

Janne Bondi Johannessen &
Kristin Hagen (red.)

Leksikografi og korpus.
En hyllest til Ruth Vatvedt Fjeld

Oslo Studies in Language

General editors: Atle Grønn and Dag Haug

Janne Bondi Johannessen &
Kristin Hagen (red.)

Leksikografi og korpus.
En hyllest til Ruth Vatvedt Fjeld

Oslo Studies in Language, 11(1), 2020.

Janne Bondi Johannessen & Kristin Hagen (eds.):
Leksikografi og korpus. En hyllest til Ruth Vatvedt Fjeld.

Oslo, University of Oslo

ISSN 1890-9639

© 2020 the authors

Set in Word, fonts Gentium Book Basic and Linux Libertine by Kristin Hagen.

Cover design by UniPub publishing house & Atle Grønn.

Printed by Print House AS from camera-ready copy supplied by the editors.

<http://www.journals.uio.no/osla>

Innhold

Om denne boka <i>Kristin Hagen</i>	1
Om Ruth <i>Boye Wangensteen</i>	3
En mer inkluderende leksikografi – bortom frågan om kvinner og män <i>Emma Sköldberg</i>	7
Skævheder og ligestilling: i ordbogen, i korpus og i samfundet <i>Lars Trap-Jensen</i>	31
Frå ordbok til korpus – og tilbake? <i>Bjørnetjeneste og forfordele</i> <i>Lars S. Vikør</i>	47
Det umulige er mulig <i>Urd Vindenes og Hans-Olav Enger</i>	61
<i>Underligt nog/klokt nog</i> – en ganske ny typ av satsadverb i svenska och norska <i>Sven-Göran Malmgren</i>	77
Store tekstkorpus til bruk i Språkrådets språkobservasjon, språknormering og språkrådgivning <i>Åse Wetås</i>	87
Leksikografisk bokmålskorpus (LBK) – bakgrunn og bruk <i>Ruth Vatvedt Fjeld, Anders Nøklestad og Kristin Hagen</i>	101

OM DENNE BOKA

I mai 2018 fylte Ruth Vatvedt Fjeld 70 år og gikk av med pensjon etter et langt arbeidsliv ved Universitetet i Oslo. For å feire Ruth arrangerte vi et symposium 1. juni 2018 med foredragsholdere fra Norge, Danmark og Sverige. Fellesnevneren skulle være Leksikografisk bokmålskorpus, et tekstkorpus på 100 millioner ord som Ruth har arbeidet med siden slutten av 90-tallet. Arrangementskomiteen bestod av fire personer som alle kjenner Ruth godt: professor Janne Bondi Johannessen fra Tekstlaboratoriet, Rikke Hauge fra Språkrådet og Nordisk forening for leksikografi, Mina Finstad Berg, som Ruth akkurat hadde veiledet masteroppgaven til, og undertegnede Kristin Hagen fra Tekstlaboratoriet, som også hadde Ruth som veileder på hovedoppgaven sin i 1991.

Etter det vellykkede symposiet var det flere foredragsholdere som ønsket at foredragene skulle bli til en publikasjon og dermed et mer håndfast minne fra dagen. Denne boka inneholder derfor artikler av Emma Sköldberg, Lars Trap-Jensen, Lars Vikør, Urd Vindenes og Hans-Olav Enger, og til slutt Åse Wetås. Innledningen er skrevet av Ruths mangeårige kollega, Boye Wangenstein.

Ikke alle foredragsholderne hadde anledning til å bidra til boka, derfor inviterte vi også Sven-Göran Malmgren, som ikke kunne delta på symposiet, samt Ruth selv. Slik har Leksikografisk bokmålskorpus fått en egen artikkel i boka der Ruth beskriver historien bak og søkegrensesnittet til korpuset sammen med Anders Nøklestad og undertegnede.

Artiklene i boka er fagfellevurdert av kollegaer i Norden. En stor takk til dem.

Vi har tatt oss god tid i arbeidet med denne boka, og jeg er glad for at den har blitt ferdig til tross for at medredaktør Janne Bondi Johannessen var mye syk i hele perioden og gikk bort før boka ble helt ferdig.

