne stammer fra 1970'erne til starten af 1990'erne. Pearce gik et skridt videre end de andre undersøgelser ved ikke blot at undersøge substantivernes kollokationer med adjektiver, men også at opdele deres kollokationer med verber efter om substantiverne forekom i subjekter (fx *The man build (something)*; *The woman consented*) eller objekter (fx *(Someone) arrested the man*; *(Someone) assaulted the woman*). Blandt de mange resultater kan man nævne at *man* som subjekt er stærkere forbundet med verber i hvis betydning der indgår fysisk styrke, fx *assault*, *climb* og *dig*, og *man* som både subjekt og objekt er stærkere forbundet med verber der har med kriminalitet og det juridiske system at gøre (fx *rape* og *strangle* som subjekt og *handcuff* og *convict* som objekt). Pearce fandt også at *man* som subjekt er stærkere forbundet med verber om magtudfoldelse eller -besiddelse (fx *dominate* og *lead*) og med verber for besiddelse (fx *possess*, *own*) end *woman* er. Desuden er *man* stærkere forbundet med adjektiver for samvittighedsfuldhed (fx *ernest*, *prudent*) og for intelligens (fx *clever*, *gifted*). *Woman* er stærkere forbundet med adjektiver som *resourceful* og *strong-minded* end *man*, men også med *daft* og *dumb*.

Holmes og Sigley (2001) fandt at frekvensen af *woman* steg til det dobbelte fra 1961 til 1991 samtidig med at frekvensen af *man* faldt i en

sådan grad at der er større brug af leksemet *woman* end leksemet *man* i 1991. Pearce (2008) fandt imidlertid at flertalsformen *women* var næsten dobbelt så frekvent som entalsformen *woman*, mens entalsformen *man* var mere end halvanden gang så hyppig som flertalsformen *men* i *British National Corpus*, dvs. op til starten af 1990'erne. Det betyder at individuelle mænd omtaltes langt hyppigere end individuelle kvinder, der oftere blev anskuet som en gruppe.

De engelske data peger altså på en vis udligning i frekvensen af omtalen af de to køn, men især Pearces (2008) og Bakers (2014) undersøgelser peger på store forskelle i de egenskaber kvinder og mænd karakteriseres ved i britisk engelsk så sent som i slutningen af 1900-tallet og i 2006.

I en dansk sammenhæng har Scheuer (2017) undersøgt forekomsten af *hun* og *han* og kollokationer med *hendes* og *hans* + et substantiv inden for en afstand af to ord i fire korpusser, bl.a. *KorpusDK* (Det Danske Sprog- og Litteraturselskab 2021) med de to delkorpusser *Korpus 90* og *Korpus 2000*. Uanset korpus forekommer *han* ca. dobbelt så mange gange som *hun*. Det samme gælder *hans* + substantiv i forhold til *hendes* + substantiv bortset fra i et rent talesprogskorpus (DGCSS 2021), hvor der kun er lidt flere forekomster af *hans* + substantiv end af *hendes* + substantiv. Scheuer sammenligner resultaterne med de resultater han selv fik ved en undersøgelse af *Den Danske Ordbogs* (DDO) oprindelige tekstkorpus, *Korpus 90* (Scheuer 1995). I 1995 fandt Scheuer at "[e]fter *hans* kommer med størst sandsynlighed [...] ord med relation til den offentlige magtsfære [...]. Efter *hendes* kommer ord med relation til kroppens fysiske attributter" (Scheuer 2017: 346). I *Korpus 2000* er der kommet flere kropsord ind efter *hans*, men der er stadig flere efter *hendes*. Mht. ord Scheuer forbinder med magt og ånd, er resultatet stort set det samme i *Korpus 2000* som i *Korpus 90*. De to korpusser er imidlertid ikke umiddelbart sammenlignelige idet *Korpus 2000* indeholder langt mere avisprog end *Korpus 90*.

I det følgende vil vi gøre rede for tre undersøgelser som kan vise om der er sket ændringer i sprogbrugen i perioden 1990-2018, dvs. fra Scheuers (1995, 2017) undersøgelser af sprogbrugen omkring 1990 og 2000. Siden år 2000 er flere kvinder blevet synlige på magtfulde poster i samfundet. Den første kvindelige formand for direktionen i

Danmarks Nationalbank besad posten fra 1995-2005, den første kvindelige statsminister i Danmark tiltrådte i 2011, en række fonde under A.P. Møller-Mærsk fik en kvindelig formand i 2012, Folketinget fik sin første kvindelige formand i 2015, i 2021 har alle fire regeringsbærende partier en kvindelig formand, og formanden for yderligere et parti i Folketinget er en kvinde. Internationalt er flere kvinder også blevet synlige på magtfulde poster som den første kvindelige kansler i Tyskland i 2005 og den første kvindelige formand for Europa-Kommissionen i 2019. I 2017 bredte der sig blandt kvinder en bølge af protester mod seksuelle krænkelse i det meste af den vestlige verden under navnet *#MeToo*. I Danmark slog denne bølge først for alvor igennem i 2020, hvilket imidlertid er for sent til at vi kan undersøge dens effekt på sprogbrugen i de tilgængelige korpusser. Ikke desto mindre mener vi at det er værd at undersøge omtalen af kvinder, mænd, piger og drenge i dag for at se om den sprogbrug som man har fundet i tidligere danske korpusser og i internationale korpusundersøgelser, har ændret sig.

Den første undersøgelse er en diakron undersøgelse af forekomsten af de personlige pronomener *hun* og *han* i to avis-korpusser fra første halvdel af 1990'erne og fra 2018; den giver et førstehåndsindtryk af udviklingen over 20-25 år (afsnit 3). Derefter følger endnu en diakron undersøgelse, der viser hvilke aldersgrupper *pige* og *dreng* bruges om i avissprog i de to perioder (afsnit 4). Den tredje undersøgelse (afsnit 5) er en samlet undersøgelse af kollokationer med *kvinde*, *mand*, *pige* og *dreng* i det største korpus vi har haft adgang til; dette korpus dækker perioden 1930-2018 med en særlig stor mængde tekster fra 1990-2018. Vi inddrager et diakront aspekt i denne undersøgelse ved også at undersøge kollokationer i en delmængde af det store korpus, nemlig kun teksterne fra 2018, og sammenligne dem med hele periodens kollokationer. Endelig diskuterer vi resultaterne og deres begrænsninger i det afsluttende afsnit 6. Først vil vi imidlertid præsentere de korpusser som bruges i undersøgelserne.

## 2 KORPUS FOR DE TRE UNDERSØGELSER

Grundlaget for undersøgelserne er Det Danske Sprog- og Litteraturselskabs korpus *Bakspejlet*, som ikke er offentligt tilgængeligt, men som vi har fået adgang til. Det offentligt tilgængelige *KorpusDK* udgør en del-

mængde af *Bakspejlet* og er et statisk korpus bestående af de to delkorpusser *Korpus 90* og *Korpus 2000* (<https://ordnet.dk/korpusdk>). Det er disse to korpusser Scheuer (1995, 2017) brugte til sine undersøgelser. Et billede af hvad *Bakspejlet* omfatter, kan man få på websiden *CoREST 2021*, men her er det stadig kun en delmængde af hele *Bakspejlet* der er tilgængelig.

Det primære formål med *Bakspejlet*, som suppleres løbende, er at danne grundlag for videreudviklingen af *Den Danske Ordbog* (<https://ordnet.dk/ddo/forside>) som en almen ordbog over dansk. Da *Bakspejlet* blev grundlagt med dette formål, tilstræbte man oprindeligt at gøre det så bredt som muligt med både tale- og skriftsprog, sprog fra private og offentlige sammenhænge, alment og specialiseret sprog osv. (Asmussen 2006). Siden har man imidlertid prioriteret kvantitet over repræsentativitet, og det betyder at *Bakspejlet* i dag rummer en meget stor andel avistekster.

I de følgende undersøgelser tager vi imidlertid udgangspunkt i et mere begrænset korpus end *Bakspejlet*, nemlig det heller ikke offentligt tilgængelige *Bak-2018*, som er *Bakspejlet* frem til og med 2018 med over en milliard løbende ord. Også *Bak-2018* er altså et statisk, men meget omfattende korpus. 98 % af indholdet i *Bak-2018* er skrevne tekster og især avistekster fra 1990-2018, resten af teksterne er fra 1930-1989 (2 %).

I de følgende afsnit beskriver vi for hver undersøgelse hvilke(n) delmængde(r) af *Bak-2018* de bygger på.

### 3 INDIVIDUELLE KVINDER OG MÆND I *BERLINGSKE* I TO PERIODER

For at få et hurtigt indtryk af udviklingen fra starten af 1990'erne til i dag undersøgte vi antal forekomster af ordformerne *han* og *hun*, dvs. pronomenerne kun i nominativ, i dagbladet *Berlingske Tidende* i 1990-1994 og *Berlingske* i 2018 (navneskift i 2011). For at udelukke *hun* og *han* brugt generisk udregnede vi den sandsynlige andel af generisk brugt *hun* og *han* ved at udvælge 300 vilkårlige eksempler med hvert af de to pronomener i hver periode, i alt 1200 eksempler. I disse eksempler blev andelen af generisk brugt *hun* samt *hun* brugt om andet end mennesker (fx dyr og skibe) udregnet som en andel af det samlede antal

*hun* i hver periode for sig, og tilsvarende med *han*. I opgørelsen af det samlede antal *hun* og *han* i de to delkorporusser af *Berlingske (Tidende)*, 1990-1994 og 2018, blev generisk brugte pronomener sat til at udgøre samme andel som i de 600 eksempler i perioden, og dette antal (se tabel 1) blev så trukket fra det samlede antal af hvert af de to pronomener i hver periode. Der var kun et eksempel på *han* brugt om en ørn i de 1200 eksempler og ingen eksempler på *hun* brugt om andet end mennesker; vi valgte helt at se bort fra denne lille andel af pronomener brugt om andet end mennesker.

Der var ingen eksempler på generisk brugt *hun* i de 300 eksempler fra den første periode, men i den anden periode var der et:

Chaufføren må hverken være fuld eller senil, og hun må ikke køre 24 timer i træk uden hverken søvn eller vådt og tørt.  
*Berlingske* 2018 (*Bak-2018*)

Eksemplet er fra en artikel om taxakørsel generelt. Der var to eksempler ud af 300 med generisk brugt *han* fra den første periode og et fra den anden periode. Resultaterne af optællingerne af ikke-generisk brugt *hun* og *han* fra *Berlingske (Tidende)* i de to perioder fremgår af tabel 1. Da antallet af generisk brugte pronomener er meget lavt, er estimerne meget usikre. For eksempel virker det tilfældigt at der er en forekomst af generisk brugt *hun* i de 300 undersøgte eksempler fra 2018. Men i betragtning af det meget store antal pronomener i det hele taget og den store forskel i antal *han* over for *hun* kan tallene i tabel 1 give et indtryk af hovedfordelingen.

Som det fremgår af tabellen, bruges ikke-generisk *han* ca. tre gange så hyppigt som *hun* i første halvdel af 1990'erne, og i 2018 var der mere end dobbelt så mange forekomster af ikke-generisk *han* som *hun*. *Hun* bruges altså mere frekvent i forhold til *han* i 2018 end i første halvdel af 1990'erne. Ændringen kan tænkes at afspejle at flere kvinder indtog magtfulde poster i samfundet i 2018 end i første halvdel af 1990'erne (se afsnit 1).



TABEL 1. FORDELINGEN AF IKKE-GENERISK BRUGT *HUN* OG *HAN* BERLINGSKE (TIDENDE) I ÅRENE 1990-1994 OG 2018.

	<i>hun</i>	<i>han</i>	andel ikke-generisk brugt <i>hun/han</i>
1990-1994	8.543 (0)	27.254 (183)	3,2
2018	8.723 (29)	21.540 (72)	2,5

Estimaterne over generisk brugte pronomener, som altså er fratrasket antal forekomster, er angivet i parentes.

#### 4 HVOR GAMLE ER *PIGER*?

I perioden 2000-2021 er der mange eksempler på at både *pige* og *dreng* bruges om voksne: DR har haft en serie om tre voksne søstre med titlen *Pigerne fra Berlin*; *Drengene fra Angora* er titlen på en anden navnkundig – og noget ældre – serie om tre voksne mænd; en fodboldsang fra 2008 hedder *Danmarks drenge*, og *Deli Drengene* er navnet på et firma der ”importerer og udvikler kvalitetsprodukter som fx trøffel, snackpølser, olivenolie...” (<https://delidrengene.dk>), næppe en sag for børn. Af en artikel på dr.dk fra februar 2020 fremgår det at unge kvinder forlader Ærø, mens unge mænd i højere grad bliver på øen:

De unge mænd kan nemlig godt få job og uddannelse inden for det, man kalder ’de traditionelle mandefag’, som diverse håndværksuddannelser og maritime uddannelser. Og det gør, at de i langt højere grad bliver hængende.

