

# Om kjønn og adjektiv: «Trond er så eksepsjonell/nydelig/skjønn/grønn»

Ingebjørg Tonne<sup>1</sup>, Helene Uri og Lars G. Bagøien Johnsen  
*Universitetet i Oslo, forfatter og Nasjonalbiblioteket*

## Abstract

Do female and male writers use adjectives differently? This article is a survey of the potentially gendered use of adjectives in Norwegian novels. It is also a tribute to Trond Tosteruds's legendary article on grammatical gender. While our concern here is biological sex of authors and their use of adjectives, the man of the hour was concerned with rule-governed gender on nouns, albeit with a biological accent. Several readers — and listeners at MONS in 1999 — took note of the bold rule that said that oblong objects, not to mention protruding natural formations, are masculine, while natural phenomena such as pits and cavities are feminine (Trosterud 2001).

Keywords: adjective, statistics, gender, literature

## 1 Introduksjon

### 1.1 Kjønn og språkvitenskapen – en kjappis

For de første språkviterne som skrev om kjønn og språk, var utgangspunktet deres eget språk: mannens språk. Kvinnespråk defineres ut fra forskjellene og avvikene – eller manglene sammenlignet med mannens språk. Det mannlige var normen, det kvinnelige måtte spesifiseres: språk og kvinnespråk – omtrent som det fremdeles gjenspeiles i hverdagspråket: *band* og *jenteband*,  *fotball* og *damefotball*, *professorer* og *kvinnelige professorer*. Otto Jespersen (1941) skriver om både språk og kvinnespråk, og kapitlet om dette siste fenomenet står sammen med kapitler om andre språklige avvik og rariteter: barnespråk, utlendingers språk og pidginspråk. Noen år senere, i 1947, kommer *Manns- og kvinnespråk*, en populærvitenskapelig fremstilling skrevet av den norske filologen Georges Abel. Verken Jespersen eller Abel bygger på annet enn egne spredte observasjoner. Det hindrer dem ikke i å komme med nokså friske uttalelser. Ifølge Abel står det dårlig til med kvinnespråk: mindre ordforråd og enklere setningsbygning. Det vi skal huske når vi, alt etter hvilket kjønn vi innehar, fniser eller humrer av disse og lignende uttalelser, er at de er skrevet i en tid der få kvinner hadde utdannelse og intellektuelt stimulerende yrker.

Robin Lakoff's bok *Language and woman's place* (1975) er en klassiker innen feltet – trass i at boken er introspektiv og ikke-empirisk. Hun startet med denne utgivelsen det som siden er blitt kalt kjønnsforskjellsparadigmet ('the gender differences paradigm'), med en språkbuksteoretisk tilnærming om mannlige dominans, altså at menn bruker språket til å dominere kvinner (Baker 2014). Fishman (1977) mente i forlengelsen av Lakoff's arbeid at kvinner bruker mer krefter på å legge til rette for god kommunikasjon: Hun mente at kvinner gjorde mer interaksjonelt drittarbeid ('interactional shitwork' og 'donkey work'). Senere har mange påpekt at Lakoff overdrev forskjellene mellom kvinnespråk og mannspråk, slik som for eksempel Cameron (2007): «Forty years after Lakoff's groundbreaking work,

<sup>1</sup> Ingebjørg Tonnes bidrag til dette arbeidet er finansiert av Norges forskningsråd gjennom ordningen Sentre for fremragende forskning, prosjektnummer 223265.



we've learned that all such generalizations are over-generalizations: none of them are true for every woman in every context (or even most women in most contexts).»

Deborah Tannens arbeider (f.eks. 1992) representerte noe nytt innen forskningen om språk og kjønn. Hun mener at kvinnens språk er et språk som vektlegger og tar utgangspunkt i et fellesskap, mens mannens språk er et språk som vektlegger uavhengighet, hierarki og status. Tannen understreker at menn og kvinner er oppdratt i forskjellige kulturer, og at dette blant annet har resultert i ulike måter å snakke på. Tannen presenterer undersøkelser som underbygger det synet at barn av ulikt kjønn blir behandlet forskjellig, også språklig.

Forskningen om språk og kjønn kan oppsummeres og skissemessig kategoriseres på denne måten (jf. f.eks. Bull 2021):

**Mangelhypotesen:** Mannens språk er norm, kvinnens språk er avvikende fra normen og derfor per definisjon mangelfullt, slik blant annet Jespersen (1941) fremlegger det.

**Dominanshypotesen:** Ifølge Tove Bull har forskningen på 70- og 80-tallet «eit klart emansipatorisk sikte og såleis eit tydeleg feministisk utgangspunkt» (Bull, 2021, s. 5). Menneskets dominans og makt blir avdekket og betraktet som kvinneundertrykkende. Formålet med forskningen er å påpeke og analysere forskjellene mellom kvinnens og mannens språk. Kvinnespråket er avmaktens språk og en avspeiling av hennes stilling i samfunnet. Hun bruker språket for å kompensere, for å gjøre seg synlig og oppnå en viss form for respekt. Representanter for denne retningen er blant andre Lakoff og Fishman.

**Forskjellshypotesen:** Menn og kvinner vokser opp i og lever i forskjellige kulturer, og derfor er språket deres ulikt. Tannen skriver innenfor denne tradisjonen.

Felles for disse tre forskningstradisjonene er at det er forskjellene mellom kjønnene som vektlegges, selv om forskjellene innad i en kvinnegruppe eller innad i en mannsguppe vil være store – kanskje like store som mellom de to gruppene.

Begrepet kjønn har endret seg og endrer seg stadig. Det er vanskelig å betrakte kjønn som en uproblematisk, todelt størrelse som menneskeheten deles inn etter, noe man har eller er. Kjønn er noe som blir skapt og gjenskap, imitert og utfordret kontinuerlig: «Vi skaper oss selv. Språk er et av verktøyene vi bruker for å gjøre dette. I nyere arbeider om språk og kjønn undersøkes også 'fotnotene' og 'unntakene'; man er ikke lenger interessert i bare den typiske heterofile mann og den typiske heterofile kvinne, men også i guttejenta og jentegutten, i transseksuelle, i homoseksuelle. Man undersøker ulike språkpraksiser som etablerer og omskaper kjønnsbegrepene» (Uri 2018).

Vi skal ikke problematisere todelingen ytterligere her. I denne festskriftsammenheng er vi tre jordnære språkvitere som har gjennomført en jordnær, liten test. For oss har det rett og slett handlet om adjektiv og om kjønn i den tradisjonelle, binære forstand – kvinne og mann – for slik var nå engang materialet vi har til rådighet, klassifisert. Med en kvantitativ tilnærming blir det vanskelig å få øye på brudd med eller sjatteringer i binariteten. Men vi vil likevel allerede nå peke på at den metoden vi har valgt, åpner for at en tradisjonell tdeling *kan* nyanseres. Det er viktig å huske at kvantitative studier som har som formål å avdekke forskjeller mellom kjønnene, også kan tilsløre forskjellene innad i kjønnskategoriene. Metoden vår bidrar til å vise om en forskjell er persistent innad i kjønnskategoriene – eller ikke. Det gjør vi ved å undersøke om mønsteret i helheten gjenspeiler seg i delene.

Bull (2021) hevder at den kvantitative variasjonslingvistikken er svekket, og at «vi faktisk visste meir om statistiske forskjellar i språket og språkbruken til kvinner og menn, gutar og jenter på 1980-talet enn vi gjer i dag». Vi bidrar med vårt lille arbeid til å gi et svar på de siste tiårenes manglende interesse for kvantitativ forskning og kjønn.

## 1.2 Kjønn og adjektiv

Jespersen (1941) var overbevist om at «Kvinnerne føler langt større trang til at gi deres varme følelser uttrykk i sproget, og da de hvert øyeblikk kan komme i begejstring selv overfor mange av livets småting, er det let forståelig, at de kommer til at slide mere på sprogets rosende ord end mændene». Abel (1947) på sin side var overbevist om at kvinnene har en hang, trolig medfødt, til overdrivelser. Det er mulig å forestille seg at både den trangen Jespersen mener å se hos kvinner, og den hangen Abel ser, kan gi seg

utslag i mange adjektiv hos kvinnelige språkbrukere. Vi himler med øynene og fniser (Lars humrer) når vi leser om Jespersens og Abels syn på kvinnespråk, men kan hende skiller ikke synet seg radikalt fra kjønnsstereotype forestillinger i dag. På høstparten 2021 uttalte en 60-årig mann, la oss kalle ham mannen i gata, om kvinners språk at det er «utflytende og beskrivende, vårt språk er mer kortfattet og to-the-point. Mindre dill og pynt og adjektiv».

Lakoff (1975) hevder at kvinner oftere enn menn bruker adjektiv som *charming*, *divine* og *cute* for å uttrykke følelser – det minner jo om Jespersens «varme følelser». Dessuten mener hun at menn (med unntak av homofile og interiørarkitekter) ikke gir uttrykk for et like finindelt fargespekter som kvinner: «words like beige, ecru, aquamarine, lavender, and so on are unremarkable in a woman's active vocabulary, but absent from that of most men» (ibid, s. 43). Det er heller ikke vanskelig å finne anekdotisk støtte for dette, her fra en kvinne i gata: «Min erfaring er at kvinner generelt har et mye større fargevokabular enn menn. Vi snakker om akvamarin, kobolt, lavendel, oker, terrakotta, lime, oliven, plomme, krem, muldvarp og røkfarget» (kvinne i Facebook-gruppen Språkspalta, sitert fra Uri 2018). Men ved nærmere ettertanke vil nok de fleste tenke som denne kvinnen, som da hun bad sin mann om å kjøpe turkise stearinlys, fikk rosa lys (ibid): «Men jeg kan love at han kan fargen turkis nå! Når det er sagt, kan faren min navnet på flere fargenyanser enn noen andre jeg kjenner. Dette har vel mest med interesser å gjøre? Far driver med rosemaling.»