Kristin Hagen, Oslo august 2020

KONTAKT

Kristin Hagen

Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo

kristin.hagen@iln.uio.no

OM RUTH

BOYE WANGENSTEEN

Tiden går, og i mai i år var det 35 år siden Ruth kom feiende som et friskt vær inn i norsk leksikografi da hun ble tilsatt som vitenskapelig assistent på Avdeling for bokmål ved det som da het Norsk leksikografisk institutt. Instituttet holdt til på Chateau Neuf i Slemdalsveien inntil det ble slått sammen med Nordisk institutt i 1990 og flyttet til Blindern i 1992.

Før Ruth gikk i gang med studiene, hadde hun jobbet i det private næringsliv, og innimellom studiene underviste hun i forskjellige skoleslag. Det ble tidlig klart at Ruth hadde faglige ambisjoner. I 1988 ble hun doktorgradsstipendiat og det ble forskerutdanning i språkvitenskap. I årene 1991-1997 fungerte hun som amanuensis og fra 1997 som førsteamanuensis ved instituttet, som skiftet navn i takt med nye omorganiseringer. Etter hvert hadde hun ferdig en stor doktoravhandling med tittelen *Rimelig ut fra sakens art. Om tolkning av ubestemte adjektiv i regelgivende språk*, som hun disputerte på i 1998. Hun ble dr.philos. og professor i nordisk språkvitenskap samme år. I 2013 ble hun valgt som prorektor ved Universitetet i Oslo.

Ruth har vært veileder for mange masteroppgaver, og hun var den første som arrangerte metodekurs for masterstudentene ved ILN. Hun har vært opponent ved flere doktordisputaser i inn- og utland, og hun har vært gjesteforeleser bl.a. i Litauen, USA, Russland og Etiopia. Hun har også vært mye benyttet som bedømmer av søknader for flere nordiske institusjoner, f.eks. Svenska Akademien og Velux-fonden i Danmark.

Ruth har alltid lagt stor vekt på nordisk og internasjonalt samarbeid. Det var hun som tok initiativet til og fikk gjennomført den første store nordiske konferansen i leksikografi, som ble arrangert i Oslo allerede i 1991. Det var på denne konferansen Nordisk forening for leksikografi (NFL) ble stiftet, og foreningen har siden arrangert nordiske konferanser i leksikografi fast hvert andre år. Hva dette samarbeidet som NFL initierte, har betydd for faget leksikografi i Norden, kan knapt overvurderes.

Ruth var med i styret i NFL i mange år og var leder fra 2003 til 2005. Hun startet og redigerte internbladet *NFL-nytt* i årene 1992-2002. Etter at hun ga seg, har bladet avgått en stille død. Hun var en av de seks nordiske redaktørene av

Nordisk leksikografisk ordbok, med hovedansvar for de semantiske termene innenfor leksikografifaget. Boka kom ut i 1997. I en årrekke var hun norsk redaksjonsmedlem i *LexicoNordica*, tidsskriftet som Nordisk forening for leksikografi har gitt ut siden 1994, og i årene 2010-2011 var hun den ene av de to hovedredaktørene. I årene 2006-2012 var hun styremedlem i den europeiske foreningen for leksikografi, EURALEX.

Det er blitt mange artikler fra Ruths hånd opp gjennom årene, i *LexicoNordica*, i rapportene fra de nordiske leksikografikonferansene og i andre språkvitenskapelige tidsskrifter. Hun redigerte også den første konferanserapporten fra konferansen i 1991 og kalte den *Nordiske studier i leksikografi*. Serien er nå kommet til nr. 14.

Ruth har lagt ned mye arbeid på å utvikle faget leksikografi, bl.a. ved at hun i mange år, og så lenge tilbudet eksisterte, hadde ansvaret for undervisningen i leksikografi ved det instituttet faget til enhver tid hørte til. Sammen med Lars Vikør har hun skrevet *Ord og ordbøker* på Høyskoleforlaget i 2008, en omfattende innføring i leksikologi og leksikografi, og den første på norsk.