Det er Ærø-drengen Johan Kiær på 24 år et omvandrende eksempel på. Han er uddannet smed på det lokale Søby Værft.

- Mange bliver ligesom tilfredse med deres liv som håndværker, hvor de drikker en øl med drengene efter fyraften, går hjem og laver aftensmad og tager på arbejde igen dagen efter. Vi snakker om mænd i 20’erne, der bare tænker, ’okay, jeg skal ikke andet resten af livet’. (<https://www.dr.dk/nyheder/indland/eli-johan-og-jesper-er-singler-paa-aeroe-de-unge-kvinder-flygter-fra-oen>)

Her benævnes en 24-årig og en gruppe håndværkere altså både ved substantivet *dreng* og substantivet *mand*.

I denne delundersøgelse ser vi på hvilke aldersgrupper ordene *kvinde*,

*mand*, *pige* og *dreng* blev brugt om i *Berlingske Tidende* i perioden 1990-1994 og i *Berlingske* i 2018 i korpusset *Bakspejlet*. Grunden til forskellen i periodens udstrækning er at tekstmassen i korpusset i starten af 1990'erne er mindre end i 2018. Programmet CoREST udvalgte i hver af de to perioder tilfældigt 500 eksempler med hvert af leksemerne *kvinde*, *mand*, *pige* og *dreng*, dvs. uanset bøjningsform, i alt 4000 eksempler. Eksemplerne blev udtrukket med en kontekst på 20 ord på hver side.

Derefter blev eksemplerne gennemgået manuelt efter den følgende frasorteringsmanual. Hvert omtalt individ eller hver gruppe blev så vidt muligt kun repræsenteret en gang per forfatter. Eksempler med faste vendinger som *dreng møder pige* og *være mand for* blev frasorteret. Desuden frasorteredes eksempler hvor *pige* betød 'datter' eller 'tjenestepige' (fx *ung pige i huset*), eksempler hvor *dreng* betød 'søn', og eksempler hvor *mand* blev brugt i betydningen 'mandlig part i et ægteskab' (DDO) eller 'person; menneske – uden tanke på køn eller alder' (DDO) som i *mellem 15.000 og 20.000 mand*. Eksempler hvor målordet indgik i en titel som *Pige med perleøring*, blev også frasorteret. Hvis målordet blev brugt om en gruppe med flere aldersgrupper (fx *en bog for piger i alle aldre*), blev eksemplet også sorteret fra. Endelig blev et eksempel sorteret fra hvis målordet blev brugt generisk.

Den manuelle sortering – og altså frasorteringen – standsede når der var 100 eksempler for hvert målord i hver periode hvor det lod sig gøre at bestemme den aldersgruppe målordet blev brugt om. Det betød altså at der indgik 100 eksempler med hvert af de fire målord inden for hver af de to perioder, i alt 800 eksempler, i denne undersøgelse. De 100 eksempler for hvert målord i hver periode blev inddelt efter tre alderskategorier baseret på det typiske barndoms- og skoleforløb med 1. barndom og folkeskole, 2. ungdomsuddannelse og 3. videreuddannelse/voksenliv, dvs. børn på 0-15 år, unge på 16-19 år og voksne fra 20 år. For at klassificere eksemplerne var det somme tider nødvendigt at lave netsøgninger, fx for at finde ud af at "den unge pige" der omtales i det følgende eksempel, og som vandt ponyløbet i hubertusjagten i 2018, var 15 år:

De jagtansvarlige takkes, og den unge pige som vandt rævehalen i ponymøbet, bliver hyldet. *Berlingske* 2018 (*Bak*-2018)

Derudover blev de følgende kriterier brugt ved klassifikationen. Individer i grundskolen blev klassificeret som børn, individer på en ungdomsskole blev klassificeret som unge, individer knyttet til en arbejdsplads, en stillingsbetegnelse eller en branche blev klassificeret som voksne. Endelig blev individer der omtales som forældre, klassificeret som voksne med mindre det eksplicit fremgik at de var under 20.

Resultaterne af undersøgelsen, der fremgår af tabel 2, viser at sandsynligheden for at *pige* bruges om voksne kvinder er større i den tidlige periode end i 2018: 47 % af eksemplerne med *pige* i 1990-1994 blev brugt om voksne, mens det kun var 15 % af eksemplerne med *pige* i 2018. Det er tydeligt at *pige* bruges meget oftere om voksne end *dreng* om voksne, især i den tidlige periode: 47 % af eksemplerne med *pige* handler som nævnt om voksne i første halvdel af 1990'erne, mens kun 12 % af eksemplerne med *dreng* handler om voksne. Samtidig med at brugen af *pige* om voksne falder fra 1990-1994 til 2018 (47 % til 15 %), falder også andelen af *dreng* brugt om voksne (fra 12 % til 6 %).

For begge køn er der en stærk tendens til at bruge *pige* og *dreng* om unge fremfor *kvinde* og *mand* i begge perioder. Der sker et lille fald i brugen af *pige* om unge (fra 20 % til 16 %) og en næsten tilsvarende meget lille stigning i brugen af *kvinde* om denne aldersgruppe (0 til 3 eksempler).

TABEL 2. Fordelingen af eksempler med benævnelse af individer ved *pige*, *dreng*, *kvinde* eller *mand* i *Berlingske Tidende* (1990-1994) og *Berlingske* (2018) på de tre aldersgrupper *børn* (0-15 år), *unge* (16-19 år) og *voksne* (20 år eller derover).

	<i>pige</i> 1990- 1994	<i>pige</i> 2018	<i>dreng</i> 1990- 1994	<i>dreng</i> 2018	<i>kvinde</i> 1990- 1994	<i>kvinde</i> 2018	<i>mand</i> 1990- 1994	<i>mand</i> 2018
børn 0-15 år	33	69	72	74	0	0	0	0
unge 16-19 år	20	16	16	20	0	3	1	1
voksne fra 20 år	47	15	12	6	100	97	99	99
i alt	100	100	100	100	100	100	100	100

Omvendt sker der en lille stigning i brugen af *dreng* om unge fra første halvdel af 1990'erne til 2018 (16 % til 20 % af eksemplerne).

Tendenserne i udviklingen over tid svarer til det Sigley og Holmes (2002) fandt for *girl* og *boy* i et britisk korpus mellem 1961 og 1991, mens Baker (2014) fandt en stigning i brugen af både *girl* og *boy* brugt om voksne i 2006.

## 5 KOLLOKATIONER MED KVINDE, MAND, PIGE OG DRENG

I undersøgelsen af kollokationer med de fire substantiver indgår to delundersøgelser: en undersøgelse af kollokationer i hele *Bak-2018* (godt 1 mia. 70 mio. løbende ord), dvs. alle tekster i det ikke offentligt tilgængelige korpus *Bakspejlet* til og med 2018, og en undersøgelse af kollokationer kun i teksterne fra 2018 i det samme korpus. Dette sidste korpus, som altså er en delmængde af *Bak-2018*, kalder vi *Del-2018*. Hovedparten af teksterne i *Bak-2018* (98 %) er fra perioden 1990-2018, men der indgår også tekster fra perioden 1930-1989 (2 %). Som nævnt i afsnit 2 udgør avistekster den dominerende del af *Bak-2018*. Grunden til at vi her ikke kun inddrager en synkron undersøgelse af *Del-2018* for at få et indtryk af det nutidige sprog, er at *Del-2018* har begrænsninger for en kollokationsundersøgelse i kraft af sin størrelse, som det vil fremgå nedenfor.

Det værktøj vi brugte til undersøgelsen (Python-pakken Gensim) tillader kun undersøgelse af to ord i sekvens, dvs. bigrammer af ordformer. Vi lagde de forskellige former af adjektivleksemerne sammen, mens substantiverne kun indgår i grundformen. For nemheds skyld taler vi om *bigrammet mørkhåret kvinde* og om *mørkhåret* som *naboord* og *kvinde* som *målord* selv om bigrammerne rettelig omfatter en række af adjektivordformer og kun substantivet i singularis nominativ. Vi valgte at undersøge de adjektiver, inkl. participier, der står attributivt umiddelbart før substantiverne, altså bigrammer bestående af en eller anden form af et adjektiv plus *kvinde*, *mand*, *pige* eller *dreng*. Det betyder fx at kun *ranglet*, og ikke *klog*, blev inddraget i analysen af kollokationer med *pige* ud fra syntagmet *en klog, ranglet pige*, hvilket naturligvis er et minus, men en konsekvens af valget af værktøj.

For hvert bigram udregnedes en værdi for om de to ordformer som er repræsenteret i bigrammet, var særligt tiltrukket af hinanden i kor-

pusset. Det skete ved at udregne en såkaldt NPMI-værdi (*Normalized (Pointwise) Mutual Information*) (Bouma 2009) for bigrammet. NPMI-værdien viser antal forekomster af bigrammet sammenholdt med antallet af gange hver af de to ordformer overhovedet forekommer i korpusset. I formlen for hvert bigrams NPMI-værdi sættes altså frekvensen af bigrammet i forhold til frekvensen af de to ordformer i korpusset i det hele taget og i forhold til det totale antal ord i korpus. Sagt på en anden måde fortæller NPMI-værdien hvor stor sandsynligheden er for at ordform X forekommer før ordform Y, idet man tager hensyn til frekvensen af X og af Y i data hver for sig og inddrager antallet af løbende ord i korpusset.

NPMI-værdien for et bigram kan være mellem -1 og +1. Hvis et bigram har værdien -1, betyder det at de to ordformer aldrig forekommer ved siden af hinanden i korpusset, og hvis værdien er +1, betyder det at de aldrig forekommer alene, men kun sammen med hinanden. Fx har bigrammet *haricots verts* NPMI-værdien 0,9697 i *Bak-2018* idet de to ordformer næsten altid forekommer sammen (113 forekomster af *haricots*, 165 forekomster af *verts* og 107 forekomster af *haricots verts*). Omvendt har bigrammet *gravid dreng* NPMI-værdien -1. NPMI-værdien er altså et mål for hvor stærkt de to ordformer tiltrækker hinanden i korpusset.

I undersøgelsen fastsatte vi den nedre grænse for relevante bigrammer til en NPMI-værdi på 0,30. Desuden skulle der være mindst fem forekomster af et bigram for at det blev inddraget i undersøgelsen af *Bak-2018* (1930-2018), og mindst tre forekomster i *Del-2018*. På grund af *Del-2018*'s beskedne størrelse talte vi i nogle tilfælde forskellige ordformer af samme naboordslekse med i antallet af bigramsforekomster for at nå tre forekomster. I analysen udelod vi bigrammer fra bog- og filmtitler (fx *Rød mand*, *sort mand* og *En frygtelig kvinde*) fra begge korpusser.

Endelig omtaler vi her kun kort bigrammer hvor naboordet var nationalitets-, etnicitets- eller religionsbeskrivende (fx *syrisk*, *afroamerikansk*, *muslimsk*). I *Del-2018* er de relevante naboord til *kvinde*: *afroamerikansk*, *asiatisk*, *australsk*, *farvet*, *muslimsk*, *norsk*, *rumænsk*, *somalisk* og *sort*, og til *mand*: *afghansk*, *albansk*, *georgisk*, *hvid*, *rumænsk*, *syrisk* og *tunesisk*. I *Del-2018* er *muslimsk* kun naboord efter kriterierne

til substantiverne *kvinde* og *pige*, og *jødisk* kun til *pige*. Det resultat kan tolkes sådan at kvinder og piger i højere grad omtales ved deres religiøse tilhørsforhold end mænd og drenge når de beskrives ved et attributivt adjektiv. Lidt påfaldende er det også at *afroamerikansk*, *sort* og *farvet* kun er naboord til *kvinde*, og *hvid* kun til *mand* i *Del-2018*.

I den følgende gennemgang har vi delt kollokationerne ind i en række emnegrupper efter naboordene (se herunder). Til sidst ser vi på alle superlativer der bruges sammen med *kvinde*, *mand*, *pige* og *dreng* uanset emne. I tabellerne over naboord til *kvinde*, *mand*, *pige* og *dreng* der har en NPMI-score på 0,30 eller derover, og som har mindst fem (*Bak-2018*) eller tre (*Del-2018*) forekomster, står de naboord der er specielle for det enkelte substantiv, først under substantivet, og de naboord der er fælles for enten *kvinde* og *mand* eller *pige* og *dreng*, følger derunder efter en tom linje (se fx tabel 4). Naboordene forekommer i grundform i tabellerne uanset hvilken bøjningsform de forekommer i i korpus. Naboordene er alfabetiseret inden for hver gruppe.

Først ser vi på hvor mange unikke naboord og hvor mange fælles naboord der forekommer med hvert af de fire målord i hvert korpus (se tabel 3). Fælles naboord er adjektiver eller participier som efter kriterierne er naboord til både *kvinde* og *mand* eller til både *pige* og *dreng*. Af tabel 3 fremgår det at der i begge korpusser er flere forskellige naboord ved *mand* end ved *kvinde*, men at forskellen kun er markant i *Del-2018*. Omvendt er der flere unikke naboord ved *pige* end ved *dreng* i *Bak-2018*. Det sidste forklarer delvist forskellen i antal ved *kvinde* og *mand* idet der er flere tilfælde af at *pige* bruges om voksne, end af at *dreng* bruges om voksne (se afsnit 4).