### 1.3 Inspirasjon og tidligere studier

I tillegg til Trosteruds filurkattlignende smil under MONS i 1999 er Ben Blatts (2017) bok *Nabokov's favourite word is mauve* vår viktigste inspirasjonskilde. Blatt undersøker engelskspråklig skjønnlitteratur ved å sette sammen store korpus av litterære tekster, og en del av det han gjør, er knyttet til forfatterkjønn. For eksempel finner han at i femti romanklassikere skrevet av mannlige forfattere er det bare seks av romanene som har flere forekomster av *she* enn av *he*. Tilsvarende – og altså motsatt – tall for romaner skrevet av kvinner er 21 bøker: Nesten halvparten av kvinneverker har flere forekomster av *he* enn av *she*. Og mer deprimerende: Blatt finner det samme mønsteret i nytutgitte, engelskspråklige romaner. Det står ikke noe bedre til i Norge (jf. Uri 2018, Johnsen og Uri 2021). Kjønnen på de fiktive karakterene i romanene spiller òg en rolle. Blatt finner blant annet at enkelte verb særlig brukes når subjektet er kvinne og andre når subjektet er mann. Lignende mønster ble også funnet i norske tekster (Uri 2018, Dyvik 2018) – i norske tekster hulker kvinner mer enn det menn gjør, mens menn på sin side (som Lars ovenfor) humrer betydelig oftere. Blatt ser også på fordeling og frekvens av ord i engelskspråklige romaner ut fra forfatterkjønn. Resultatene hans er ikke overraskende, og fordelingen av ord signaliserer mer enn noe annet tradisjonelle domener, og enda mindre overraskende er det når vi tar inn over oss at mange av romanene Blatt har søkt i, er fra 1800-tallet. Ordene *pillows*, *lace*, *curls* tyder på at romanen trolig er skrevet av en kvinne. Ordene *chief*, *rear*, *civil* peker i retning av en mannlige forfatter. Blatt serverer oss to lister med henholdsvis de ti mest typiske dameordene og de ti mest typiske herreordene. Adjektivfangsten er imidlertid liten – ingen adjektiv hos kvinnenes ti på topp – og adjektivene begrenser seg til disse tre på herrenes liste: *civil*, *bigger*, *public*.

En stor amerikansk undersøkelse av Facebook-oppdateringer (Park m.fl. 2011) viser at kvinner bruker flere positivt ladde adjektiv enn det menn gjør. Det samme viser en liten norsk studie av journalist-språket (Johnsen og Karlsen 2017).

Svendsen (2019) diskuterer undersøkelser som har tatt utgangspunkt i Robin Lakoffs studie, men der undersøkelsene er gjennomført med mer robuste metoder, særlig korpusbaserte. Disse undersøkelsene viser ifølge Svendsen at språkbruksbildet er komplekst og sammensatt når det gjelder hvordan kvinner og menn velger ord, og at det kan være vanskelig å skjelle mellom hva som kan skyldes kjønn, og hva som kan skyldes andre faktorer, som utdanningsnivå, etnisitet og alder (Svendsen 2019).

For engelsk har også Barczewska og Andreassen (2018) ved hjelp av korpusstudier forsøkt å besvare spørsmål som ligner våre, om adjektivbruk og kjønn. Deres undersøkelse er motivert av det de omtaler som uavklarte og selvmotsigende resultater fra tidligere studier: «Although a great number of scholars highlight the differences between women's and men's speech patterns, there are studies which call into question the conclusion of differentiating between female-male styles of communication. [...] Lakoff

(1973; 1975), Brandis and Henderson (1970), Entwisle and Garvey (1972), Hartman (1976), Bamman et al. (2014), Schmid (2003) and Amir et al. (2012) argue that adjective use differs between the sexes, whereas Kr[a]mer (1974) and Cameron (2008) claim that it does not, and Bamman et al. (2014)'s study seems to suggest that not all differences align with the stereotypes as men may be equally, if not more, expressive than women. Baker (2014) is hesitant to come down strongly in support of either position.» (Barczewska og Andreasen 2018, s. 201).

Mangelen på avklaring i tidligere forskning knytter Barczewska og Andreasen i noen grad til det de mener er den tidlige forskningens overdrevne ekstrapolering fra små mengder data og anekdotiske bevis fra muntlig språkbruk. Forskernes intuisjoner, brukt som metode, kan ha gitt de sprikende resultatene, fremholder de, og mener derfor at korpusbaserte studier er det som skal til. De er dermed på linje med Baker (2014), som i sin bok kombinerer korpuslingvistik og kjønnsanalyse. Og da blir spørsmålet hva slags språklig korpus som brukes (muntlig eller skriftlig, for eksempel) og hvor stort (eller variert) det er. Dessuten er forskningsspørsmålet selvsagt viktig. Det er særlig to typer spørsmål eller tilnærminger som er interessante når det gjelder kjønn og korpusstudier, som Baker (2014) skisserer det. På den ene siden undersøker man kjønn og språkbruk, det vil si at man vil finne ut hvordan de to kjønnene selv bruker språket. På den annen side kan man ta for seg spørsmål om representasjon av kjønn gjennom språk, altså hvordan kjønnene omtales. Det er den første typen tilnærming som er vårt anliggende i denne artikkelen, slik det er for Barczewska og Andreasen som vi nå ser litt nærmere på.

Barczewska og Andreasen bruker korpuset MICASE (Michigan Corpus of Academic Spoken English), et korpus med transkribert tale fra akademiske situasjoner i perioden 1997 til 2002, med ca 1,85 millioner ord. Korpuset har ulike variabler markert, som type interaksjon, om talerne er studenter eller akademisk ansatte, grad av engelskkompetanse og: kjønn. Kvinner står for 54 % av ordmengden i korpuset, mens menn står bak 46 %. Barczewska og Andreasen understreker at de to kjønnene er jevnt fordelt på de ulike situasjonstypene, slik at for eksempel kvinnestemmene ikke først og fremst er fra lunsjpausene mens mannsstemmene foredrar om egne forskningsresultater. Barczewska og Andreasen stiller følgende forskningsspørsmål: Bruker kvinner flere og et bredere utvalg av adjektiv enn menn? De tar for seg åtte adjektiv, som de kaller «basic adjectives»: *good, bad, big, small, pretty, ugly, different, important*. Til disse grunnadjektivene legger de en mengde (nær-)synonymer. Her er nettopp synonymene interessante, i og med at de søker å undersøke om kvinner bruker et bredere utvalg adjektiv. (Nær-)synonymene har de funnet i Thesaurus.com. De er, for eksempel for *good*: *acceptable, fine, excellent, exceptional, favorable, great, marvelous, positive, satisfactory* og *wonderful*, mens det for eksempel for *ugly* er *awful, grisly, hideous, horrid, unseemly, unsightly, disgusting, terrible, gross* og *unpleasant*.

De finner at kvinnene i korpuset har en høyere adjektivfrekvens enn mennene, og særlig bruker kvinnene flere av de positivt ladde adjektivene enn de negativt ladde (både grunnadjektivene og nær-synonymene). Men menn bruker også flere positivt enn negativt ladde adjektiv. Bruken av adjektivet *marvelous* trekkes fram som overraskende og på tvers av de mer generelle funnene, idet det er mennene som bruker dette mest. På den annen (negative) side finner de at mennene i undersøkelsen bruker adjektivet *wrong* mer enn kvinnene. For øvrig bølger det litt frem og tilbake i funnene: For eksempel bruker kvinner flest nærsynonymer for *big*, mens menn bruker flest nærsynonymer for *small*. I motsetning til hva Lakoff fant, eller snarere mente, finner de at menn bruker *charming* og *lovely* oftere enn kvinner. Konklusjonen til Barczewska og Andreasen er at de finner forskjeller i adjektivbruk hos kvinner og menn, men også at det er en god del likheter. De avslutter slik: «The mixed results of this study, although limited to one genre, do suggest that Lakoff's intuitions, and the stereotypes they represent, do not always hold and more work should be done into the notion of 'feminine' or 'masculine' adjectives.» (s. 211).

Så da griper vi den stafettspinnen (og legger samtidig merke til at dette avlange objektet er et maskulinum).

#### 1.4 Det vi lurte på

Vi søker etter adjektiv i store, norskspråklige skjønnlitterære tekstkorpus, og vi spør: I hvilken grad kan forfatteres bruk av adjektiv kobles til forfatterens kjønn? Bruker kvinner adjektiv oftere enn det menn

gjør? Er adjektivinventaret et annet hos menn enn hos kvinner? Stemmer det at kvinner bruker flere positivt ladde adjektiv enn det menn gjør? Bruker menn færre fargeord enn det kvinner gjør?