Ruth har vært hovedarkitekten bak det store prosjektet *Leksikografisk bokmålskorpus* (LBK), som ble ferdig i 2013. Korpuset består av over 100 millioner ord, og det er et fantastisk hjelpemiddel både for ordboksarbeid og annen språkforskning.

I 2018 kom *Norsk banneordbok* på Humanistisk forlag, et nybrottsarbeid som gir en omfattende dokumentasjon og analyse av banneord og kraftuttrykk fra hele landet.

Feltet språkrøkt bør også nevnes. Det begynte med mange kurs i praktisk norsk rundt omkring alt i studietiden, og den første boka Ruth ga ut, *Norsk til kontorbruk* på Universitetsforlaget, kom for over 30 år siden. Den kommer stadig i reviderte opplag, den siste i 2017 med tittelen *Skriv godt og riktig på jobben*. Her kan også nevnes *Lingua Bokmålsordliste med skriveregler* på Kunnskapsforlaget i 2008.

I årene 2000-2012 var Ruth språkkonsulent i NRK, der hun fulgte etter størrelser som Finn-Erik Vinje og Dag Gundersen. Hun har også vært benyttet som ekspertvitne i mange rettssaker om ord og varemerker.

Ruth har aldri vært redd for å tone flagg og ta en frisk diskusjon. Hun har alltid hatt mot til å stå fram og målbære det hun mener er rett og aldri gått på akkord med sine prinsipper. (Selv om hun er leksikograf, er det mye som tyder på at hun ikke kjenner ordet og begrepet opportuniste.) Det hadde vært underlig om slike egenskaper ikke hadde ført til konflikter av og til, særlig når de er representert av en kvinne. Det har nok Ruth opplevd fra tid til annen, men med sitt mot og sin fightervilje har hun alltid stått fjellstøtt på sitt.

Ruth har stor arbeidskapasitet og en imponerende energi, og hun har vært engasjert på flere samfunnsområder også utenom det rent faglige. Hun var leder i Bymålslaget i ei årrekke (1997-2013), hun har vært nestleder i Oslo-avdelingen av Foreningen Norden i to perioder og i regi av Human-Etisk Forbund har hun vært fast taler både på navnefester og ved begravelser. For Ruth er ingen oppgave for stor og ingen for liten, enten det dreier seg om å ha hovedansvaret for store internasjonale konferanser (bl.a. Euralex-konferanse på Blindern i 2012) eller å koke risengrynsgrøt til juleavslutningen på avdelingen. Og enten arrangementet er stort eller lite, så liker hun å ha regien. Og da kan vi andre ta det rolig, vi vet at den er i de beste hender.

Jeg husker Ruth også som en miljøbygger. Mens vi holdt til nede på Chateau Neuf, inviterte hun i flere år ikke bare oss på Avdeling for bokmål, men hele språkmiljøet som holdt til der, Norsk Ordbok, Avdeling for gammelnorsk og mellomnorsk og Institutt for navnegransking, på hagefester hjemme hos seg i Mylskerudveien, fester som jeg minnes med stor glede.

Ruth har gjennom mange år vært en god og inspirerende kollega, og jeg drister meg til å si også en god venn. Hun fortjener både et jubileumssymposium og en hyllest i bokform!

Boye Wangenstein
mangeårig kollega

KONTAKT

Boye Wangenstein
Pensjonert leksikograf, Institutt for lingvistiske og nordiske studier,
Universitetet Oslo
boye.wangenstein@hotmail.no

EN MER INKLUDERANDE LEXIKOGRAFI – BORTOM FRÅGAN OM KVINNOR OCH MÄN

EMMA SKÖLDBERG

SAMMANFATTNING

I artikeln behandlas ordböcker ur ett genusperspektiv. I första hand diskuteras, utifrån befintlig genuskritisk forskning, innehållet i det enspråkiga verket *Svensk ordbok utgiven av Svenska Akademien* (SO, 2009). Vidare presenteras resultaten av en undersökning som gäller användningen av substantiven *tjej* och *kille* i SO:s språkprov. Dessa resultat utgör underlag för resonemang kring några av de förändringar som är aktuella inför nästa upplaga av ordboken, som är under utarbetande.