I gennemgangen af naboordene inden for de forskellige kategorier trækker vi på betydningsdefinitionerne i *DDO* fordi vi ikke har haft mulighed for at lave en kvalitativ analyse af alle målord og naboord i kontekst. Det mener vi kan forsvares fordi ordbogen bygger på det samme korpus som vi har brugt.

### 5.1 Alder

*Ung* har den højeste NPMI-score af alle målord sammen med *kvinde*, *mand* og *pige* i *Bak-2018*, men bigrammet *ung dreng* falder under grænsen på 0,30 (se tabel 4). Selv når man frasorterer alle eksempler med

TABEL 3. ANTAL UNIKKE OG FÆLLES NABOORD FOR HVERT MÅLORD I DE TO KORPUSSE.

	<i>Bak-2018</i>		<i>Del-2018</i>	
	unikke	fælles	unikke	fælles
<i>kvinde</i>	74	14	24	9
<i>mand</i>	86		40	
<i>pige</i>	37	17	9	4
<i>dreng</i>	27		7	

*ung pige i huset*, har *ung pige* stadig en NPMI-score over 0,50. Grunden til forskellen på *pige* og *dreng* er sandsynligvis at samme aldersgruppe for de to køn ofte omtales som henholdsvis *ung pige* og *ung mand* og ikke som *ung kvinde* og *ung dreng* (se også afsnit 4), men *ung kvinde* når dog også en NPMI-score på 0,50. Skævheden i alder for hvornår et individ benævnes ved *pige*, og hvornår ved *dreng*, kan også aflæses af det forhold at der i *Del-2018* er knap 2,25 gange så mange bigrammer bestående af [adjektiv] *mand* (1410) som bigrammer bestående af [adjektiv] *kvinde* (631), mens der er godt 2,5 gange så mange bigrammer med [adjektiv] *pige* (495) som [adjektiv] *dreng* (189). Som det fremgår af de absolutte tal, skyldes det ikke at personer af hunkøn omtales oftere i *Del-2018* end personer af hankøn: Der er 1,4 gange så mange eksempler med [adjektiv] *mand/dreng* (1599) som [adjektiv] *kvindel/pige* (1126) i *Del-2018*.

I omtalen af drenges alder lægges der især i hele perioden vægt på at de er på vej til at være voksne (*halvvoksen*, *halvstor*), mens pigerne i højere grad ses som unge (*præpubertær*, *purung*, *ung*, men dog også *halvstor*). Adjektivet *halvgammel* betyder ifølge DDO 'som endnu ikke er gammel, men har levet eller eksisteret i forholdsvis lang tid; temmelig gammel', dvs. at det i kombination med *dreng* må antages at være blevet brugt om voksne mænd som i det følgende eksempel, hvor *han* refererer til Dennis Knudsen på 54 år:

Selvom han er ved at være en halvgammel dreng, træner han p. t. på Lanzarote for at komme i form til sin femte ironman.  
*Ekstra Bladet* 2018 (*Bak-2018*)

Naboordene til *mand* lægger vægt på at mænd er fuldvoksne, mens kvinder med en metafor fra biologien betegnes som *modne*. Det tyder på at biologisk udvikling ses som vigtigere for kvinder, mens status som voksen i form af myndighed og selvstændighed er vigtigere for mænd.

*Gammel* danner kollokation med *mand*, hvorimod NPMI-scoren for *gammel kvinde* kun er 0,20. I *Bak-2018* er *ældre kvinde* syv gange hyppigere end *gammel kvinde*. Ifølge *DDO* kan *ældre* bruges i betydningen

TABEL 4. NABOORD DER VEDRØRER ALDER.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>moden</i>	<i>fuldvoksen</i>	<i>præpubertær</i>	<i>halvgammel</i>
	<i>gammel</i>	<i>purung</i>	<i>halvvoksen</i>
		<i>ung</i>	
<i>midaldrende</i>		<i>halvstor</i>	
<i>n-årig</i>		<i>jævnaIdrende</i>	
<i>ung</i>		<i>lille</i>	
<i>voksen</i>		<i>mindreårige</i>	
<i>ynge</i>		<i>n-årig</i>	
<i>ældre</i>		<i>nyfødt</i>	
Del-2018			
<i>moden</i>	<i>gammel</i>	<i>ung</i>	<i>jævnaIdrende</i>
<i>midaldrende</i>		<i>lille</i>	
<i>n-årig</i>		<i>mindreårig</i>	
<i>ung</i>		<i>nyfødt</i>	
<i>voksen</i>		<i>n-årig</i>	
<i>ynge</i>			
<i>ældre</i>			

Betegnelsen *n-årig* sammenfatter alle de brugte aldersbetegnelser, dvs. *5-årig*, *6-årig*, *80-årig* osv.



'forholdsvis gammel', dvs. ikke helt så gammel som hvis man brugte *gammel*. Muligvis fordi kvinder, i hvert fald når de benævnes ved *kvinde*, er stærkere forbundet med attraktivitet og biologi end mænd (se senere), og fordi yngre alder derfor regnes som væsentligere for kvinder end for mænd, bruges udtrykket *gammel kvinde* i mindre grad end udtrykket *gammel mand*. I *Del-2018* (se tabel 4) adskiller kvinder og mænd sig med hensyn til aldersbetegnelser: Personer benævnt ved *kvinde* beskrives ved *moden*, og personer benævnt ved *mand* ved *gammel*.

Sammenfattende tyder kollokationerne med aldersbetegnende adjektiver på at der i hele perioden lægges vægt på ungdom hos individer der betegnes ved *kvinde* eller *pige* idet de især beskrives ved adjektiverne *moden*, *præpubertær*, *purung* og *ung*. Ved individer der betegnes ved *mand* eller *dreng*, lægges der mere vægt på den selvstændighed og myndighed der følger med at være voksen, jf. adjektiverne *fuldvoksen* og *halvvoksen* om henholdsvis mænd og drenge. Desuden peger fordelingen af *moden* (med *kvinde*) og *gammel* og *halvgammel* (med henholdsvis *mand* og *dreng*) på at alder omtales eufemistisk i forbindelse med kvinder. Der er dog også mange fælles naboord mellem *kvinde* og *mand* og mellem *pige* og *dreng*.

## 5.2 Udseende

Dette og det følgende afsnit handler om udseende og påklædning. Vi har skilt påklædning ud i et særligt afsnit af praktiske grunde.

I kategorien *Udseende* (se tabel 5) er det første der slår en, at der er flere naboord til *mand* end til *kvinde*. Der er altså større variation i karakteriseringen af individer der omtales ved *mand*, end af individer der omtales ved *kvinde*.

Listerne over naboord til henholdsvis *kvinde* og *mand* der har med udseende at gøre, viser desuden meget lidt overlap (tabel 5). Dog er der nogenlunde det samme fokus på håret hos de to grupper. Blandt naboordene til *pige* er der større fokus på håret end blandt naboordene til *dreng*. Fysisk størrelse uden muskelmasse kommer til udtryk i de følgende naboord til *kvinde*: *frodig*, *buttet*, *barmfager* og *storbarmet*, og *spinkel* fremhæver den manglende styrke. Ved *mand* fokuserer de fleste ord der har med størrelse at gøre, på fysisk styrke: *bredskuldret*, *granvoksen* (DDO: 'høj og stor og som regel stærk og muskuløs'), *muskuløs*,

*senet, sportstrøjet* og måske *tætbygget*, men dog også *fedladen, firskåret, trind* og *lavstammet*.

Ingen af de ord hvor der lægges vægt på omfanget af fedtvævet, går igen som naboord for *kvinde* og *mand*. Kvinders høje vægt angives ved ordene *buttet* og *frodig*, mænds ved *fedladen* og *trind*. Ifølge DDO betyder *buttet* 'med fyldig og rund form – især om menneske eller legemsdel', og *frodig* 'med svulmende former; fyldig, stor – især om (dele af) kvindekrop'. Det første eksempel under *buttet* i ordbogen er: "*Du er ikke fed,*" *sagde jeg sødt. "Du er højst lidt buttet.*" fra Bjarne Reuters *Mig og Albinoni*. Eksemplet peger på at *buttet* ligesom *frodig* er et eufemistisk udtryk om stor kropsvægt. *Frodig* peger desuden ligesom *moden* på kvinders biologi, DDO første betydning under *frodig* er 'frugtbar, næringsrig', dog om planter.

Der er ikke nogen eufemistisk omskrivning af stor kropsvægt hos individer der betegnes ved *mand*, sammenlignet med *buttet* og *frodig* om individer der betegnes ved *kvinde*. Tilsvarende beskrives individer der benævnes ved *pige*, med adjektiverne *buttet* og *barmfager* (DDO: 'med stor og flot barm'), mens individer der benævnes ved *dreng*, og som menes at have meget fedtvæv beskrives ved *tyk* i korpusset. Dog bør det nævnes at DDO beskriver et af naboordene til *mand*, nemlig *trind*, neutralt eller positivt som 'rund og fyldig; velnæret om person, dyr eller legemsdel' og med *buttet* som synonym.

En hel række adjektiver med positiv værdi indgår i kollokation med *pige*, nemlig *dejlig, køn, laber, pæn, smuk, sød* og *yndig*. Af disse er kun *sød* genganger ved *dreng*. Kollokationen af ordformerne *sød* og *pige* er imidlertid meget stærkere end kollokationen af ordformerne *sød* og *dreng*. *Sød dreng* har en NPMI-score på 0,30 og er nr. 68 ud af 70 kollokationer målt efter styrke i *Bak-2018*, mens *sød pige* har en NPMI-score på 0,43 og er nr. 10 ud af 87 kollokationer.

*Del-2018* adskiller sig fra det samlede korpus især ved at der ikke er så meget omtale af kvinders runde former (jf. *frodig, buttet, barmfager* og *storbarmet* som naboord i *Bak-2018*, men ikke i *Del-2018*), mere af deres hårfarve (*blond, lyshåret, rødhåret, sorthåret*), mens mænd og drenge også i *Del-2018* især karakteriseres ved deres kropsfylde. Især lyse hårfarver anses for relevante når en person af hunkøn omtales ved *pige* (*blond, lyshåret*). Et af de adjektiver der har positiv værdi, og som

TABEL 5. NABOORD DER VEDRORER UDSEENDE.

	<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018				
	<i>barmfager</i>	<i>bredskuldreret</i>	<i>barmfager</i>	<i>blåøjjet</i>
	<i>blond</i>	<i>enarmet</i>	<i>blond</i>	<i>rangellet</i>
	<i>buttet</i>	<i>etbenet</i>	<i>brunøjjet</i>	<i>svagelig</i>
	<i>frodig</i>	<i>fedladen</i>	<i>buttet</i>	<i>tyk</i>
	<i>korthåret</i>	<i>firskåren</i>	<i>dejlig</i>	
	<i>lyshåret</i>	<i>granvoksen</i>	<i>krølhåret</i>	
	<i>mandhaftig</i>	<i>gråskegget</i>	<i>køn</i>	
	<i>modebevidst</i>	<i>halvskuldreret</i>	<i>laber</i>	
	<i>rygvendt</i>	<i>hvidskegget</i>	<i>lyslokket</i>	
	<i>rødhåret</i>	<i>krøghåndet</i>	<i>pæn</i>	
	<i>smuk</i>	<i>kronraget</i>	<i>smuk</i>	
	<i>sorthåret</i>	<i>langskegget</i>	<i>sorthåret</i>	
	<i>storbarmet</i>	<i>muskulos</i>	<i>yndig</i>	
	<i>spinkel</i>	<i>mørklødet</i>		
	<i>underskøn</i>	<i>rødmosset</i>		<i>halvstor</i>
	<i>velklædt</i>	<i>rødskegget</i>		<i>lille</i>
		<i>senet</i>		<i>lyshåret</i>
		<i>skuldreret</i>		<i>mørkhåret</i>
		<i>skegget</i>		<i>rødhåret</i>
		<i>sportstrænet</i>		<i>spinkel</i>
		<i>statelig</i>		<i>sød</i>
		<i>trind</i>		<i>velskabt</i>
		<i>tyndhåret</i>		
		<i>tætbygget</i>		
		<i>gråhåret</i>		
		<i>hvidhåret</i>		
		<i>mørkhåret</i>		
Del-2018				
	<i>blond</i>	<i>bredskuldreret</i>	<i>blond</i>	<i>dejlig</i>
	<i>lyshåret</i>	<i>gråhåret</i>	<i>lyshåret</i>	<i>krølhåret</i>
	<i>mørkhåret</i>	<i>korpulent</i>	<i>pæn</i>	<i>tyk</i>
	<i>rødhåret</i>	<i>muskulos</i>	<i>smuk</i>	
	<i>rank</i>	<i>skegget</i>	<i>sød</i>	
	<i>smuk</i>	<i>tætbygget</i>		
	<i>spinkel</i>			

er naboord til *pige* i hele *Bak-2018*, nemlig *dejlig*, forekommer nu kun som naboord med tilstrækkelig frekvens til *dreng*.