I vår undersøkelse speiler vi delvis Barczewska og Andreasens studie av engelsk muntlig språkbruk. Vi tar utgangspunkt i bokmåloversettelser av seks av de åtte engelske adjektivene fra studien deres, inkludert nærsynonymene. Vi bruker et stort korpus, men vi dykker også ned i mindre delkorpus for å se om adjektivfordelingen i det store korpuset gjentar seg når det brytes ned.

## 2 Metode

Nasjonalbiblioteket har digitalisert hele den norske litteraturen og gjort bøkene tilgjengelig for søk. Den skjønnlitterære delen består i skrivende stund av rundt 55 000 bøker<sup>2</sup>. Digitale søk i store mengder bøker kan avsløre mønstre som for alle praktiske formål ellers ville ha vært skjult for oss. Det er stor variasjon individuelt i språkbruk, men når tekstmengden er stor nok, er det mulig å si noe om trender i bruken, og det er også mulig å dykke ned i mengden for å finne eventuell variasjon. Undersøkelsen vår er korpusbasert, som vil si at man bruker korpus for å sjekke ut om visse hypoteser holder i store språklige tekstsamlinger (Baker 2014, s. 15–16).

### 2.1 Korpusdata

Datagrunnlaget i undersøkelsen er sammensatt: Det dreier seg om et kuratert, høylitterært korpus som består av hundre norskskrevne romaner, der de eldste bøkene er fra 1890-årene og de nyeste fra 2015. Dessuten har vi brukt korpus hentet fra Nasjonalbibliotekets samlinger, valgt ut ved hjelp av tilgjengelige metadata. Vi har plukket ut de bøkene publisert mellom 1960 og 2010 som er klassifisert under deweynummer 839 (skjønnlitteratur), har målform bokmål, og som har informasjon om kjønnet til forfatteren. Resultat er om lag 21 000 bøker. Den klassen av bøker refererer vi til som «totalkorpuset» eller «totalen». Fra totalkorpuset har vi så konstruert, altså «samplet», fem vilkårlige korpus, hver på 2000 skjønnlitterære bøker.<sup>3</sup>

Felles for det kuraterte korpuset, totalkorpuset og de fem samplede korpusene er at halvparten av bøkene er skrevet av kvinnelige forfattere, halvparten av mannlige. De samplede korpusene er fra perioden 1960 til 2010 og er laget ved at vi fem ganger har stukket neven ned i den digitale haugen med bøker skrevet av kvinner og like mange ganger ned i den tilsvarende haugen med bøker skrevet av menn. En håndfull er altså tusen bøker, så de ti håndfullene som utgjør de fem samplede korpusene, består av til sammen 10 000 bøker. Det kuraterte korpuset er bygd opp med romaner skrevet av anerkjente, norske forfattere. I sample-korpusene og totalkorpuset er det mer blandet, en del er oversatt, og mange av bøkene minner om dem vi har stående på hytta: Her er det både romantikk, erotikk og western. Mens det er 16 forlag, samtlige såkalte seriøse, i sving for det kuraterte korpuset, er det 1523 ulike forlag for totalen – flere av dem med utgivelseslister spekket av underlødige titler.

Vi tar ikke ut forfatterinformasjon annet enn kjønn, slik at hver bok i undersøkelsen ikke har noe annen metadata knyttet til seg. Det er ingen gruppering av bøkene utover forfatterens kjønn. Bøkene bidrar i sin tur til helheten med relative frekvenser. Om boken er stor eller liten, vil det kun være proporsjonen av adjektiv den bidrar med i tellingen frem til totalen for et korpus. Et korpus måles med gjennomsnittet av de relative frekvensene.

### 2.2 Språkdata

Våre seks oversatte adjektiv fra Barczewska og Andreasens studie er *god*, *dårlig*, *stor*, *liten*, *pen* og *stygg*. I det følgende kaller vi disse grunnadjektivene for B&A-adjektivene. B&A-adjektivene er knyttet til følgende seks sett nærsynonymadjektiv som telles opp samlet innad i gruppen:

<sup>2</sup> Det er tall slik de fremkommer gjennom nb.no i mars 2022 ved å velge kategorien skjønnlitteratur på enten bokmål eller nynorsk.

<sup>3</sup> Alt av data og programkode ligger tilgjengelig via Johnsen (2022).

**God:** akseptabel, fin, utmerket, eksepsjonell, gunstig, flott, fantastisk, positiv, tilfredsstillende

**Dårlig:** grusom, forferdelig, fryktelig, elendig, uakseptabel, dårlig, feil

**Stor:** kolossal, betydelig, gigantisk, massiv, enorm

**Liten:** trang, begrenset, mager, mikroskopisk, beskjeden, kort

**Pen:** vakker, sjarmerende, elegant, grasiøs, kjekk, nydelig, ryddig, attraktiv, søt

**Stygg:** upassende, ekkel, grov, ubehagelig

I valget av fargeadjektiver har vi tatt utgangspunkt i Newtons fargesirkel (unntatt det litt mindre frekvente ordet *indigo*): *blå, fiolett, grønn, gul, oransje* og *rød*, og så har vi lagt til *hvit* og ellers en del nyanser: *beige, burgunderrød, grå, lilla, rosa* og *turkis*.

De adjektivene vi henter ut kvantitative data for, er utvidet med informasjon fra Norsk Ordbank<sup>4</sup> slik at hvert adjektiv bidrar med en liste av fulle former. Det er de fulle formene som er telt mot korpusene og summert opp for hvert enkelt adjektiv. For eksempel vil adjektivet *god* bidra med ordformene 'beste', 'best', 'gode', 'bedre', 'godt' og 'god', og alle bøyingsformer blir med i analysen.

### 2.3 Eksperimenter

Vi sammenligner totalkorpuset med det kuraterte korpuset og de samlede delkorpusene i tre eksperimenter og vurderer holdbarheten ved å se på variasjonen mellom korpusene. I det første eksperimentet telles B&A-adjektivene med nærsynonymer for henholdsvis kvinner og menn. I eksperiment 2 telles for henholdsvis kvinner og menn bare nærsynonymer. I det tredje eksperimentet telles fargeadjektivene fordelt etter kjønn. I tellingen måles adjektivene med relativ frekvens, altså antall adjektiv per ti tusen ord. Informasjonen er tilgjengelig via Nasjonalbibliotekets DH-lab<sup>5</sup>.

Om vi finner at et adjektiv brukes mer av kvinner enn av menn i totalkorpuset, ønsker vi å undersøke om det er en faktisk kjønnsforskjell, eller om det for eksempel er noen få forfattere som gjennom en meget omfattende bruk av visse adjektiv gir oss et inntrykk av at det er en kjønnsforskjell. Det undersøker vi ved å splitte opp korpuset i mindre biter for å se om forskjellen fremdeles er til stede i de mindre korpusene. Mens en vanlig statistisk praksis er å resonnerer fra del til hele, prøver vi altså her i vår studie å si om de forskjellene som observeres i helheten, også er til stede i delene. Resonnementet blir at dersom forskjellen består på tvers av korpusene, så kan forskjellen tilskrives til kjønn.

I tillegg til å sammenligne helheten med delene vil vi se på i hvor stor grad fordelingen av adjektivene i statistisk forstand viser forskjeller. Hvis enkelte adjektiv har større forekomst hos kvinner enn hos menn, kan vi for det første spørre om hvor store de observerte frekvensforskjellene er (statistisk effektstørrelse), men også i hvor stor grad fordelingene er separert fra hverandre, målt i termer av varians eller standardavvik. Hver bok gir adjektivet en frekvens.<sup>6</sup> En gruppe bøker, for eksempel et av våre korpus, vil tilordnes et gjennomsnitt av frekvensene for hver bok som inngår i gruppen; samtidig vil det være variasjon innad i gruppen rundt det gjennomsnittet. Den variasjonen måles i standardavvik.<sup>7</sup> Ved å benytte standardavvik kan vi måle i hvor stor grad to grupper av bøker er separert fra hverandre: Separasjon er knyttet til hvor mange standardavvik det er mellom gjennomsnittet av gruppene; jo flere standardavvik, desto høyere separasjon.

Separasjonen som mål er dermed distinkt fra den gjennomsnittlige forskjellen i frekvens mellom observasjonene. Et annet alternativ for separasjon kunne ha vært å benytte en statistisk test og se på såkalte p-verdier, altså sannsynligheter for observasjoner. Men ulempen med dette siste alternativet er at p-verdier er sensitive for gruppenes datastørrelse, slik at samme gjennomsnitt og standardavvik kan gi lave eller høye p-verdier avhengig av hvor mange bøker som inngår i et (del)korpus. I og med at det også

<sup>4</sup> Norsk Ordbank er beskrevet her <https://www.nb.no/sprakbanken/ressurskatalog/oai-nb-no-sbr-5/>

<sup>5</sup> Les mer om DH-lab ved Nasjonalbiblioteket og dhlab på <https://nb.no/dh-lab>

<sup>6</sup> Mer presist - en relativ frekvens, altså hvor stor proporsjon adjektivet tar i boken.

<sup>7</sup> Konsulter gjerne en innføring i statistikk for presise definisjoner og formler.

er variasjon i størrelse på korpusene i undersøkelsen, har vi derfor heller valgt å sammenligne dem gjennom gjennomsnitt av frekvenser og standardavvik fremfor gjennomsnitt av frekvenser og p-verdier.