[1] INLEDNING

Studier av ordböcker ur ett genuskritiskt perspektiv visar, som Fjeld (2001:12–13) har konstaterat, att ”kvinnene og deres språkbruk ofte blir dårlig behandlet, selv om ikke alle ordbøker er direkte sexistiske”. Kritik har riktats mot bl.a. svenska verk, exempelvis *Svensk ordbok utgiven av Svenska Akademien* (SO, 2009). En av de frågeställningar som kritiska läsningar av ordböcker väcker är om ordböcker såsom SO upprätthåller, eller kanske t.o.m. befäster, stereotypa skillnader mellan könen.

I det följande redogör jag kortfattat för det lexikografiska forskningsläget ur ett genusperspektiv. Därefter följer en kort diskussion om olika typer av ordböcker och vilken eller vilka roller de förväntas spela i förhållandet till språket och det omgivande samhället. Vidare diskuterar jag, utifrån befintlig genuskritisk forskning, innehållet i SO (2009), som tjänar som utgångspunkt för denna artikel. Därpå följer en mindre empirisk studie baserad på innehållet i SO (2009). Studien, som gäller användningen av substantiven *tjej* och *kille* i språkprov i ordboken, utgör underlag för vidare resonemang kring några av de förändringar som är aktuella inför nästa upplaga av verket, som är under utarbetande.

De frågor som tas upp i artikeln är näppeligen enkla att besvara och jag vill förutskicka att artikeln inte innehåller något uttömmande facit. I den meningen innehåller följande text således fler frågor än svar. De problematiserande resonemang som förs torde dock visa på den komplexitet som många av dessa frågeställningar har.

Innan redogörelsen för forskningsläget följer nu en kortfattad presentation av den aktuella ordboken. SO kom ut i tryckt format år 2009 och i form av appar 2015. SO är en vidareutveckling av *Svensk ordbok* (SOB 1986) och *Nationalencyklopedins ordbok* (NEO 1995–96), och den bygger på en databas som utarbetats och vidareutvecklas vid Institutionen för svenska språket, Göteborgs universitet. Huvudredaktör för ordboken var professor Sven-Göran Malmgren. SO är en definitionsordbok som avser, bl.a. genom betydelsebeskrivningar och språkexempel, att i första hand beskriva svenskt samtidsspråk. Vidare är SO en korpusbaserad ordbok. De korpusar som användes under arbetet inför 2009 års upplaga (och dess föregångare) innehöll huvudsakligen svenska dagstidningar och romaner, dvs. etablerat, redigerat skriftspråk (se Inledning, 2015:XII; se vidare nedan). Det är värt att framhålla att de språkprov som ges oftast är redigerade. Korpusarna utgör ett viktigt stöd när det gäller att finna illustrativa exempel men många autentiska språkanvändningar har modifierats innan de lagts in i ordboksdata-basen (se vidare bl.a. Hult 2010 för en diskussion kring autentiska vs. redigerade språkprov och tankarna bakom SO:s exempelgivning).

Tillsammans med Akademiens andra två ordböcker, rättskrivningsordboken SAOL och den historiska ordboken SAOB, är SO tillgänglig via webbplatsen svenska.se. SO har genom ordboksportalen fått större spridning. Avståndet mellan redaktion och användare har tack vare portalen i förening med ordboksapparna minskat. Användare kan nu på ett enklare sätt än tidigare förmedla sina synpunkter på ordböckernas innehåll till lexikograferna.

[2] GENUSKRITISKA STUDIER AV ORDBÖCKER

Enligt Hageberg (1990:49–50) ökade intresset i slutet av 1960-talet för att närma sig språket på nya sätt. I kölvattnet av kvinnorörelsen på 1970-talet publicerades en rad forskningsartiklar m.m. om språk och kön och språk och makt (se även Epple 2000; Edlund et al. 2007: 47–54; Ishikawa 2013:222). Forskningsinriktningar som psykolingvistik och sociolingvistik etablerades därmed som egna kunskapsfält. Anmärkningsvärt nog kom emellertid inte innehållet i ordböcker – under denna tid eller under de kommande decennierna – att granskas eller ifrågasättas i någon nämnvärd omfattning.