Sammenfattende karakteriseres håret oftere i omtalen af individer der benævnes ved *pige*, end i omtalen af individer der omtales ved *dreng*, mens der ikke er nogen forskel med hensyn til hår for individer der benævnes ved *kvinde* eller *mand*. Der lægges vægt på fysisk styrke ved individer der omtales ved *mand*. For individer der omtales ved *pige* eller *kvinde*, karakteriseres stor kropsfylde ved eufemistiske ord, mens stor kropsfylde hos individer der omtales ved *dreng* eller *mand*, karakteriseres ved negative eller neutrale ord.

### 5.3 Påklædning

Med hensyn til påklædning scorer *kvinde* højest på antal naboord (se tabel 6). De specifikke naboord til *kvinde* fokuserer på tøjets karakter (type, farve eller materiale) og på muslimske kvinders tildækning med naboordene *burkaklædt*, *niqabklædt* og muligvis også *tilsløret*, *tørklædeklædt* og *tilhyllet*, men bl.a. *tilhyllet* bruges også fx i omtalen af tildækkede kvinder i kunst. En række naboord til *kvinde* peger på i hvilken grad tøjet dækker kroppen, nemlig *afklædt*, *halvnøgen*, *let-påklædt*, *topløs* og *nøgen*. Kun *nøgen* er også naboord til *mand* i *Bak-2018*, og der er forskel på styrken af de to kollokationer: I *Bak-2018* er NPMI-scoren for *nøgen mand* 0,31 med kollokationen på en 120. plads målt efter styrke, mens NPMI-scoren for *nøgen kvinde* er 0,44 med kollokationen på sjette plads blandt *kvinde*'s kollokationer. Det ser altså ud til at skribenterne i *Bak-2018* oftere finder anledning til at omtale graden af påklædthed ved individer der omtales ved *kvinde*, end ved individer der omtales ved *mand*. I *Del-2018* er kun *nøgen kvinde* tilstrækkeligt frekvent efter kriterierne. De to adjektiver der handler om grad af tildækning ved *mand* i *Bak-2018*, nemlig *maskeret* og *hætteklædt*, hører typisk hjemme i kontekster hvor mænd ses som (potentielle) gerningsmænd.

TABEL 6. NABOORD DER VEDRØRER PÅKLÆDNING.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>afklædt</i>	<i>bebrillet</i>	<i>hvidklædt</i>	<i>bebrillet</i>
<i>burkaklædt</i>	<i>grønklædt</i>	<i>tørklædeklædt</i>	
<i>halvnøgen</i>	<i>habitklædt</i>		
<i>hvidklædt</i>	<i>hætteklædt</i>		
<i>letpåkklædt</i>	<i>maskeret</i>		
<i>niqabklædt</i>	<i>mørkkklædt</i>		
<i>pelsklædt</i>			
<i>tilhyllet</i>			
<i>tilsløret</i>			
<i>topløs</i>			
<i>tørklædeklædt</i>			
	<i>nøgen</i>		
	<i>sortklædt</i>		
Del-2018			
<i>burkaklædt</i>	<i>klædt</i> <sup>2</sup>		
<i>nøgen</i>	<i>mørkkklædt</i>		
<i>tørklædeklædt</i>	<i>maskeret</i>		

#### 5.4 Køn, seksualitet og familierelationer

*Kvinde* har absolut flest naboord i kategorien *Køn, seksualitet og familierelationer* i *Bak-2018*, men ikke i *Del-2018* (se tabel 7). Flere af naboordene i *Bak-2018* kan af biologiske grunde næsten kun bruges i bigrammer med *kvinde*: *aborterende*, *ammende*, *fødende*, *gravid* og *højgravid*. Hvis ordene bruges i bigrammer med *mand*, har de metaforisk betydning, og denne brug er ikke hyppig nok sammenlignet med kollokationer med *kvinde* til at de er naboord til *mand* efter denne undersøgelses kriterier. Der er ikke biologiske grunde til at *lesbisk* kun er naboord til *kvinde*; en googlesøgning (22/10 2021) finder enkelte eksempler på *lesbisk mand*, men *lesbisk* har stærke selektionsrestriktioner med *kvinde*, og der er ingen eksempler på *lesbisk mand* i *Bak-2018*.

2 Klædt som naboord til *mand* stammer fra udtrykkene *bedst klædt*, *værst klædt*.

Nogle adjektiver, som handler om familierelationer, og som hverken af biologiske grunde eller på grund af selektionsrestriktioner må forventes at være begrænset til *kvinde*, er ikke desto mindre kun naboord til *kvinde* og ikke til *mand* i *Bak-2018*. Det drejer sig om *barnløs*, *enlig* og *ugift*. Det eneste familierelaterede naboord til *mand* er *frasepareret*. I alle forekomster af *frasepareret mand* bruges bigrammet i *Bak-2018* imidlertid til at karakterisere en kvindes tidligere ægtefælle (nominaler som *sin/hendes/kvindens fraseparerede mand*). At dømme ud fra kollokationerne karakteriseres personer der omtales som *mand*, sjældent ved deres familierelationer. I en undersøgelse af britisk engelsk i *British National Corpus* fandt Pearce (2008) tilsvarende at *married woman* var tre gange så hyppigt som *married man*, og at *childless woman* var femten gange hyppigere end *childless man*. I *Del-2018* er *fraskilt* dog naboord til både *kvinde* og *mand* efter kriterierne.

I *Bak-2018* tiltrækker *kvinde* en række adjektiver der beskriver kvinders seksualitet (*sensuel*) og deres normfølgende eller tilbageholdende (*dydig*, *ærbart*) eller grænseoverskridende (*falden*, *sexgal*) seksualitet (jf. beskrivelsen af ordenes betydninger i *DDO*). *Transkønnet* og *transseksuel* kan bruges som attributiv til både *kvinde* og *mand* (se fx *sexogsamfund.dk*), men bruges i *Bak-2018* kun sammen med *kvinde* i tilstrækkeligt omfang til at være naboord til *kvinde* efter kriterierne. I *Del-2018* er kun *sexet* naboord til *kvinde* og samtidig også til *mand*.

I *Bak-2018* peger to adjektiver i retning af mænds seksualitet, nemlig *metroseksuel* og *überseksuel*. Ifølge *DDO* kendes *metroseksuel* fra 2003 og har betydningen 'som er ung, heteroseksuel og mand og går op i udseende, påklædning, boligindretning og den moderne storbys *fashionable* livsstil'. Adjektivet handler altså ikke entydigt om mænds seksualitet. *Überseksuel* kendes fra 2005 og har ifølge *DDO* to betydninger: 1. 'som både udfylder den klassiske manderolle med vægt på maskulinitet, handlekraft og ansvarsbevidsthed og giver plads til træk der traditionelt har været betragtet som feminine, fx omsorg, forfængelighed og medfølelse' og 2. 'som i overdreven grad udstiller eller spiller på sex'. Bigrammet *überseksuel mand* bruges imidlertid kun med den første betydning i *Bak-2018*. *Überseksuel kvinde* er ikke et bigram i *Bak-2018* efter kriterierne, men tre ud af fire eksempler på den anden betydning af ordet i *DDO* handler om kvinder, det sidste eksempel om nogle

TABEL 7. NABOORD DER VEDRØRER KØN, SEKSUALITET OG FAMILIERELATIONER.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>aborterende</i>	<i>frasepareret</i>		
<i>abortsøgende</i>	<i>kønsskifteopereret</i>		
<i>ammende</i>	<i>metroseksuel</i>		
<i>barnløs</i>	<i>überseksuel</i>		
<i>dydig</i>			
<i>emanciperet</i>			
<i>enlig</i>			
<i>falden</i>			
<i>fraskilt</i>			
<i>frigjort</i>			
<i>fødende</i>			
<i>gravid</i>			
<i>højgravid</i>			
<i>lesbisk</i>			
<i>sensuel</i>			
<i>sexet</i>			
<i>sexgal</i>			
<i>transkønnet</i>			
<i>transseksuel</i>			
<i>ugift</i>			
<i>ærbar</i>			
	<i>heteroseksuel</i>		<i>bortadopteret</i>
			<i>faderløs</i>
			<i>forældreløs</i>
Del-2018			
<i>gravid</i>	<i>fremmed</i>	<i>dydig</i>	
	<i>heteroseksuel</i>		
	<i>homoseksuel</i>		
	<i>liderlig</i>		
	<i>uimodståelig</i>		
	<i>fraskilt</i>		
	<i>sexet</i>		

kalkmalerier. Hverken *metroseksuel* eller *überseksuel* handler altså om mænds seksualitet.

I *Del-2018* er billedet helt anderledes: Her er der fem naboord der efter kriterierne er specielle for *mand*, og som har noget med seksualitet at gøre, nemlig *heteroseksuel*, *homoseksuel*, *liderlig*, *sexet* og *uimodståelig*. Forskellen mellem de to korpusser er altså her virkelig slående. I 2018 er mandens seksualitet tilsyneladende blevet et tema på en helt anden måde end den var tidligere, hvor det primært var kvinder der blev karakteriseret ved deres seksualitet i hele perioden.

### 5.5 Personlighed og mental kapacitet

I kategorien *Personlighed og mental kapacitet* er der slet ikke overlap blandt naboordene til *kvinde* og til *mand*, hverken i *Bak-2018* eller i *Del-2018* (se tabel 8). Der er flere naboord om intelligensniveau og intellektuel aktivitet ved *mand* end *kvinde* og ved *dreng* end *pige*. *Mand* har naboordene *belæst*, *forstandig*, *klog*, *kundskabsrig*, *lærd* og *vis* i *Bak-2018*, *kvinde* kun naboordet *veluddannet* om intellektuel aktivitet. *Livsklog* (som naboord til *kvinde*) betegner stor erfaring snarere end høj intelligens (DDO: 'som har (eller vidner om) stor viden om tilværelsen og livets glæder og sorger'). *Dreng* har naboordene *fremmelig*, *gammelklog*, *opvakt*, *velbegavet* og *videbegærlig*, og *pige 12-tals* (med mange sta-vevarianter)<sup>3</sup> og *tænksom*.

Bortset muligvis fra *gammelklog* (DDO: 'med større erfaring og viden end hvad der svarer til personens alder især om barn') har naboordene til *dreng* en positiv værdi (*fremmelig*, *opvakt*, *velbegavet*, *videbegærlig*), mens eksempler fra avistekster viser at *12-tals* ikke er en positiv karakteristik af *piger*:

Hun kalder sig selv en 'tidligere 12-talspige', da hun i gymnasietiden pressede sig selv meget for at sikre 12-taller i sit karakterblad. *Jyllands-Posten* 2017 (*Bak-2018*)

[...] fx i form af pigen, som gerne vil have gode karakterer, men som rubriceres som en ulykkelig 12-talspige i en robotificeret upersonlig fremtid. *Berlingske* 2018 (*Bak-2018*)



TABEL 8. NABOORD DER VEDRØRER PERSONLIGHED OG MENTAL KAPACITET.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>gæv</i>	<i>belest</i>	<i>12-tals</i>	<i>adfærdsvanskelig</i>
<i>livsklog</i>	<i>besindig</i>	<i>gæv</i>	<i>dagdrømmende</i>
<i>synsk</i>	<i>ensom</i>	<i>livsglad</i>	<i>fremmelig</i>
<i>veluddannet</i>	<i>flink</i>	<i>tænksom</i>	<i>fræk</i>
<i>viljestærk</i>	<i>forstandig</i>	<i>ubeslutsom</i>	<i>gammelklog</i>
	<i>glad</i>		<i>indestuttet</i>
	<i>gudfrygtig</i>		<i>leende</i>
	<i>klog</i>		<i>opvakt</i>
	<i>kundskabsrig</i>		<i>slem</i>
	<i>lærd</i>		<i>velbegavet</i>
	<i>modig</i>		<i>videbegerlig</i>
	<i>morsom</i>		
	<i>ond</i>		<i>artig</i>
	<i>retskaffen</i>		<i>genert</i>
	<i>uberygtet</i>		<i>velopdragen</i>
	<i>uheldig</i>		<i>uartig</i>
	<i>uligevægtig</i>		
	<i>usnild</i>		
	<i>vis</i>		
Del-2018			
<i>mystisk</i>	<i>flink</i>	<i>følsom</i>	<i>begavet</i>
<i>napkæftet</i>	<i>forbitret</i>		<i>fræk</i>
<i>sej</i>	<i>førfængelig</i>		<i>slem</i>
<i>stum</i>	<i>klog</i>		
	<i>kræsen</i>		

3 *12-tals* er ikke et adjektiv og burde efter dansk retskrivning skrives i et ord med det substantiv det modificerer, hvilket imidlertid langt fra altid er tilfældet i avistekster; derfor dukkede *12-tals* op i undersøgelsen af bigrammer. Vi har tilladt os at tage det med her fordi det siger noget væsentligt om synet på intelligens hos piger og unge kvinder (Hansen & Blom 2017).