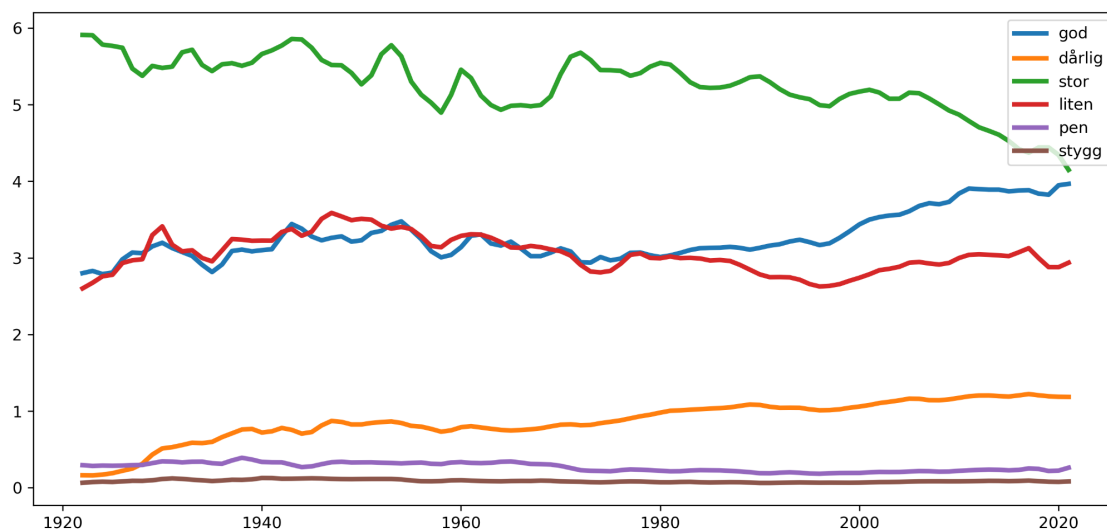
Graden av separasjon presenteres i tre tabeller. De vises som fargekart som letter lesning av mønstre, og for å finne avvikende og lignende tall. Tallene i tabellene er antall standardavvik mellom kjønnene i et korpus. Fargeleggingen er slik at intensiteten følger kolonnen, der adjektivet med størst separasjon vil stå frem med den mørkeste avskygningen. Se for eksempel tabell 2 nedenfor.

### 3 Resultater

I vår undersøkelse spør vi altså om kvinner bruker flere adjektiv enn menn, og om kvinner bruker andre adjektiv enn menn. Med adjektiv forstås her listene av adjektiv gitt ovenfor (B&A-adjektivene, nærsynonymene og fargeadjektivene). Som nevnt telles B&A-adjektivene med nærsynonymer i eksperiment 1. I eksperiment 2 telles bare nærsynonymer. I det tredje eksperimentet telles fargeadjektivene. Vi gyver løs.

#### 3.1 Eksperiment 1

I dette eksperimentet teller vi B&A-adjektivene og nærsynonymene og sammenligner frekvensen hos mannlige og kvinnelige forfattere. B&A-adjektivene, altså *god*, *dårlig*, *stor*, *liten*, *pen* og *stygg*, har hatt en relativt stabil frekvens i norsk de siste hundre årene som vist i figur 1 nedenfor med en graf fra NB-N-gram som viser trenden for alle bøker uansett sjanger. Selv om trendlinjene for *stor* og *god* endrer seg mest, og tilsynelatende på bekostning av hverandre, er det liten variasjon i frekvens. Eventuelle forskjeller mellom adjektivene kan tilskrives andre variabler. For eksempel er det slik at trendlinjen for adjektivet *dårlig* endrer seg en del fra 1920 til 1930, noe som skyldes etterheng i forbindelse med overgangen fra *aa* til *å* i norsk.

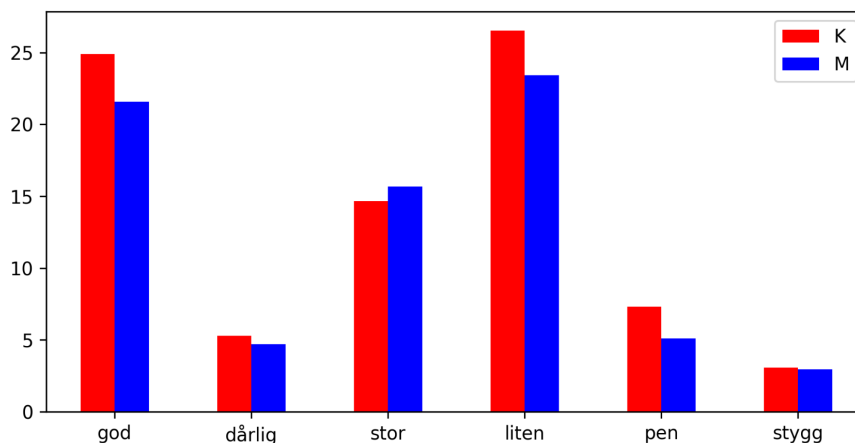


Figur 1: Trendlinjer for adjektivene i studien. Tallene på y-aksen er frekvens pr. 10 000 løpeord.

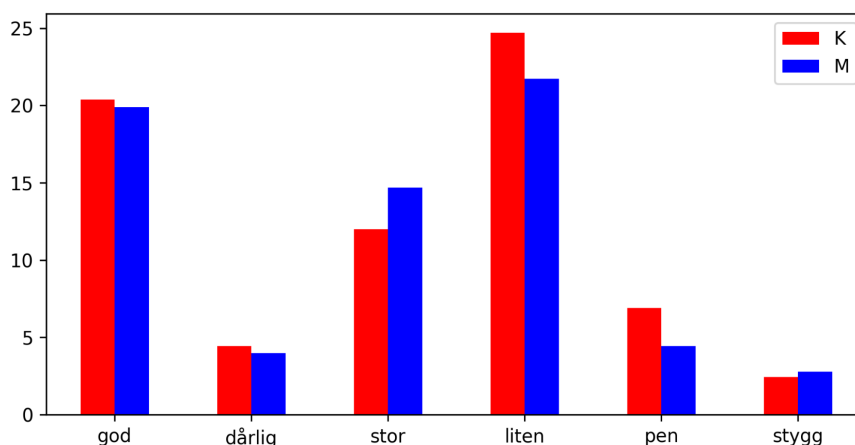
Figur 2 nedenfor viser gjennomsnittlig bruk av adjektiv i totalkorpuset for henholdsvis kvinner og menn, mens figur 3 viser adjektivbruken i det kuraterte korpuset. Her er B&A-adjektivene og nærsynonymene telt, slik at nærsynonymene er innbakt i søylene for sine respektive grunnadjektiv. Alle bøyingsformene telles med. Summen er relativisert til antall ord i teksten de er telt opp i, slik at vi for eksempel ser at alle

formene for *liten* (og nærsynonymene til *liten*) til sammen forekommer ca. 25 ganger<sup>8</sup> pr. 10 000 ord, det vil si omtrent ett adjektiv pr. bokside.

Frekvens av adjektiv fordelt på forfatterens kjønn viser seg å være lik i de to korpusene, slik det fremgår av figurene 2 og 3. De kvinnelige forfatterne bruker B&A-adjektivene (med sine respektive nærsynonymer) noe mer enn menn. Unntakene er adjektivet *stor* som gjennom sine nærsynonymer brukes noe mer av menn enn av kvinner i begge korpusene, og *stygg*, som sammen med nærsynonymene brukes litt mer av menn i det kuraterte korpuset.



Figur 2: Gjennomsnittlig bruk av adjektiv i totalkorpuset, for henholdsvis kvinner (K) og menn (M). Her er B&A-adjektivene og nærsynonymene telt. Y-aksen viser antall adjektiv pr. 10 000 løpeord.



Figur 3: Søylene viser gjennomsnittet for B&A-adjektivene og nærsynonymene, for henholdsvis kvinner (K) og menn (M), i det kuraterte korpuset. Tallene langs y-aksen er antall adjektiv pr 10 000 løpeord.

<sup>8</sup> Økning av den prosentvise andelen sammenlignet med NB-ngram skyldes at i søylediagrammene er alle bøyningsformene både for adjektivet og for nærsynonymene med i optellingen.



Vi sammenligner i tabell 1 nedenfor også de fem samlede korpusene med hverandre for å undersøke variasjonen av enkeltadjektiv mellom hvert av korpusene. Mens figur 2 gir oss et overblikk over helheten i totalkorpuset, spør vi nå om delene, samplene fra totalkorpuset, oppfører seg likedan. Hvis de gjør det, gir det en indikasjon på om adjektivene er typiske for en kjønnskategori.

For hvert adjektiv måles forskjellen mellom kjønnene som en ratio, der den mest frekvente kategorien divideres på den mindre frekvente. Alle tallene er derfor større enn 1, og jo større tall, desto større forskjell.