Hansen og Blom (2017) har undersøgt mediernes brug af begrebet 12-talspige, og de fandt at 12-talspiger fremstilles som syge piger og kvinder, der bukkes under for en samfundskultur med stærkt præstationspres. I *DDO* beskrives betydningen af *12-talspige* sådan: 'pige der er flittig og dygtig i skolen og ofte får topkarakteren 12 – evt. opfattet som alt for perfektionistisk og pligtopfyldende'; *12-talsdreng* forekommer ikke i *DDO*.

Det er altså især individer der omtales ved *mand* og *dreng*, der tillægges intelligens og intellektuel aktivitet, mens intellektuel aktivitet hos individer der omtales ved *pige*, karakteriseres som noget potentielt negativt for individet selv. Dog er *tænk* også naboord til *pige*. Individer der omtales ved *kvinde*, karakteriseres ved uddannelsesniveau (*veluddannet*) eller stor erfaring (*livsklog*). I *Del-2018* peger ingen af naboordene til *kvinde* og *pige* på intelligens, mens *klog* er naboord til *mand* og *begavet* til *dreng*.

Adjektiverne *adfærdsvanskelig*, *slem* og *fræk* i *Bak-2018* og *fræk* og *slem* i *Del-2018* er specifikke naboord til *dreng*, mens *dreng* og *pige* deler *artig*, *uartig* og *velopdragen* i *Bak-2018*. Dog har bigrammet *uartig dreng* en NPMI-score på 0,45 og står øverst på listen over naboord til *dreng*, mens bigrammet *uartig pige* har en NPMI-score på 0,30 og altså kun lige kvalificerer til at tælle med i undersøgelsen. *Pige* har stærkere bånd til adjektiver der karakteriserer positive egenskaber som *gæv* og *livsglad* i *Bak-2018*. I *Del-2018* er det kun *følsom* der er naboord til *pige*.

Bortset fra *livsglad* som naboord til *pige* i *Bak-2018* er adjektiver der har med munterhed og glæde at gøre, forbeholdt individer der omtales ved *mand* og *dreng*. *Dreng* og *mand* tiltrækker *leende*, *glad* og *morsom*, mens *kvinde* ikke har noget naboord der handler om glæde eller latter. I *Bak-2018* er adjektiverne *ensom*, *uheldig*, *uligevægtig* og *usnild* dog også naboord til *mand*. En række naboord til *mand* handler om generel moral, nemlig adjektiverne *flink*, *gudfrygtig*, *modig*, *uberygtet*, *retskaffen* og *ond*. Naboord til *kvinde* er *gæv* og *viljestærk* i *Bak-2018* og *sej* og *rapkæftet* i *Del-2018*. I *Del-2018* er *forbitret* naboord til *mand*, og der er ikke nogen adjektiver om munterhed.

Under *Personlighed og mental kapacitet* ser vi altså at der er langt flere ord for intelligens og intellektuel aktivitet blandt naboordene til *mand* og *dreng* end blandt naboordene til *kvinde* og *pige*. Ved *kvinde* ses kun

*veluddannet* med fokus på intelligensens resultat, og ved *pige* ses intellektuel aktivitet som noget potentielt negativt for individet selv (*12-tals*). Drenges adfærd karakteriseres i højere grad som grænseoverskridende (*adfærdsvanskelig, fræk, slem*) end pigers. Personer der omtales ved *mand* og *dreng*, karakteriseres i højere grad ved adjektiver der peger på munterhed (*glad* og *morsom* som naboord til *mand* og *leende* som naboord til *dreng*) end personer der omtales ved *kvinde* eller *pige*, idet *livsglad* dog er naboord til *pige*. I *Del-2018* er der ingen naboord der peger i retning af munterhed.

### 5.6 Magt, rigdom og erhverv

Som det fremgår af tabel 9, er der ingen naboord til *kvinde* der har med magt og rigdom at gøre i *Bak-2018*. Naboordene *selverhvervende* og *udearbejdende* til *kvinde* karakteriserer kvinder som økonomisk uafhængige, noget som især er relevant i en tid hvor kvinder ellers ofte er økonomisk afhængige. I *Del-2018* er *selverhvervende* og *udearbejdende* ikke længere naboord efter kriterierne til *kvinde*, hvilket kan tænkes at afspejle det forhold at det tages for givet at kvinder har et arbejde.

*Mand* har naboord der betegner rigdom og i mindre udstrækning magt og uafhængighed i *Bak-2018* (*næstmægtigst, selfmade* med betydningen 'som har opnået rigdom og succes udelukkende ved egen indsats' – DDO). Dog forekommer *mand* også sammen med *håndgangen*, der fokuserer på manden som afhængig af en anden – magtfuld – person (DDO: 'lydig over for en magtfuld person og villig til at medvirke ved gennemførelsen af dennes planer'). Samlet set forekommer ordene *rig, næstrigest, tredjerigest, fjerderigest, femterigest, mægtig* og *næstmægtigst* attributivt til *mand* 2.294 gange i *Bak-2018*, 9,5 gange hyppigere end som attributiv til *kvinde* (240 gange). På dette punkt er der ikke sket nogen ændring til 2018; i *Del-2018* er ord der karakteriserer magt og rigdom også kun naboord til *mand* efter kriterierne.

### 5.7 Kriminalitet

I kategorien *Kriminalitet* er der ikke nogen fælles naboord mellem *kvinde* og *mand* eller mellem *pige* og *dreng* (se tabel 10). Naboordene til *mand* handler om den omtalte mand eller de omtalte mænd som potentielle voldsudøvere med adjektiver som *knivbevæbnet, pistolbevæbnet*,

TABEL 9. NABOORD SOM VEDRØRER MAGT, RIGDOM OG ERHVERV.

	<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018				
	<i>selverhvervende</i>	<i>femterigeste</i>	<i>fattig</i>	
	<i>prostitueret</i>	<i>fjerderigeste</i>		
	<i>udearbejdende</i>	<i>holden</i>		
		<i>håndgangen</i>		
		<i>næstmægtigst</i>		
		<i>næstrigest</i>		
		<i>rigest</i>		
		<i>selfmade</i>		
		<i>tredjerigest</i>		
		<i>våbenfør</i>		
Del-2018				
	<i>placeret<sup>4</sup></i>	<i>femterigest</i>		
	<i>undertrykt</i>	<i>holden</i>		
		<i>mægtfuld</i>		
		<i>mægtigst</i>		
		<i>næstrigest</i>		
		<i>rigest</i>		

*rockerrelateret* og *voldsudøvende* eller som underkastet samfundets rets-udøvelse: *drabsdømt*, *drabssigtet*, *drabstiltalt*, *eftersøgt*, *pædofilidømt* og *terrordømt* i *Bak-2018*. Naboordene til *kvinde* og *pige* anskuer det eller de omtalte individer som ofre med naboordene *kidnappet*, *overfalden*, *voldsramt* og *voldtagen kvinde* og *indemuret*, *overfalden*, *sexmisbrugt* og *voldtagen pige*. I *Del-2018* genfindes dette mønster med *forulempet* og *kidnappet* som naboord til *kvinde*, og *anholdt*, *eftersøgt*, *farlig*, *fængslet*, *medsigtet* og *voldelig* som naboord til *mand*.

Der er ingen naboord til *dreng* i denne kategori, muligvis fordi drenge der er involveret i kriminalitet som udøvere eller ofre, omtales ved andet end *dreng*.

<sup>4</sup> *Placeret* som naboord til *kvinde* stammer fra udtryk som *højt/højst/højt placeret*, *lavest placeret*, *centralt placeret*.

TABEL 10. NABOORD DER VEDRØRER KRIMINALITET.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>kidnappet</i>	<i>drabsdømt</i>	<i>indemuret</i>	
<i>overfalden</i>	<i>drabssigtet</i>	<i>overfalden</i>	
<i>voldsramt</i>	<i>drabstiltalt</i>	<i>sexmisbrugt</i>	
<i>voldtagen</i>	<i>eftersøgt</i>	<i>voldtagen</i>	
	<i>knivbevæbnet</i>		
	<i>pistolbevæbnet</i>		
	<i>pedofilidømt</i>		
	<i>rockerrelateret</i>		
	<i>terrordømt</i>		
	<i>voldsudøvende</i>		
Del-2018			
<i>forulempet</i>	<i>anholdt</i>		
<i>kidnappet</i>	<i>eftersøgt</i>		
<i>tilfældig</i>	<i>farlig</i>		
	<i>fængslet</i>		
	<i>medsigtet</i>		
	<i>ukendt</i>		
	<i>voldelig</i>		

Vi har også anført *tilfældig* som naboord til *kvinde* og *ukendt* som naboord til *mand* i *Del-2018* i denne kategori: *Tilfældig kvinde* bruges ofte til at karakterisere kvinder som ofre, og *ukendt mand* bruges ofte om gerningsmænd som i de følgende eksempler fra *Bak-2018*.

Kort tid senere, i november 2016, blev en tilfældig kvinde - for øjnene af sit barn - ramt i brystet af skud ved Magasin i Lyngby. *Jyllands-Posten* 2018 (*Bak-2018*)

Der er tale om et kynisk og planlagt seksuelt overgreb og drab af særdeles brutal karakter på en tilfældig kvinde. *Jyllands-Posten* 2018 (*Bak-2018*)

En ukendt mand gik ind i kvindens bolig og tog nogle penge, der lå i en kasse i stuen. *JP Aarhus* 2018 (*Bak-2018*)  
[...] en ukendt mand indhentede hende og befølte hende forskellige steder på kroppen. *JP Aarhus* 2018 (*Bak-2018*)

*Ukendt* og *tilfældig* bruges imidlertid ikke kun i forbindelse med kriminalitet.

### 5.8 Aktivitet og passivitet

I kategorien *Aktivitet og passivitet* har vi opstillet de naboord som fremstiller et individ som handlende eller passiv, uden dog at gentage ord der allerede indgår i andre kategorier som fx *leende* eller alle *participierne* under *kriminalitet*. Som det fremgår af tabel 11, er der ikke overlap mellem naboordene til *kvinde* og *mand* med hensyn til aktivitet og passivitet, hverken i *Bak-2018* eller i *Del-2018*, og kun *stakkels* går igen som naboord til både *pige*, *dreng* og *mand*.

Det må antages at både piger og drenge bliver forbrændt og mobbet, men *forbrændt* og *mobbet* er kun naboord til *dreng*, muligvis fordi *pige* i højere grad end *dreng* bruges om voksne, og der tales måske mere om at blive forbrændt eller mobbet når det drejer sig om børn end om voksne.

Kvinder karakteriseres især ved deres specifikke typer af aktiviteter i *Bak-2018*: *cigaretrygende*, *skakspillende* og *solbadende*. Det samme gælder drenge (*fløjtespillende*, *tissende*), mens mænd karakteriseres ved generelle egenskaber der peger på aktivitet, *driftig*, *foretagsom*, *travl* og muligvis *hurtig*, i *Del-2018* ved *travl* og *hurtigst*. Kvinder karakteriseres dog også generelt ved *handlekraftig* i *Bak-2018*. Naboordene til *pige* handler alle om passivitet, nemlig *forskræmt*, *forsvunden*, *skrækslagen* og *stakkels* i *Bak-2018* og *forulempet* i *Del-2018*.

Sammenfattende karakteriseres især mænd ved ord der betegner generelle egenskaber af aktivitet, mens kvinder og drenge mere karakteriseres ved specifikke aktivitetstyper, og piger kun ved passivitet.

### 5.9 Sygdom

I kategorien *Sygdom* har vi sat adjektiver der handler om sygdom eller om kroppen som biologisk, men ikke seksuel, enhed (se tabel 12).

TABEL 11. NABOORD DER VEDRØRER AKTIVITET OG PASSIVITET.

	<i>aktivitet</i>	<i>passivitet</i>	<i>aktivitet</i>	<i>passivitet</i>
	<i>kvinde</i>		<i>mand</i>	
Bak-2018				
	<i>cigaretrygende</i>	<i>nødstedt</i>	<i>driftig</i>	<i>slagen</i>
	<i>handlekraftig</i>	<i>tilredt</i>	<i>foretagsom</i>	<i>stakkels</i>
	<i>skakspillende</i>		<i>hurtig</i>	
	<i>solbadende</i>		<i>travl</i>	
	<i>pige</i>		<i>dreng</i>	
		<i>forskræmt</i>	<i>fløjtespillende</i>	<i>forbrændt</i>
		<i>forsvunden</i>	<i>tissende</i>	<i>mobbet</i>
		<i>skrækslagen</i>		
		<i>stakkels</i>		<i>stakkels</i>
Del-2018				
	<i>kvinde</i>		<i>mand</i>	
	<i>handlekraftig</i>	<i>forulempet</i>	<i>hurtigst</i>	
			<i>travl</i>	
	<i>pige</i>		<i>dreng</i>	
	<i>krænket</i>			

*Kvinde* tiltrækker i højere grad adjektiver om sygdomme end *mand* i *Bak-2018*: *kræftamt*, *kræftsyg*, *skleroseramt*. *Pige* er også specifikt forbundet med *kræftamt* sammenlignet med *dreng*, mens *kræftsyg* indgår i kollokation med både *pige* og *dreng*. Piger og kvinder beskrives altså oftere som ofre for kræft end drenge og mænd relativt til hvor mange gange de fire grupper betegnes med substantiverne *kvinde*, *mand*, *pige* og *dreng*. Mænds kropslige problemer beskrives i højere grad som selvforskyldt: *alkoholpåvirket*, *beruset*, *medicinpåvirket*. Ifølge Sundhedsstyrelsen drikker kvinder og mænd lige meget, men mænd drikker oftere store mængder alkohol på en gang. Dermed kommer de også i højere grad i konflikt med loven og bliver avisstof; ni ud af ti spritbilister er mænd (Rådet for Sikker Trafik 2016).

I *Del-2018* er der kun et specifikt naboord til *kvinde* i denne kategori, nemlig *dement*, mens *mand* tiltrækker *beruset*, *handicappet*, *kræftsyg* og

TABEL 12. NABOORD DER VEDRØRER SYGDOM.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
<i>bevidstløs</i>	<i>afdød</i>	<i>anorektisk</i>	<i>handicappet</i>
<i>gangbesværet</i>	<i>alkoholpåvirket</i>	<i>bevidstløs</i>	<i>hjerneskadet</i>
<i>kræftamt</i>	<i>beruset</i>	<i>døvstum</i>	<i>livløs</i>
<i>kræftsyg</i>	<i>evnesvag</i>	<i>evnesvag</i>	<i>multihandicappet</i>
<i>skleroseramt</i>	<i>medicinpåvirket</i>	<i>kræftamt</i>	<i>svagelig</i>
<i>retarderet</i>		<i>udviklingshæmmet</i>	
<i>udviklingshæmmet</i>			
	<i>dement</i>		<i>autistisk</i>
	<i>sindsforvirret</i>		<i>kræftsyg</i>
			<i>retarderet</i>
Del-2018			
<i>dement</i>	<i>afdød</i>		
	<i>beruset</i>		
	<i>handicappet</i>		
	<i>kræftsyg</i>		
	<i>kvæstet</i>		
	<i>sindsforvirret</i>		

*kvæstet*. Alkoholpåvirkning er altså også i *Del-2018* især knyttet sammen med mænd, mens *kræftsyg* her er naboord til *mand*.

### 5.10 Superlativer

Under kategorien *Superlativer* har vi samlet alle morfologiske superlativer der opfylder kriterierne for en NPMI-score på mindst 0,30 og en forekomst med substantivet på mindst fem i *Bak-2018* og mindst tre i *Del-2018*, uanset om de har optrådt i andre kategorier (se tabel 13).

Det interessante ved superlativer er at de i reglen udpeger et enkelt individ som scorer i yderområdet på en eller anden egenskab. Der er ikke noget superlativ der indgår i kollokation med *kvinde* efter kriterierne i *Bak-2018*. I *Del-2018* er der en delvist skjult superlativ idet *placeret* fra *bedst/højest/højest placeret* (og *højt/centralt placeret*) er naboord



TABEL 13. NABOORD DER ER SUPERLATIVER.

<i>kvinde</i>	<i>mand</i>	<i>pige</i>	<i>dreng</i>
Bak-2018			
	<i>ensomst</i>	<i>dejligst</i>	<i>slemmest</i>
	<i>femterigest</i>	<i>kønnest</i>	<i>uartigst</i>
	<i>fjerderigest</i>	<i>sødest</i>	
	<i>flinkest</i>		
	<i>gladest</i>		
	<i>hurtigst</i>		
	<i>morsomst</i>		
	<i>mægtigst</i>		
	<i>næstmægtigst</i>		
	<i>næstrigest</i>		
	<i>ondest</i>		
	<i>rigest</i>		
	<i>tredjerigest</i>		
	<i>visest</i>		
Del-2018			
<i>placeret</i>	<i>farligst</i>	<i>smukkest</i>	
	<i>femterigest</i>		
	<i>hurtigst</i>		
	<i>mægtigst</i>		
	<i>næstrigest</i>		
	<i>rigest</i>		

til *kvinde* efter kriterierne. Her spores altså en stigende interesse for kvinders placering i hierarkier.

De parametre mænd bedømmes på i *Bak-2018*, er rigdom og magt, klogskab, munterhed og hurtighed, men også ondskab og ensomhed. Naboordene i *Del-2018* viser at mænd også i 2018 sammenlignes på rigdom og magt.

I den yngre aldersgruppe vurderes både piger og drenge på skalaer i *Bak-2018*, men helt forskellige skalaer. Piger rangordnes på deres udseende og sympatiske egenskaber (*dejligst*, *kønnest*, *sødest* – og i *Del-2018* *smukkest*), drenge på deres normbrydende adfærd i *Bak-2018* (*slemmest*, *uartigst*), i *Del-2018* er der ingen superlativer blandt naboordene til *dreng*.

## 6 DISKUSSION

I undersøgelsen af antallet af ordformerne *han* og *hun* i *Berlingske* (*Tidende*) i perioderne 1990-1994 og 2018 frasorterede vi en andel af det samlede antal pronomener som vi skønnede ville være generisk brugt på grundlag af en stikprøve. Da stikprøven var ret lille og antallet af generisk brugte pronomener tilsvarende lavt, er skønnet forbundet med usikkerhed, og på grundlag af stikprøverne så vi helt bort fra muligheden af pronomener brugt om andet end mennesker. Det meget store antal pronomener der indgår i undersøgelsen (66.060), taget i betragtning mener vi at resultaterne – 3,2 gange så mange forekomster af *han* som *hun* i 1990-1994 og 2,5 gange så mange *han* som *hun* i 2018 – kan tolkes som en klar overvægt i omtalen af individuelle hankønsindivider sammenlignet med individuelle hunkønsindivider i begge perioder, men med mindre forskel mellem de to grupper i anden periode.

Dette resultat stemmer i store træk overens med hvad Scheuer (2017) fandt ved en undersøgelse af pronomenernes forekomster i et korpus fra 1995 og *Korpus 2000*, nemlig ca. dobbelt så mange forekomster af *han* som *hun* i begge korpuser. Tidsforskellen i Scheuers undersøgelse er tilsyneladende så lille at der ikke kunne ses noget fald i det relative forhold. Baker (2014) fandt et fald i andelen af hankønspronomen i forhold til hunkønspronomen i britisk engelske korpuser fra 1961 og 2006. Desuden fandt han at der i 2006 blev brugt næsten 1,75 gange så mange hankønspronomen som hunkønspronomen. Den lavere andel af hankønspronomen i det britiske korpus fra 2006 sammenlignet med det danske korpus fra 2018 kan skyldes at Baker inkluderede både akkusativformerne *him* og *her* og de possessive pronomen *his* og *hers*. Man kan især forestille sig at *her* er mere frekvent i forhold til *him* end *she* i forhold til *he*, fordi det kan tænkes at der oftere refereres til individer af hunkøn i objektposition end i subjektposition, og omvendt med hensyn til individer af hankøn. Det kan altså tænkes at antallet af omtaler af kvinder og piger ville øges i en dansk undersøgelse hvis man inddrog pronomenerne i ikke-subjektposition (se også Pearce 2008).

Den anden undersøgelse handlede om hvilke aldersgrupper *pige* og *dreng* bruges om i de to perioder. DDO angiver fire hovedbetydninger af *pige* og syv for *dreng*. Tre af de fire hovedbetydninger af *pige* går

i store træk igen blandt hovedbetydningerne af *dreng*, nemlig 'barn eller ung person' af det relevante køn, 'person af [det relevante køn] som man er far eller mor til' og 'tjenestepige' eller 'barn eller ung person af hankøn der hjælper til eller arbejder et bestemt sted, fx på en gård'. Den sidste hovedbetydning af *pige* har nogle overlap med en af de andre betydninger af *dreng*. For *pige* er det '(yngre) kvinde', og der står at denne betydning undertiden er 'forskønnende'. For *dreng* er det 'person af hankøn, uden hensyn til vedkommendes alder' med anmærkningen 'ofte spøgende eller forskønnende'. Begge hovedbetydninger har forskellige underbetydninger, bl.a. '(yngre) kvinde betragtet som (mulig) seksualpartner' og '(yngre) mand betragtet som (mulig) seksualpartner'. Ordbogen mener altså overvejende at der er en parallel mellem de to ords betydning.

Resultaterne af vores optælling af hvilke aldersgrupper de to ord bruges om, peger imidlertid på en forskel i brugen og dermed betydningen. Begge ord bruges i begge perioder i *Berlingske (Tidende)* om voksne, dvs. individer fra 20 år og opefter. Men hvor 47 af 100 eksempler på *pige* handler om en voksen kvinde, handler kun 15 af 100 eksempler på *dreng* om en voksen mand i første periode. Der sker et fald i brugen af ordene om voksne fra første halvdel af 1990'erne til 2018, og faldet er størst for *pige*. Endnu i 2018 blev *pige* brugt om voksne en del hyppigere end *dreng* om voksne (15 % af forekomsterne af *pige* mod 6 % af forekomsterne af *dreng* bedømt ud fra stikprøver).

Ordbogen peger på at brugen af leksemerne *pige* og *dreng* om voksne "undertiden" er "forskønnende". I afsnit 5.1 pegede vi på skævheden i brugen af *gammel* som naboord efter kriterierne til *kvinde* sammenlignet med *mand*. Bigrammet *gammel kvinde* opfylder ikke kriterierne for at indgå i undersøgelsen, mens *gammel mand* gør, og *ældre kvinde* bruges syv gange oftere end *gammel kvinde* i *Bak-2018*. DDO mener at *gammel* bruges om mennesker på mere end ca. 75 år, og *ældre* om mennesker på mere end ca. 60 år. Både brugen af *ældre* sammenlignet med *gammel* og brugen af *pige* sammenlignet med *dreng* om voksne peger altså på at sprogbrugen omtaler kvinder som om de var yngre end mænd. Det kan tolkes som forskønnende givet den store vægt der lægges på ungdom i samfundet.

Alternativt kan man sige at sprogbrugen viser leksemernes betyd-

ning, og at *pige* og *dreng* simpelthen ikke har parallel betydning. Der er god tradition i forskningen for at analysere leksemernes betydning gennem sprogbrugen. Fx viser Jensen (2001) hvordan bl.a. *ret* gennemgår en udvikling fra et adjektiv med en konkret og objektiv betydning (fx *sijn rette mening*) til et adverbium med en subjektiv betydning (fx *en ret god vin* og *at hun ret skulle føle*). Men Fillmore (1982: 126) mener at "[o]ne important type of change consists in reconstituting the motivating circumstances while preserving the lexical item and its basic fit with the associated scene." Dvs. nogle leksemer er så bevidst en del af sprogbrugernes sprogviden at de kan ændre deres sprogbrug i overensstemmelse med ændringer i deres opfattelse af "the motivating circumstances". Fillmores eksempel er netop nogle sprogbrugeres ændrede brug af *girl* og *woman* sådan at aldersgrænsen mellem hvem leksemerne bruges om, afpasses efter aldersgrænsen for *boy* og *man*. Som argument for sin semantiske teori *frame semantics* fremhæver Fillmore at det ville være mangelfuldt blot at sige at der her er sket en semantisk forandring i leksemernes betydning som viser sig i sprogbrugen: "the full explanation must assign the change to the underlying schematization on the part of the language user" (Fillmore 1982: 126); ændringen skyldes sprogbrugernes opmærksomhed på det felt som ordene bruges om. At brugen af *pige* om voksne i dansk er noget i hvert fald nogle sprogbrugere fokuserer på, fremgår af den følgende titel på en aviscommentar: "Må man sige hvid, trans, brun, broget eller pige om en voksen kvinde?" (Hendricks & Marker 2019). Det kan imidlertid tænkes at aldersgrænsen for brug af *dreng* øges sådan at forskellen mellem de to køn udlignes den modsatte vej (se også Baker 2014). Eksemplerne i afsnit 4 (*Delidrengene*, *Drengene fra Angora* osv.) kunne imidlertid tolkes sådan at der er en forskel på *pige* brugt om voksne og *dreng* brugt om voksne (jf. DDO's bemærkninger om sprogbrugen: for *pige* "forskønnende", for *dreng* "ofte spøgende eller forskønnede"). Men det kan kun en kvalitativ analyse vise.