	1	2	3	4	5
<b>god</b>	1.2 ♀	1.2 ♀	1.1 ♀	1.2 ♀	1.2 ♀
<b>dårlig</b>	1.2 ♀	1.2 ♀	1.1 ♀	1.1 ♀	1.1 ♀
<b>stor</b>	1.1 ♂	1.0 ♂	1.0 ♂	1.0 ♂	1.0 ♂
<b>liten</b>	1.1 ♀	1.1 ♀	1.1 ♀	1.1 ♀	1.1 ♀
<b>pen</b>	1.6 ♀	1.4 ♀	1.5 ♀	1.5 ♀	1.4 ♀
<b>styg</b>	1.0 ♀	1.1 ♀	1.0 ♂	1.1 ♀	1.0 ♀

Tabell 1. Hvem bruker adjektiv mest? Tallene viser forskjellen i snitt mellom kvinner og menn målt i ratio, der den mest frekvente kategorien divideres på den mindre frekvente. Alle tallene er derfor større enn 1, og jo større tall, desto større forskjell mellom kjønnene. Røde tall med symbolet ♀ viser til kvinner, blå tall med symbolet ♂ viser til menn. Tallene (1-5) over kolonnene refererer til de fem samlede korpusene.

Resultatene i tabell 1 viser tydelig forskjell mellom kvinner og menn. Det kan være verdt å ha i bakhodet at om det var vilkårlig hvilken kjønnsgruppe som fikk høyest score, vil det være litt over 3 % sjans for at en hel rad skulle være helt lik. I tabell 1 er hver rad sær lik innad. Legg også merke til raden for *stor*, der frekvensforskjellen er forholdsvis liten<sup>9</sup>, men likevel overvekt i mannlig retning i hvert delkorpus. *Styg* med nærsynonymene *upassende*, *ekkel*, *grov*, *ubehagelig*, som hadde ulik fordeling i totalkorpuset og det kuraterte, bryter mønsteret med ett kjønn over hele linja – bokstavelig talt! Dette er ikke akkurat adjektiv som vi assosierer med jubilanten, men vi vender tilbake til det senere.

I tabell 2 nedenfor gis det første av tre datasett for separasjonsverdier. Tabellen viser hvor mange standardavvik som skiller mellom ♀ og ♂ illustrert i et fargekart, og bør sees i sammenheng med tabell 1 og søylediagrammene i figur 2 og 3. Forskjellen i frekvens (effektstørrelse) kan være stor eller liten og kan variere uavhengig av separasjon, selv om en kan forvente at små forskjeller i frekvens også går sammen med lav separasjonsgrad,<sup>10</sup> altså at om forskjellen er liten den ene veien, kan den lettere gå andre veien, om alt annet er likt. Vi ser for eksempel at separasjonsverdiene for *stor* passer godt sammen med den lille forskjellen i frekvens det er mellom kjønnene for det adjektivet, og at det er det kuraterte korpuset som har høyest grad av separasjon. Samme korrespondanse mellom lav forskjell i frekvens og lav separasjon observeres for det kuraterte korpuset for *god* og *dårlig*, der begge får lav forskjell i frekvens og lav grad av separasjon.

<sup>9</sup> Så liten at forskjellen har forsvunnet i avrundingen.

<sup>10</sup> Det bør sies at alle separasjonsverdiene er lave i og for seg. Med lav forstås her lav sammenlignet med de andre verdiene. En forskjell i standardavvik regnes vanligvis som stor om det er snakk om to eller mer, altså tre til fire ganger så stor som de høye verdiene vi har i våre data.

## OM KJØNN OG ADJEKTIV

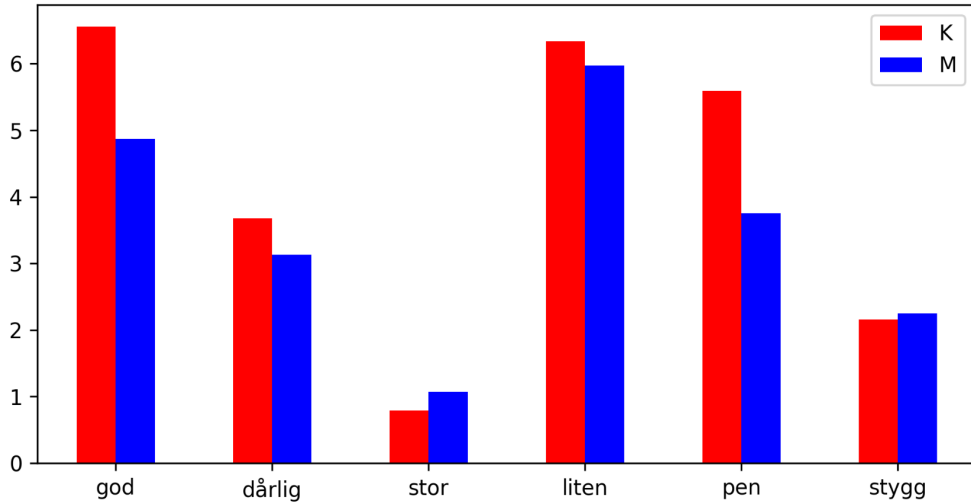
	Total	Kuratert	1	2	3	4	5
god	0,29	0,3	0,31	0,32	0,33	0,34	0,35
dårlig	0,2	0,22	0,24	0,26	0,28	0,3	0,32
stor	0,11	0,4	0,09	0,06	0,04	0,06	0,05
liten	0,21	0,36	0,25	0,22	0,2	0,23	0,23
pen	0,43	0,75	0,51	0,45	0,46	0,51	0,44
stygg	0,06	0,17	0,02	0,09	0	0,06	0,04

Tabell 2: En oversikt over separasjonsverdier i de forskjellige korpusene for S&A-adjektivene med nærsynonymer. Kolonnene indikerer hvilke korpus verdiene kommer fra.

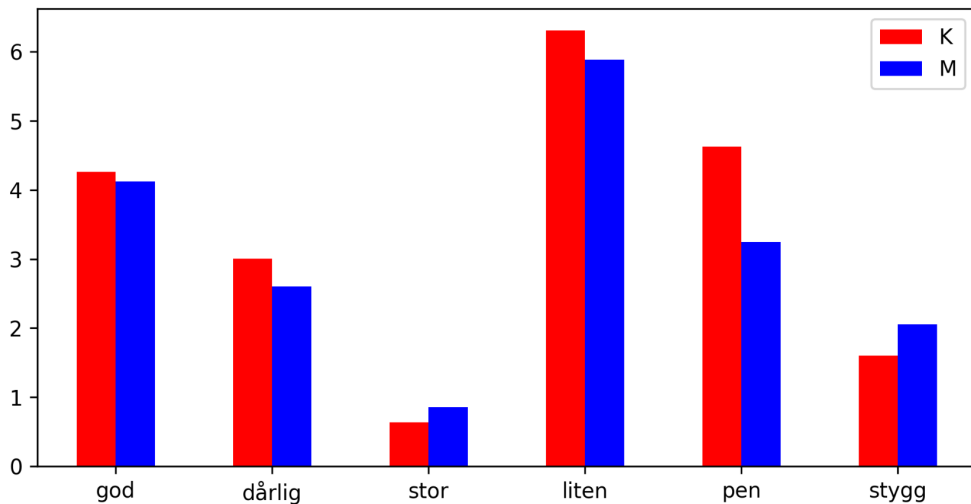
Separasjonstabellene og frekvenstabellene som de fremgår i tabell 1 og 2, samt figur 2 og 3, brukes i resonnementet på den måten at høye verdier forsterker hverandre i retning av en konklusjon om at adjektivene er knyttet til det ene eller andre kjønn. Fra tabellene og figurene over kan vi konkludere at *pen*, *god*, *dårlig* og *liten* er knyttet til kvinnelige forfattere, og at adjektivet *stor* er knyttet til mannlige: Figurene 2 og 3 viser at *pen*, *god*, *dårlig* og *liten* frekvensmessig er i kvinnenens favør med (relativt) høye separasjonsverdier, mens *stor* går til mannlige forfattere med forholdsvis liten separasjon i total- og delkorpusene, men høy for det kuraterte. Adjektivet *stygg* har vekslende data i sammenligningen og har også gjennomgående lave separasjonsverdier, og peker dermed ikke på noe kjønn. For eksperiment 1 kan vi si at kvinner bruker adjektivene mer enn det menn gjør, unntatt adjektivet *stor* som menn bruker mest, og *stygg* som ikke kan knyttes til noen av kjønnene.

### 3.2 Eksperiment 2

Som i eksperiment 1 over går vi her gjennom opptellingen i alle korpusene, altså totalkorpuset, det manuelt kuraterte korpuset og det samlede, men nå begrenser vi oss til nærsynonymene og unntar selve grunnformen fra telling. Figur 4 og 5 nedenfor viser resultater for søk i henholdsvis totalkorpuset og det kuraterte korpuset, igjen henholdsvis for kvinner og menn. Søylene er merket med grunnadjektivet, men det er altså nærsynonymene for dette adjektivet som er undersøkt – det betyr for eksempel at søylen merket *pen* viser resultatet av tellingen av *vakker*, *sjarmerende*, *elegant*, *grasiøs*, *kjekk*, *nydelig*, *ryddig*, *attraktiv*, *søt*. De nærsynonyme adjektivene forekommer betydelig sjeldnere enn grunnadjektivene.



Figur 4: Opptelling av nærsynonymer i totalkorpuset for henholdsvis kvinner (K) og menn (M). Antall pr. 10 000 løpeord.



Figur 5: Opptelling av nærsynonymer i det kuraterte korpuset, for henholdsvis kvinner (K) og menn (M). Antall pr. 10 000 løpeord.