Inden vi diskuterer resultaterne af den tredje undersøgelse, kollokationsundersøgelsen, vil vi repetere hvad man skal forstå ved *naboord* i denne artikel. Et *naboord* til et af substantiverne *kvinde*, *mand*, *pige* eller *dreng* i grundform er et adjektiv (eventuelt participium) der står attributivt til et af substantiverne. Derudover skal der være en særlig

tiltrækning mellem adjektivet og substantivet målt ved bigrammets NPMI-score. NPMI-scoren beregnes ved en formel der viser hyppigheden af sekvensen [specifik ordform + *kvinde/mand/pige/dreng*] sat i relation til de to ordformers samlede frekvens i korpuset hver for sig og det totale antal ord i korpus. Vi satte så to yderligere kriterier: Vi var kun interesseret i bigrammer – kollokationer – med en NPMI-score på mindst 0,30 og en frekvens i *Bak-2018* på mindst fem og i det mindre korpus *Del-2018* på mindst tre.

En konsekvens af udregningen af NPMI-scorerne er at adjektiver som næsten kun forekommer med et af målordene, vil have en høj NPMI-score med dette målord uanset hvor få gange adjektivet forekommer i korpus. Fx må *gudfrygtig*, *retskaffen*, *uberygtet* og *usnild* antages at være infrekvente i absolut forstand, specielt når teksterne fra før 1990 kun udgør 2 % af tekstmassen i *Bak-2018*. *Usnild* findes ikke i *DDO*, men i den ældre *ODS* findes det med betydningen 'uklog; uforstandig; ubehændig'. Når det alligevel optræder som naboord til *mand*, skyldes det at adjektivet stort set ikke forekommer i andre funktioner end som attributiv til *mand*.

Et andet forhold ved beregningen af naboordsfunktionen som man skal være opmærksom på, er at specifikke begivenheder i verden eller Danmark kan påvirke frekvensen af et bigram kraftigt, især fordi *Bak-2018* fortrinsvis inkluderer avistekster. I år 2010 sprængte en mand med en benprotese en bombe i et københavnsk hotel og blev mistænkt for at være terrorist, hvilket kan have givet anledning til at *etbenet mand* er et bigram efter kriterierne. Jo kortere periode et korpus af avistekster dækker, des mere vil kollokationerne og dermed kategorien naboord afspejle den slags tilfældigheder. Det kunne tænkes især at ramme delkorpuset *Del-2018*, som jo kun omfatter et enkelt år. Men ingen af naboordene i *Del-2018* ser ud til at afspejle enkeltbegivenheder givet at vi sætter NPMI-scoren til 0,30 og har frasorteret bog- og filmtitler. Et andet aspekt af korpusstørrelsen er at diversiteten i kollokationerne er mindre i et lille korpus simpelthen fordi der tales om færre fænomener med tilstrækkelig frekvens. Dette problem ser ud til at have ramt *Del-2018*, som rummer meget færre naboord end det store korpus som *Del-2018* er en delmængde af. Det ideelle til at undersøge udviklingen over tid ville derfor være at lave undersøgelsen i to korpusser bestående

af fx 10 års tekster med 10 år imellem, altså fx et korpus 1990-1999 og et korpus 2010-2019. Den mulighed havde vi imidlertid ikke, og det betyder at det diakrone aspekt af undersøgelsen af naboord skal tolkes med forsigtighed.

Et tredje forhold ved beregningen af naboordsfunktionen er at når fx *kræftamt* er naboord til *kvinde* i *Bak-2018*, men ikke til *mand*, behøver det ikke at betyde at kræftamte mænd slet ikke eller næsten ikke omtales i *Bak-2018*; faktisk kan sekvensen *kræftamt mand* være mere frekvent end *kræftamt kvinde* i korpusset. Men hvis *mand* forekommer i mange andre kontekster med eller uden attributiv, bliver tiltrækningen mellem adjektivet *kræftamt* og substantivet *mand*, eller altså NP-MI-scoren for *kræftamt mand*, for lav til at *kræftamt* er naboord til *mand*. Modsat *usnild* forekommer *kræftamt* også i andre kontekster end som attributiv til *mand*, og så bliver ordets tiltrækning specifikt til *mand* for svag. At *kræftamt* er naboord til *kvinde* og ikke til *mand*, betyder altså at der er større sandsynlighed for at når *kvinde* bruges, bruges det sammen med *kræftamt* end sandsynlighed for at *mand* bruges sammen med *kræftamt*.

Man kunne så sige at disse forbehold gør undersøgelsen af bigrammer irrelevant, men pointen er at bigrammer der opfylder de kriterier vi har opstillet, kan tænkes at få en særlig status på grund af det vi har kaldt den særlige tiltrækningskraft mellem ordene. Nyere kognitiv og sprogbrugsbaseret lingvistik (fx Croft & Cruse 2004) peger på at frekvente kombinationer i sprogborgen kan blive forankret (*entrenched*) psykolingvistisk. Denne forankring viser sig psykolingvistisk fx ved at frekvente kombinationer læses hurtigere, huskes bedre i forsøg med sætningsgenkaldelse og produceres hurtigere end mindre frekvente kombinationer (for en oversigt over denne type forskning, se Tremblay 2012). Dette tolkes som at kombinationerne lagres som enheder i hukommelsen; de bliver forankret. Vi har intet belæg for at sige at bigrammer med en vis frekvens og NP-MI-score har samme status i forhold til forankring som de grammatiske og højfrekvente leksikalske fænomener der er undersøgt psykolingvistisk, men pointen fra den psykolingvistiske forskning er at menneskers sproglige hukommelse er sensitiv over for sprogbrugsfænomener, og NP-MI-scorer måler et sprogbrugsfænomen.

En yderligere pointe ved NPMI-scorer er at de afspejler hvad sprogbrugerne især taler om. I forhold til grammatik og fordeling af informationer på leksemer i konstruktioner har Slobin (fx 1996) fremsat en teori om *thinking for speaking*. Han mener at hvert sprogs grammatik og leksikon strukturerer indholdet på en særlig måde (jf. Dansk Funktionel Lingvistik's begreb om indholdsstruktur, fx Harder 2006). Denne måde kalder Slobin det enkelte sprogs særlige retoriske stil, dvs. den måde sprogbrugerne typisk strukturerer indholdet i grammatiske og leksikalske konstruktioner i diskurs. Når sprogbrugerne skal sige noget, består deres *thinking for speaking* i at udvælge "those characteristics of objects and events that (a) fit some conceptualization of the event, and (b) are readily encodable in the language" (Slobin 1996: 76). Det er en vigtig pointe at Slobin ikke mener at sprogets grammatik eller leksikon determinerer den måde sprogbrugerne udtrykker sig på. Den retoriske stil er typiske måder at udtrykke sig på i et givent sprog, men altså ikke de eneste, jf. ændringen i aldersgrænse for brug af *pige* om voksne.

Den kognitive sprogbrugsbaserede lingvistik og psykolingvistikken påpejning af forankring af frekvente kombinationer og Slobins teori om sprogbrugernes sprogspecifikke *thinking for speaking* lader sig forene i en påstand om at bigrammer med høj frekvens får en særlig psykolingvistisk og samfundsmæssig status gennem den vedvarende gentagelse i offentlige tekster som avistekster. Det betyder imidlertid ikke at de alle kan skæres over en kam. En del af de bigrammer vi har fundet, skyldes forskelligt syn på samme egenskab hos kvinder og mænd, fx alder og intelligens. Andre skyldes forskelle mellem kvinder og mænd som ikke er sproglige. Kvinder bliver gravide og ammer, mænd har skæg og er statistisk set mere kriminelle end kvinder, og deres alkoholforbrug er mere synligt i samfundet end kvinders. Men selv for adjektiver der af biologiske grunde kun bruges om kvinder som fx *ammende* eller *barmfager*, kan man spørge om relevansen af at kommentere netop disse aspekter ved kvinder i en given situation. Der skal en egentlig kvalitativ undersøgelse til at finde ud af om det i den enkelte tekst er relevant for tekstens indhold at en kvinde er barmfager eller en mand belæst, eller om der er en tendens til at opsøge og omtale begivenheder hvor disse træk er relevante (se fx en historie med titlen *TV: Her bliver gigantisk, barmfager kvinde påkørt bagfra* – Ekstra Bladet 2013). Det kan

en korpusundersøgelse ikke vise; den kan kun vise at *barmfager kvinde* og *belæst mand* er bigrammer efter kriterierne. I det følgende eksempel fra *Bak-2018* viser en gennemlæsning af hele teksten at kvindens barm ikke er relevant for historien. Alligevel omtales hun som *den barmfagre kvinde* gennem hele historien.

For en del år siden havde jeg en retssag, hvor en midaldrende barmfager kvinde blev dømt til at betale ca. 20.000 kr. til en førende restaurant på Fyn. Hendes datter var blevet gift, og hun havde inviteret bryllupsselskabet på en fin middag uden at kunne betale. *Politiken* 2000 (*Bak-2018*)

Selv i en vis forstand biologisk funderede bigrammer kan altså blive forankret psykolingvistisk så at de gentages hvis ikke sprogbrugerne træffer et aktivt valg.

I modsætning til Baker (2014) fandt vi at der i hele perioden ikke bruges markant flere attributive adjektiver ved *mand* end ved *kvinde*. Derimod fandt vi klare forskelle på hvilke emneområder der havde flere unikke naboord til *kvinde* end til *mand*, og omvendt. Inden for kategorierne *Påklædning* og *Køn, seksualitet og familierelationer* var der flere naboord til *kvinde* end til *mand*. Omvendt var der flere naboord til *mand* inden for *Udseende* og især *Personlighed og mental kapacitet*, *Magt, rigdom og erhverv*, *Kriminalitet* og *Superlativer*. Man kunne forestille sig at de relativt få naboord til *kvinde* inden for disse kategorier, skyldtes at *pige* i vid udstrækning bliver brugt om voksne. Det kan måske forklare at der er flere naboord til *mand* end *kvinde* i kategorien *Udseende*; i kategorien er der lige så mange unikke naboord til *kvinde* og *pige* tilsammen som til *mand*, nemlig 24. Men der er ikke særlig mange naboord til *pige* inden for de andre kategorier; faktisk er der klart flere naboord til *dreng* end til *pige* inden for *Personlighed og mental kapacitet*. Individuer der benævnes ved *kvinde*, beskrives altså især med adjektiver inden for områderne påklædning, seksualitet og familierelationer, mens individer der benævnes ved *mand*, især beskrives ved adjektiver inden for kategorierne *Personlighed og mental kapacitet*, *Magt, rigdom og erhverv*, *Kriminalitet* og *Superlativer*.

I en del tilfælde omtales det samme træk ved de to køn systematisk



forskelligt. Kvinder med mere kropsfedt end flertallet omtales ved adjektiverne *butikket*, *frodig* og *barmfager* (DDO: 'med en stor og flot barm' – vores fremhævnings), mens mænd omtales ved *fedladen* og *trind*. *Buttet* kan siges at være forskønnende, mens *frodig* associerer til kvinders biologi, og *barmfager* fremhæver et attribut ved kvinder som kønsvæsner. Til forskel herfra fokuserer *fedladen* og *trind* kun på fedtet og den runde form hos mænd, og disse adjektiver står i modsætning til naboordene *bredskuldret*, *muskuløs*, *senet* og *sportstrænet* til *mand*. Det eneste naboord til *kvinde* om kropsfylde der ikke handler om rigeligt fedtvæv, er *spinkel*, som beskriver mangel på muskler ud over det manglende fedt. Sammenfattende kan man sige at kvinders store kropsfylde omtales forskønnende eller ved et ord der fremhæver deres attraktivitet, mænds ved negative eller neutrale ord, mens lidt fedtvæv omtales neutralt hos kvinder og overvejende positivt hos mænd.

Alder har også forskelligt sprogligt udtryk hos de to grupper. Individer der benævnes ved *kvinde*, omtales som *modne* (biologi) og ældre (knap så gamle som *gammel*), mens individer der benævnes ved *mand*, betegnes som *gamle* og – med en understregning af at målet er nået – som *fuldvoksne* (DDO: 'som har fulden sin vækst og har nået voksenstadiet'). Ved *pige* fremhæves det endnu umodne (*præpubertær*, *purung* og *ung*), mens individer der benævnes ved *dreng*, ses som på vej til det fuldvoksne stadie (*halvvoksen*). Fordelingen af naboord kan tolkes sådan at biologi ses som vigtigt ved bedømmelsen af kvinders og pigers alder, mens (manglende) myndighed og selvstændighed er vigtigt i omtalen af individer der benævnes ved *mand* eller *dreng*.

Med hensyn til graden af påklædthed er der ingen biologisk bestemt forskel på kvinder og mænd, ligesom leksemerne *afklædt*, *halvnøgen*, *letpåkledt*, *topløs* og *nøgen* ikke har selektionsrestriktioner med *kvinde*. Alligevel er de alle med undtagelse af *nøgen* kun naboord til *kvinde*, og NPMI-scoren for *nøgen kvinde* er langt højere (0,44) end for *nøgen mand* (0,31). Den høje frekvens af *nøgen* blandt de karakteriserende adjektiver til *kvinde* kan tænkes at smitte af på det Slobin kalder den retoriske stil – det bliver i så fald selvforstærkende at tale om *nøgen kvinde*. I *Del-2018* er *nøgen* kun naboord til *kvinde*. Det samme misforhold mellem de to køn ses ved *sød* i forhold til *pige* og *dreng*. *Sød* er naboord til begge substantiver, men NPMI-scoren for *sød pige* (0,43)

er højere end for *sød dreng* (0,30). Omvendt har *uartig dreng* (0,42) en højere NPMI-score end *uartig pige* (0,30). Resultaterne peger i retning af faste forestillinger om kvinder, mænd, piger og drenge. Forestillingen om drenge understreges af at *adfærdsvanskelig*, *fræk* og *slem* kun er naboord til *dreng* i hele perioden og *fræk* og *slem* også i 2018 alene.

Naboordene til *kvinde* og *pige* beskriver i højere grad kvinder og piger som ofre for kriminalitet og for sygdom end mænd og drenge, men denne skævhed modsvarer næppe af de faktiske forhold. Her er det imidlertid igen relevant at huske på metoden til beregning af NPMI-scorer. Eftersom målordet *mand* forekommer i så mange andre kontekster, har adjektiver som beskriver mænd som ofre for kriminalitet eller sygdom, sværere ved at nå en tilstrækkelig NPMI-score. Skævheden bliver først udlignet når substantiverne *kvinde* og *pige* bruges oftere og i andre kontekster.

Individer der benævnes ved *mand*, karakteriseres i højere grad end individer der benævnes ved *kvinde*, ved munterhed, intelligens, rigdom, magt og moralsk stillingtagen og ved egenskaber der karakteriserer aktivitet, men også ved ensomhed (*ensom*) og ynkkelighed (*stakkels*). Høj præstation i skole- og uddannelsessystemet af individer der benævnes ved *pige*, ses som noget potentielt skadeligt (*12-tals*). Scheuer (2017) fandt at der både i 1995 og 2000 indgår langt flere substantiver efter *hans* end efter *hendes* inden for det område han kalder ”magt og ånd” (fx *værker*, *opfattelse*, *parti* i 1995 og *mening*, *parti*, *karriere* i 2000). I de britiske korpusser i Bakers (2014) og Pearces (2008) undersøgelser blev mænd også i højere grad end kvinder beskrevet som magtfulde, fysisk stærke og intelligente.

Det største skift vi kan registrere fra hele perioden til kun 2018, er brugen af adjektiver der karakteriserer individer der benævnes ved *mand*, som seksuelle væsner. I hele perioden er de eneste naboord til *mand* der handler om seksualitet, *kønsskifteopereret* og *heteroseksuel*, mens naboordene til *kvinde* er *lesbisk*, *heteroseksuel*, *sensuel*, *sexet*, *sexgal*, *transkønnet* og *transseksuel*; hertil kommer den kontrollerede seksualitet med *dydig* og *ærbær* som naboord til *kvinde*. Det billede er helt ændret i 2018, hvor naboordene til *mand* i denne kategori er *heteroseksuel*, *homoseksuel*, *liderlig*, *uimodståelig* og *sexet*, mens *sexet* er det eneste naboord til *kvinde* med tilstrækkelig høj frekvens der handler om køn

eller seksualitet. Dette skifte viderefører den udvikling Scheuer (2017) fandt, nemlig at der kom flere kropsord efter *hans* i 2000 sammenlignet med 1995, fx *ansigt, hånd, krop* og *hoved*.

Et andet tegn på et skift fra hele perioden til 2018 er at *fraskilt* er naboord til både *kvinde* og *mand* i 2018, hvor det i hele perioden kun var naboord til *kvinde*. Ændringerne i naboord er muligvis et udtryk for et nyt syn på mænds status i relation til familie og seksualitet (jf. diskussioner i nyere tid om barsel til mænd). Hvor *glad* og *morsom* var naboord til *mand* i hele perioden, er *forbitret* og ikke *glad* og *morsom* naboord i 2018. Mænd har tilsyneladende bevaret deres magt og rigdom i 2018; i hvert fald karakteriserer en række naboord til *mand* rigdom og magt i 2018. Naboordet *placeret* til *kvinde* antyder dog at kvinders relative position er blevet mere væsentlig i 2018. Men det andet naboord til *kvinde* i denne kategori i 2018 er *undertrykt*.

En vigtig pointe ved den metode vi har brugt i kollokationsundersøgelsen, er at øget omtale af individer der benævnes ved *kvinde* eller *pige*, er en nødvendig, men ikke tilstrækkelig forudsætning for at forankret sprogbrug brydes op. En yderligere forudsætning er at skribenterne begynder at bruge attributiver inden for alle emneområder til karakteristik af mennesker af begge køn.

Fremtidige undersøgelser kunne udbygge de undersøgelser vi har fremlagt her, fx med analyse af perifrastiske superlativer, dvs. trigrammer af *mest/mindst* + [adjektiv] + *kvinde/mand/pig/dreng*, en undersøgelse af sociolingvistisk variation i brugen af *pige* om voksne og kvalitative undersøgelser af relevansen af de attributiver der bruges med *kvinde*, *mand*, *pige* og *dreng* i forhold til teksternes emne. Men en fordel ved at gentage den kollokationsundersøgelse vi har lavet her, med nye avis-korpusser om fx ti år ville være en bedre mulighed for sammenligning.

## TAK

Artiklen bygger på Ida Flörkes kandidatspeciale i lingvistik ved Københavns Universitet med Elisabeth Engberg-Pedersen som vejleder. Vi vil gerne takke Det Danske Sprog- og Litteraturselskab for at give Ida adgang til det ellers ikke offentligt tilgængelige korpus *Bak-2018*. Vi vil desuden gerne takke Nicolai Hartvig Sørensen for hjælp med

udregning af NPMI-scorerne for bigrammer, Jørg Asmussen for at sikre at oplysningerne om korpusserne under Det Danske Sprog- og Litteraturselskab er korrekte, og Sanni Nimb for løbende opbakning. Endelig skylder vi de to redaktører Michael Nguyen og Jacob Thøgersen og en anonym fagfællebedømmer stor tak for meget grundige kommentarer og forslag til videre analyser. Alle resterende mangler og fejl er alene vores ansvar. Man kan få Idas speciale ved at kontakte hende på mail.

Ida Flörke, assisterende redaktør  
Det Danske Sprog- og Litteraturselskab  
floerkeida@gmail.com

Elisabeth Engberg-Pedersen, professor  
Institut for Nordiske Studier og Sprogvidenskab  
Københavns Universitet  
eep@hum.ku.dk

## ORDBØGER

*DDO (Den Danske Ordbog)*. <https://ordnet.dk/ddo/forside> (tilgået 26. oktober 2021).  
*Merriam-Webster*: <https://www.merriam-webster.com/dictionary/boy> (tilgået 20. oktober 2021).  
*ODS (Ordbog over det Danske Sprog)*. <https://ordnet.dk/ods/forside> (tilgået 26. oktober 2021).

## LITTERATUR

- Asmussen, J. 2006. At måle og veje korpusser: Et aspekt af arbejdet bag de store almensproglige korpusser for dansk. A. Brasch (red.), *Sprogteknologi i dansk perspektiv: En samling artikler om sprogforskning og automatisk sprogbehandling*, 157–177. København: C.A. Reitzel.
- Baker, P. 2014. *Using corpora to analyze gender*. London: Bloomsbury Publishing.
- Bebout, L. 1984. Asymmetries in male-female word pairs. *American Speech* 59(1). 13–30. DOI: 10.2307/454991.
- Bebout, L. 1995. Asymmetries in male/female word pairs: A decade of change. *American Speech* 70(2). 163–185. DOI: 10.2307/455814.

- Bouma, G. 2009. Normalized (pointwise) mutual information in collocation extraction. C. Chiarcos, R. Eckart de Castilho & M. Stede (red.), *From form to meaning: Processing texts automatically* (Proceedings of the Biennial GSCL Conference 2009), 31–40. Tübingen: Gunter Narr Verlag.
- CoREST 2021. <https://korpus.dsl.dk/corest/index.html> (tilgået 26. oktober).
- Croft, W. & D.A. Cruse. 2004. *Cognitive linguistics*. Cambridge: Cambridge University Press.
- Det Danske Sprog- og Litteraturselskab. 2021. *KorpusDK*. <https://ordnet.dk> (tilgået 26. oktober 2021).
- DGCSS. 2021. *LANCHART-korpusset*. Sprogforandringscentret LANCHART, Det Humanistiske Fakultet, Københavns Universitet. <https://dgcss.hum.ku.dk> (tilgået 26. oktober 2021).
- Fillmore, C.J. 1982. Frame semantics. The Linguistic Society of Korea (red.), *Linguistics in the morning calm: Selected papers from SICOL-1981*, 111–137. Seoul: Hanshin Publishing Company.
- Gensim. <https://radimrehurek.com/gensim/models/phrases.html> (tilgået 26. oktober 2021).
- Hansen, K.R. & J.N. Blom. 2017. 12-talspigerne i medierne: Rammesætning og stereotypificering af højtpræsterende piger og kvinder i det danske uddannelsessystem. *MedieKultur: Journal of Media and Communication Research* 63. 102–129. <https://tidsskrift.dk/mediekultur/article/view/26451/146875?acceptCookies=1> (tilgået 26. oktober 2021).
- Harder, P. 2006. Dansk Funktionel Lingvistik. *NyS. – Nydanske Sprogstudier* 34–35. 92–130. DOI: 10.7146/nys.v34i34-35.13459.
- Hendricks, V. & S. Marker. 2019. Må man sige hvid, trans, brun, broget eller pige om en voksen kvinde? *Berlingske* 20. august 2019. <https://www.berlingske.dk/kommentarer/maa-man-sige-hvid-trans-brun-broget-eller-pige-om-en-voksen-kvinde> (tilgået 26. oktober 2021).
- Holmes, J. & R. Sigley. 2001. What's a word like *girl* doing in a place like this? Occupational labels, sexist usages and corpus research. *Language and Computers* 36. 247–263. DOI: 10.1163/9789004334113\_017.
- Jacobsen, L. 1912. *Kvinde og Mand: En Sprogstudie fra dansk Middelalder*. København og Kristiania: Gyldendalske Boghandel Nordisk Forlag.
- Jensen, E.S. 2001. Om subjektivitet som motivation for sproglig forandring. P. Jarvad, F. Gregersen, L. Heltoft, J. Lund & O. Togeby (red.), *Sproglige åbninger: E som Erik, H som Hansen. Festskrift til Erik Hansen, 18. september 2001*, 194–204. København: Hans Reitzels Forlag.

- Jørgensen, J.B. 2013. TV: Her bliver gigantisk, barmfager kvinde påkørt bagfra. *Ekstra Bladet*. <https://ekstrabladet.dk/Haandvaerkeren/article4566475.ece> (tilgået 26. oktober 2021).
- Kerrn-Jespersen, R. 2016. Hvor mange køn findes der? Videnskab.dk. <https://videnskab.dk/krop-sundhed/hvor-mange-koen-findes-der> (tilgået 27. oktober 2021).
- Pearce, M. 2008. Investigating the collocational behaviour of MAN and WOMAN in the BNC using Sketch Engine1. *Corpora* 3(1). 1– 29. DOI: 10.3366/E174950320800004X.
- Rådet for Sikker Trafik. 2016. Ni ud af ti spritbilister er mænd. <https://www.sikkertrafik.dk/presse/pressemeddelelser/ni-ud-af-ti-spritbilister-er-maend> (tilgået 26. oktober 2021).
- Scheuer, J. 1995. Hans hustru og hendes bryster. M. Kunøe & E.V. Larsen (red.), 5. *Møde om Udforskningen af Dansk Sprog, MUDS 5*, 246–265. Aarhus: Institut for Nordisk Sprog og Litteratur, Aarhus Universitet. [https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS\\_5.pdf](https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS_5.pdf) (tilgået 26. oktober 2021).
- Scheuer, J. 2017. Hans mening om hendes krop: Portrætter af to køn i danske tekst-korpusser. I.S. Hansen, T.T. Hougaard & K.T. Petersen (red.), 16. *Møde om Udforskning af Dansk Sprog, MUDS 16*, 343–366. Aarhus: Institut for Kommunikation og Kultur, Aarhus Universitet. [https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS\\_16.pdf](https://projekter.au.dk/fileadmin/projekter/Muds.dk/rapporter/MUDS_16.pdf) (tilgået 26. oktober 2021).
- Sigley, R. & J. Holmes. 2002. Looking at girls in corpora of English. *Journal of English Linguistics* 30(2). 138–157. DOI: 10.1177/007242030002004.
- Slobin, D.I. 1996. From “thought” and ”language” to “thinking for speaking”. J.J. Gumperz & S.C. Levinson (red.), *Rethinking linguistic relativity*, 70–96. Cambridge: Cambridge University Press.
- Tremblay, A. 2012. Empirical evidence for an inflationist lexicon. *Yearbook of Phraseology* 3. 109–126. DOI: 10.1515/phras-2012-0006.