Figur 4 og 5 viser at resultatene av søkene i de to korpusene er ganske like, både når det gjelder bruksfrekvensen for adjektivgruppene og når det gjelder fordelingen mellom kvinner og menn. Sammenlignet med resultatene i eksperiment 1 (jf. figurene 2 og 3) er det en betydelig nedgang for gruppen til *stor*, som har gått fra å være den største gruppen til å bli den minste. Det betyr at forfatterne – av begge kjønn – foretrekker å bruke adjektivet *stor* fremfor nærsynonymene, mens forfatterne for de andre adjektivene bruker både grunnadjektivene og nærsynonymene. Igjen ser vi kvinnes preferanse for det som kan beskrives som lite, pent og godt. De adjektivene menn har en (svak) tendens til å bruke mer av, sorterer under kategoriene *stygg*, *dårlig* og *stor*.

## OM KJØNN OG ADJEKTIV

I tabell 3 under er delkorpuserne telt opp og sammenlignet. Den tabellen viser en stor forskjell mellom menn og kvinner for nærsynonymene til *stor*. Nærsynonymene til *stor* – *kolossal*, *betydelig*, *gigantisk*, *massiv*, *enorm* – brukes betydelig mer av menn enn av kvinner. Adjektivet *stygg* forsterker det som ble antydnet i eksperiment 1, altså at menn bruker det mer, altså noe forsterket her (blå, men lave tall hele veien) hvor det bare dreier seg om nærsynonymene – *upassende*, *ekkel*, *grov*, *ubehagelig*. For de resterende adjektivene i hvert delkorpus er det kvinnene som bruker nærsynonymene til grunnadjektivene mest, og kan altså sies å ha en større spredning i adjektivinventaret.

	1	2	3	4	5
<b>god</b>	1.3 ♀	1.4 ♀	1.3 ♀	1.4 ♀	1.3 ♀
<b>dårlig</b>	1.2 ♀	1.3 ♀	1.2 ♀	1.2 ♀	1.2 ♀
<b>stor</b>	1.5 ♂	1.3 ♂	1.3 ♂	1.4 ♂	1.5 ♂
<b>liten</b>	1.1 ♀	1.1 ♀	1.1 ♀	1.1 ♀	1.0 ♀
<b>pen</b>	1.5 ♀	1.4 ♀	1.4 ♀	1.5 ♀	1.4 ♀
<b>stygg</b>	1.0 ♂	1.0 ♂	1.0 ♂	1.1 ♂	1.1 ♂

Tabell 3: Hvem bruker flest av nærsynonymene i hvert av de fem samlede korpusene? Tallene viser forskjellen i snitt mellom kvinner og menn målt i ratio, røde tall med symbolet ♀ viser til kvinner, blå tall med symbolet ♂ viser til menn.

Tabell 4 under viser hvilke separasjonsverdier som korresponderer med data i tabell 3. Det er noen endringer sammenlignet med tabell 2, spesielt ser vi at distribusjonene for *stor* har fått høyere separasjonsverdier, slik at totalkorpus med delkorpus er likt med det kuraterte. Samtidig har separasjonene for *liten* minnet.

	Total	Kuratert	1	2	3	4	5
<b>god</b>	0,29	0,3	0,31	0,32	0,33	0,34	0,35
<b>dårlig</b>	0,23	0,24	0,25	0,26	0,27	0,28	0,29
<b>stor</b>	0,28	0,33	0,33	0,24	0,26	0,34	0,34
<b>liten</b>	0,07	0,17	0,08	0,1	0,08	0,1	0,06
<b>pen</b>	0,36	0,59	0,44	0,39	0,39	0,44	0,38
<b>stygg</b>	0	0,25	0,05	0,02	0,03	0,08	0,05

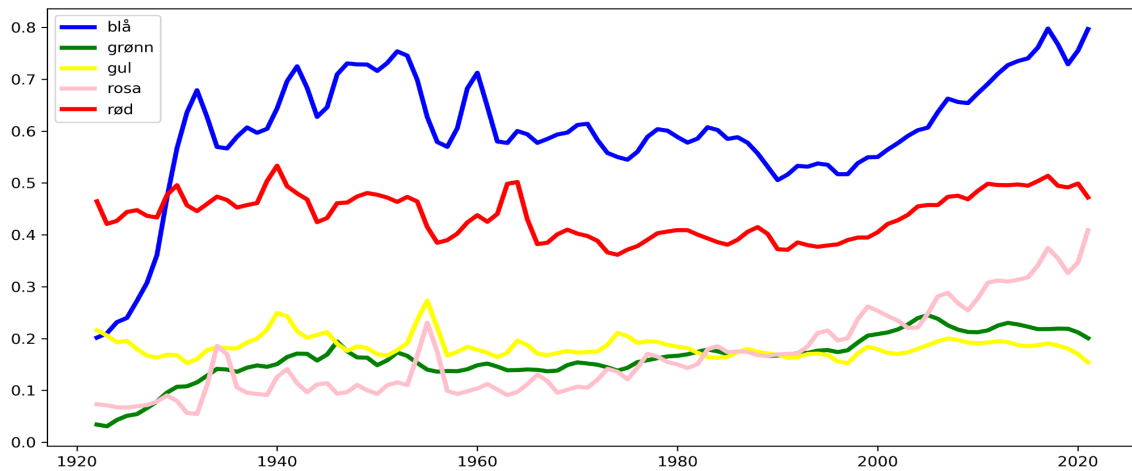
Tabell 4: En oversikt over separasjonsverdier i de forskjellige korpusene for nærsynonymene. Verdiene er utstyrt med en gradert fargestyrke.

Eksperiment 2 viser det samme som eksperiment 1, adjektivene har den samme kjønnsstilknytningen som i eksperiment 1. Det er marginale forskjeller i frekvens (mellom kjønnene) for adjektivene, bortsett fra for nærsynonymene til *stor* der forskjellen har økt en del. Forskjellene med og uten nærsynonymer viser seg i grad av separasjon, kvinner og menn nærmer seg hverandre for *liten*, men går fra hverandre for *stor*.

### 3.3 Eksperiment 3

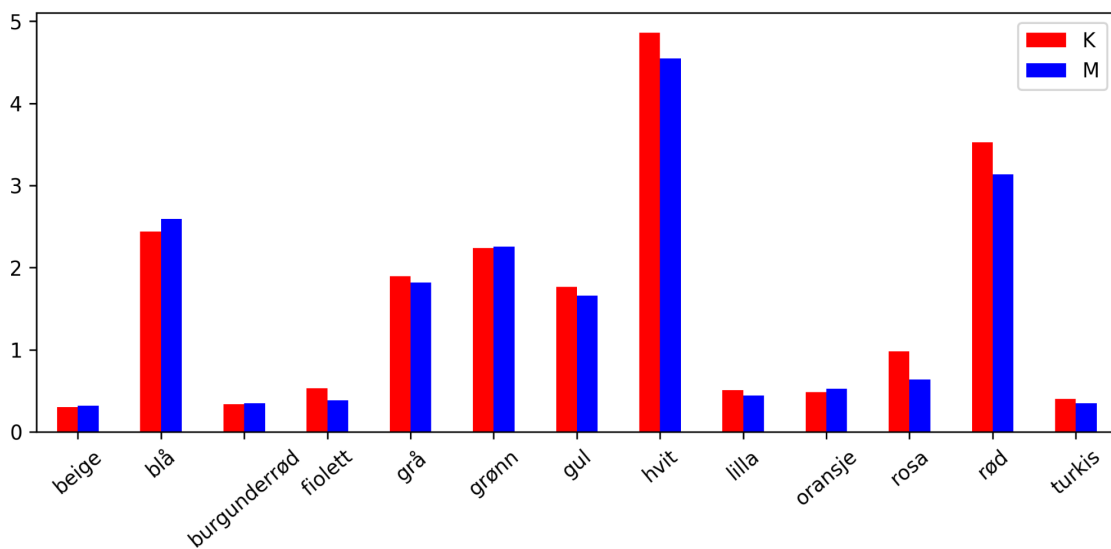
I det tredje eksperimentet teller vi fargeadjektivene for å se om kvinner bruker flere adjektiv enn menn, og hvilke farger som eventuelt foretrekkes av hvilket kjønn.

Bruken av sentrale fargenavn er nokså stabil i norske bøker (figur 6). Vi ser en rosa og gul topp på femtallet, og det er et lite rosa og blått oppsving på nittitallet.



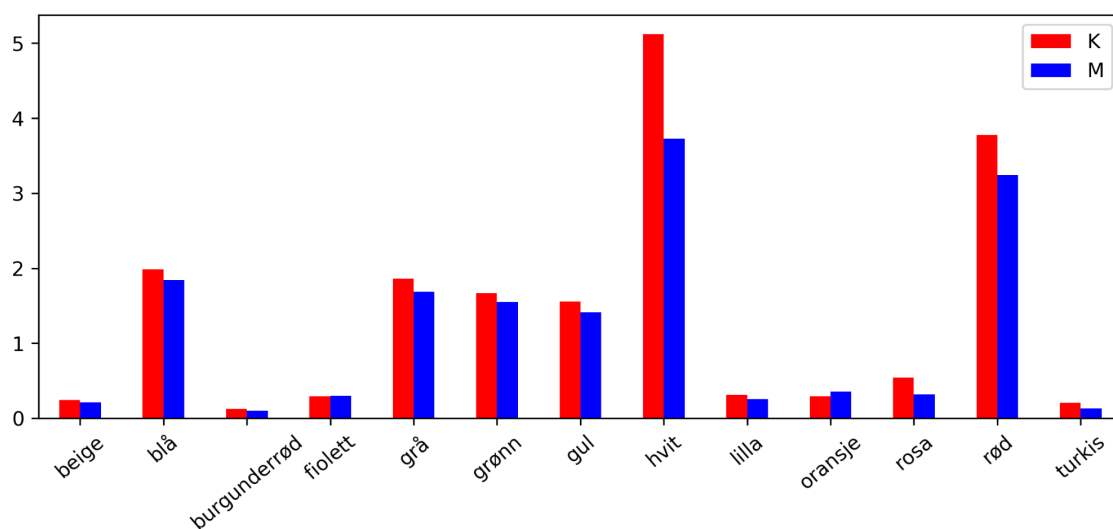
Figur 6: Trendlinjer fra NB-ngram for fem utvalgte farger, vist i antall pr. 10 000 ord.

Figur 7 og 8 under viser distribusjonen i totalkorpuset og i det kuraterte korpuset. Telling av alle våre valgte farger i totalkorpuset er vist i figur 7 og viser stor variasjon mellom fargeadjektivene, det vil si at forholdet mellom menns bruk og kvinners bruk varierer en del fra fargeadjektiv til fargeadjektiv. Figur 8 viser at adjektivet *oransje* har en liten mannlig overvekt i det kuraterte korpuset, og vi ser samme tendens i figur 7 i totalkorpuset. Mens det for B&A-adjektivene (med nærsynonymer) var høy overensstemmelse mellom totalkorpuset og det kuraterte korpuset, er det større variasjon for fargeadjektivene. Noe av variasjonen kan kanskje tilskrives at fargeadjektivene har betydelig lavere frekvens enn B&A-adjektivene med nærsynonymer. Likevel kan vi se et visst mønster.



Figur 7: Oversikt over fargebruken hos kvinner (K) og menn (M) i totalkorpuset. Tallene langs y-aksen er antall forekomster pr. 10 000 ord.

## OM KJØNN OG ADJEKTIV



Figur 8: Fargebruken hos kvinner (K) og menn (M) i det kuraterte korpuset, der y-aksen viser antall forekomster pr. 10 000 ord.

Vi kan fastslå at kvinner er mer fargeglade enn det menn er: Både i figur 7 og 8 ser vi at de fleste fargene er brukt mest av kvinner. I det kuraterte korpuset (figur 8) er 12 av 13 fargeadjektiv hyppigere brukt av de kvinnelige forfatterne; i totalkorpuset (figur 7) er de tilsvarende tallene 10 av 13. Analysen i tabell 5 under viser imidlertid at det er stor variasjon mellom de samlede korpusene med hensyn til hvilket kjønn som bruker de ulike fargeadjektivene mest. Trass i denne variasjonen er det kvinnene som bruker flest fargeadjektiv.

	1	2	3	4	5
<b>beige</b>	1.1 ♂	1.1 ♀	1.4 ♂	1.1 ♀	1.2 ♀
<b>blå</b>	1.0 ♀	1.1 ♂	1.0 ♀	1.2 ♀	1.0 ♀
<b>burgunderrød</b>	1.8 ♀	2.1 ♀	1.4 ♀	1.4 ♂	1.9 ♂
<b>fiolett</b>	1.1 ♀	1.2 ♀	1.0 ♂	1.3 ♂	1.2 ♀
<b>grå</b>	1.1 ♀	1.0 ♂	1.1 ♀	1.1 ♀	1.0 ♂
<b>grønn</b>	1.0 ♂	1.0 ♀	1.0 ♀	1.0 ♀	1.1 ♂
<b>gul</b>	1.1 ♀	1.1 ♀	1.2 ♀	1.2 ♀	1.1 ♀
<b>hvit</b>	1.0 ♂	1.0 ♂	1.1 ♀	1.0 ♀	1.0 ♀
<b>lilla</b>	1.5 ♀	1.1 ♂	1.0 ♂	1.1 ♂	1.1 ♀
<b>oransje</b>	1.5 ♀	1.1 ♀	1.1 ♀	1.1 ♀	1.7 ♀
<b>rosa</b>	1.4 ♀	2.0 ♀	1.4 ♀	1.3 ♀	1.5 ♀
<b>rød</b>	1.1 ♀	1.2 ♀	1.2 ♀	1.3 ♀	1.2 ♀
<b>turkis</b>	1.8 ♀	1.0 ♀	1.4 ♀	1.1 ♂	1.4 ♀

Tabell 5. Gruppevis fordeling av kjønn over farger. Tallene viser forskjellen i snitt mellom kvinner og menn målt i ratio. For eksempel viser tallet 2 i raden for rosa at det adjektivet forekommer over dobbelt så hyppig hos kvinner som hos menn. Røde tall med symbolet ♀ viser til kvinner, blå tall med symbolet ♂ viser til menn.

Tabell 6 under viser oversikten over separasjonsverdier for fargeadjektiv i alle korpusene. På samme måte som over benyttes fargekode for å angi grad av separasjon, sammen med en tallfesting.

	Total	Kuratert	1	2	3	4	5
beige	0,07	0,2	0,03	0,1	0,14	0,09	0,17
blå	0,03	0,11	0,03	0,05	0,01	0,09	0,01
burgunderrød	0,1	0,36	0,32	0,45	0,26	0,12	0,17
fiolett	0,01	0,03	0,05	0,11	0,02	0,12	0,09
grå	0,02	0,14	0,05	0,03	0,11	0,07	0
grønn	0,02	0,08	0	0	0,02	0	0,04
gul	0,04	0,13	0,05	0,04	0,07	0,09	0,04
hvit	0,03	0,4	0,01	0,01	0,06	0,03	0,02
lilla	0,06	0,18	0,21	0,06	0,01	0,05	0,05
oransje	0,09	0,14	0,11	0,12	0,06	0,08	0,29
rosa	0,14	0,56	0,18	0,16	0,17	0,17	0,24
rød	0,15	0,23	0,11	0,17	0,15	0,22	0,17
turkis	0,06	0,69	0,24	0,01	0,15	0,07	0,14

Tabell 6: En oversikt over separasjonsverdier i de forskjellige korpusene for fargeadjektiv. Kolonnene indikerer hvilke korpus verdiene kommer fra.

Den mest typiske kvinnefargen er rosa, som er hyppigst brukt i bøker skrevet av kvinner, og som i tillegg har høy grad av separasjon i alle korpusene. Fargen *rød* knytter seg også tydelig til kvinnelige forfattere, med høy frekvens og høy grad av separasjon.

La oss fra fargefunnene leke oss med en rangering av fargeadjektivbruk og kjønn. Det kan vi gjøre med utgangspunkt i resultatene vi ser i tabellene 5 og 6, samtidig som vi skjeler til tallene i figurene 7 og 8, med henholdsvis totalkorpuset og det kuraterte korpuset. Leken går sånn: Vi ordner fargeadjektivene etter hvor mange «stemmer» de får fra å ha høyest verdi i de forskjellige korpusene (ett korpus = én stemme), og der fargene scorer likt, ordnes de igjen etter en vurdering av frekvens og grad av separasjon. Rangeringen gir toppscore til kvinner for adjektivene *rosa*, *rød* og *gul*, rangert etter frekvens. Deretter ser vi på de adjektivene som har minst seks kvinnestemmer, altså der kun ett korpus avviker, der finner vi *turkis* og *oransje*, begge med varierende frekvens og separasjon. Blant de som avviker med to stemmer finner vi *hvit*, *grå* og *blå* alle for kvinner, der *hvit* tar plassen foran de andre med størst frekvensforskjell i det kuraterte. Til slutt, om vi fordeler adjektiv som har fire kvinnestemmer og tre mannsstemmer, finner vi *burgunderrød*, *fiolett*, *grønn* og *lilla*. Adjektivet *grønn* skiller seg ut med lavest separasjonsgrad, mens *burgunderrød* er interessant ved at det skifter mellom korpusene, og tillegg har høy grad av separasjon, selv når det skifter side. Så topplisten for feminint foretrukne farger er: *rosa*, *rød*, *gul*, *turkis* og *oransje*. På ventelisten setter vi *blå*, *grå*, *burgunderrød*, *hvit* og *turkis*. For menn kunne vi våge oss til å sette opp de to adjektivene *grønn* og *lilla*, som får tre stemmer hver for mest mannebruk, og har mest konsistent separasjon.

#### 4 Mer fargelegging og *Den gode, den onde og den grusomme*<sup>11</sup>

Søkene i eksperiment 1, med utgangspunkt i B&A-adjektivgruppene, viser at kvinner bruker hver adjektivgruppe mer enn det menn gjør. Unntaket er adjektivgruppen for *stor*. Om vi titter inn i hvert av de fem samlede korpusene, ser vi at kvinner bruker adjektiv mer frekvent enn det menn gjør – bortsett fra for adjektivgruppen til *stygg*, som i tre av de fem samlede korpusene brukes mest av menn, og som også har lave separasjonsverdier. For adjektivgruppen til *stygg* finner vi dermed ingen klar kobling til kjønn. I eksperiment 2, der bare nærsynonymene er telt, er kvinnelige forfatters tendens til å bruke flere adjektiv enn sine mannlige kolleger enda klarere enn i eksperiment 1. For tre av adjektivene – *god*, *liten*, *pen* – er tallenes tale ekstra tydelige. Unntakene fra kvinnedominansen er nærsynonymene til *stor*, som brukes mest av menn, og *stygg*, som ikke har noe dominant kjønn. For de resterende adjektivene er det kvinnene som bruker nærsynonymene mest.

Dersom vi later som om verden er inndelt i det som er godt og det som er dårlig, og lett kan beskrives enten med positive adjektiv eller negative adjektiv, ser vi at kvinner bruker positivt ladde adjektiv i enda større grad enn sine mannlige kolleger. Både grunnadjektivene *god* og *pen* og nærsynonymene til disse brukes mer frekvent av kvinner. At mennene bruker nærsynonymene til *stygg* mer enn kvinner gjør, peker i retning av at menn i større grad enn kvinner bruker de negativt ladde adjektivene. Men tallene er ikke krystallklare for positiv-negativ ladning: Kvinner bruker *dårlig* og og dennes nærsynonymer oftere enn det menn gjør.

Eksperiment 3 viser at de fleste fargene er brukt mest av kvinner. Likevel er det ikke mange adjektiv som kan sies å være *typiske* for kvinner. For selv om kvinner bruker flere av omtrent alle fargeadjektivene, vil variasjonen i gruppene være såpass stor at få adjektiv peker seg ut som først og fremst kvinnelige – rett og slett fordi mannlige forfattere også bruker dem ganske mye. Det er kanskje bare *rosa* og *rød* vi rimelig trygt kan hevde er typisk kvinnelige farger, mens *blå*, *oransje* og *gul* er de neste kvinnefargekandidatene. Og for menn: Kanskje vi kan driste oss til å hevde at *grønn* – og muligens *lilla* – er maskuline adjektiv. Våre hetero-normative sjeler stusser over at lilla ser ut til å være en mannefarge, men forskning skal jo nettopp flytte forestillinger og mulige fordommer.

For fargeadjektivene er det altså mer kvinnebruk enn mannebruk, men ikke mye mer og ikke for alle fargene. Resultatene kan dermed bare i noen grad sies å stemme med det Lakoff (1975) hevdet om fargeadjektiv. Lakoff brukte ikke korpus, men egne intuisjoner og inntrykk fra bekjente, som minner om hvordan Jespersen (1941), Abel (1947) og vårt 2021-eksemplar av *arten mannen i gata* gikk til verks.

Oppsummert finner vi at hovedspørsmålet vi har stilt, om kvinner bruker adjektiv mer enn menn i vårt datamateriale, kan besvares med «ja». Når det gjelder spørsmålet om kvinner bruker andre adjektiv enn menn, om de varierer mer, er svaret også «ja», kanskje i retning av «tja», idet kvinner bruker nærsynonymer oftere enn mennene, selv om noen få adjektiv brukes oftere av menn.

Resultatene fra vår undersøkelse viser på mange måter det Barczewska og Andreasen (2018) konkluderer med for sin del – at ulike tilnærminger og metoder vil gi ulike resultater. Vi kan legge til at også ulike korpustyper og adjektivutvalg sannsynligvis også har mye å si for resultatet (jf. også Baker 2014). Vi har brukt korpus med skriftlig tekst, et skjønnlitterært basert korpus. Fordelen med store, skjønnlitterære korpus er at man kan forvente et vell av ulike situasjoner, et hav av mellommenneskelige interaksjonstyper og relasjoner. Ulempen er at forskjeller innad i en gruppe kan utlignes; noen deler av korpus har høye verdier, mens andre deler har lave, og snittet som presenteres, er kanskje litt villedende. Med vår samplingsmetode har vi søkt å gjøre noe med denne ulempen. Samplingen av korpuset viser at det kan være noen individuelle forskjeller hos de ulike forfatterne når det gjelder bruken av adjektiv, men også at det store korpuset er relativt homogent med hensyn til kvinners og menns adjektivbruk. Svendsen (2019), omtalt tidligere, er inne på denne problematikken når hun fremholder: «While all studies discussed in this article take an empirical approach to examine Lakoff's claims about women's language, they still share a general methodological problem: all investigate the differences between the speech of

<sup>11</sup> Italiensk westernfilm fra 1966 – som jubilanten bør se hvis han ikke allerede har gjort det – regissert av Sergio Leone, originaltittel *Il buono, il brutto, il cattivo*.



men and women assuming that men and women form homogenous groups.» Hun mener at det er vanskelig å isolere kjønnseffekt i og med at faktorer som «ethnicity, age, educational level, socio-economic status, sexuality, etc. vary within the groups and may indeed affect the speech style of the individual» (Svendsen, 2019, s. 8). Også vi forholder oss til en todeling av menneskeheten i enten kvinner eller menn, men vi går altså i vår undersøkelse med samplingen også inn i delkorpus og finner da relativt stor grad av homogenitet, med noen punkter med intern heterogenitet.

Samtidig er vårt korpus delt inn etter forfatterens kjønn, ikke kjønn til den som har synsvinkelen i romanen. Hvis en mannlig forfatter ønsker å skrive ut fra en kvinnelig synsvinkel, hva da med adjektivbruken? Det blir kanskje en annen skål. I mellomtiden skåler vi for jubilanten i alkoholfri vin: Skål! Og hvis vi skulle valgt tre adjektiv som beskrev ham, ville vi verken ha valgt de typiske manneadjektivene *lilla*, *stygge* eller *stor*, men kanskje snarere *rød*? Og dessuten *dyktig* og *sympatisk*.

## 5 Referanser

- Abel, Georges. 1947. *Manns- og kvinnespråk: en populær fremstilling*, Oslo.  
<https://www.nb.no/nbsok/nb/4a41b0e618c97415d4c32872da0a8670?lang=no#0>
- Amir, Zaini, Hazirah Abidin, Saadiyah Darus, and Kemboja Ismail. 2012. Gender differences in the language use of Malaysian teen bloggers. *GEMA Online™ Journal of Language Studies* 12(1): 105–124.
- Baker, Paul. 2014. *Using corpora to analyze gender*. Bloomsbury, London, New York.
- Bamman, David, Jacob Eisenstein, and Tyler Schnoebelen. 2014. Gender identity and lexical variation in social media. *Journal of Sociolinguistics* 18(2): 135–160. <https://doi.org/10.1111/josl.12080>
- Barczewska, Shala og Andreassen, Agata. 2018. “Good or marvelous? Pretty, cute or lovely? Male and female adjective use in MICAS”. *Suvremena Lingvistika* 44(86):194-213.  
<https://doi.org/10.22210/suvlin.2018.086.02>.
- Blatt, Ben. 2017. *Nabokov’s Favorite Word is Mauve*. New York: Simon & Schuster.
- Bull, Tove. 2021. Kjønnen språk og språkbruk før og no. *Målbryting* nr.12(2021): 1–24.  
<https://doi.org/10.7557/17.5787>.
- Cameron, Deborah. 2008. *The Myth of Mars and Venus*. Oxford University Press.
- Dyvik, Helge. 2018. <https://nogramtall.w.uib.no/om-denne-bloggen/>
- Fishman, Pamela M. 1977. ‘Interactional Shitwork’. *Heresies* 2: 99–101.
- Hartman, Maryann. 1976. A descriptive study of the language of men and women born in Maine around 1900 as it reflects the Lakoff hypotheses in “Language and Women’s Place”. Paper presented at the Conference on the Sociology of the Languages of American Women. New Mexico State University, Las Cruces, New Mexico, January 16–17. Available online at  
<https://files.eric.ed.gov/fulltext/ED130316.pdf>.
- Jespersen, Otto. 1941. *Sproget. Barnet, kvinden, slægten*. Gyldendalske boghandel.
- Johnsen, Lars G. Bagoien. 2022. Datasett <https://doi.org/10.5281/zenodo.6425086>
- Kramer, Cheri. 1974. Women’s speech. Separate but unequal? *Quarterly Journal of Speech* 60(1): 14–24. <https://doi.org/10.1080/00335637409383203>
- Lakoff, Robin T. 1975. *Language and woman’s place*. New York: Harper & Row.
- Schmid, Hans-Jörg 2003. Do women and men really live in different cultures? Evidence from the BNC. In *Corpus Linguistics by the Lune: A Festschrift for Geoffrey Leech* edited by Andrew Wilson, Paul Rayson, and Tony McEnery, 185–221. Peter Lang.
- Svendsen, Amalie Due. 2019. Lakoff and Women’s Language: A Critical Overview of the Empirical Evidence for Lakoff’s Thesis. *Leviathan: Interdisciplinary Journal in English*, 4, 1-11.  
<https://doi.org/10.7146/lev.v0i4.112651>
- Tannen, Deborah. 1990. *You just don’t understand: Women and men in conversation*. New York: Morrow.
- Trosterud, Trond. 2001. Genustilordning i norsk er regelstyrt. *Norsk lingvistisk tidsskrift*, 19: 29–58.
- Uri, Helene. 2018. *Hvem sa hva? Kvinner, menn og språk*. Oslo: Gyldendal.