Under de senaste decennierna, inte minst på 1990-talet, har det dock forskats kring nordiska ordböcker ur ett könsperspektiv (se exempelvis Hageberg 1990, Nikula 1997, Kram 1998, Fjeld 2015). Forsknings slutsatserna om nordiska ordböcker går i linje med iakttagelser i internationella lexikografiska studier av bl.a. Whitcut (1983), Epple (2000), Landau (2001), Ishikawa (2013), Moon (2014) och Arimbi & Kwary (2016).

Hageberg (1990) problematiserar det sätt som kvinnor presenteras i norska

ordböcker. Hon konstaterar bl.a. att kvinnor överlag är relativt frånvarande i verken och att närvarande kvinnor framställs de på ett onyanserat sätt. Hon ifrågasätter även varför antalet nedsättande benämningar på kvinnor är klart högre än nedsättande benämningar på män i lexikografiska verk.

Också Fjeld (2001) studerar norska ordböcker ur ett genusperspektiv. Hon konstaterar bl.a. att det ofta finns stora luckor i ordböckernas lemmauppsättningar vad gäller uppslagsord kopplade till bl.a. smink och kläder, kvinnliga kroppsfunktioner och traditionella kvinnoyrsslor. Fjelds forskningsgenomgång visar också att kvinnor i ordböcker ofta presenteras som om de vore avhängiga av män. De beskrivs generellt stereotypiskt exempelvis genom sina yrken (såsom sekreterare) eller genom de yrsslor de ägnar sig åt (såsom dammsugning) (2001:11–13).

Nikulas studie (1997) visar i sin tur på en rad skevheter i språkprov som ingår i svenska SOB (1986). Hon konstaterar att sexism inte nödvändigtvis uppträder öppet i den meningen att kvinnor negligeras i exempelmeningarna eller att de omtalas som bihang till män. Det förefaller däremot finnas gott om exempel på dold och mera försåtlig diskriminering eller smygsexism. Kännetecknande för denna typ är, enligt Nikula, att kvinnan beskrivs som underordnad, passiv, som objekt, och bunden till de gamla könsrollsmönstren (1997:196). Vidare konstateras att kvinnor också ofta uppträder i exempel som är negerade och att de därför framstår som oförmögna i något avseende (t.ex. ”en kvinna i karriären har inte tid med barn”). Denna typ av sexism är långt mindre iögonenfallande än när kvinnor omtalas med nedsättande beteckningar som *ragata*, *satkärning* och *skvalermoster* (se även Kram 1998; Edlund et al. 2007:179–205, 192–194).

Nikula (2008) har även studerat svenska ordboksillustrationer ur ett könsperspektiv och kommer då fram till att illustrationer främst gäller ord som hör hemma i en manlig värld. Enligt författaren är det en konsekvens av en allvarlig och grundläggande obalans vid valet av lemman.

De problem som lyfts fram ovan gäller främst enspråkiga ordböcker men också tvåspråkiga verk uppvisar brister ur genusperspektiv. Bl.a. har Pilke (2009) synliggjort den skeva bild av kvinnor och män i arbetslivet som återges i en större finsk-svensk ordbok från 1990-talet.

Genuskritiska studier av ordböcker kan och bör sättas in i ett större sammanhang. Studierna av manligt och kvinnligt i ordböcker och ordboksexempel har t.ex. tydliga beröringspunkter med undersökningar av innehållet i exempel som ges i grammatikböcker. Bland annat har Ohlander (1995) studerat exempel som återfinns i engelska grammatikor under en tidsperiod av 400 år. Lewandowski (2014) har å sin sida studerat innehållet i exempelmeningar i liknande böcker sedan 1970-talet och framåt. Enligt Ohlander (1995) speglar flertalet exempel i

grammatikböckerna religiösa och politiska idéer under olika perioder. Exempelvis anspelar många äldre exempel på välkända teman i Bibeln. I yngre böcker kan exemplen istället gälla miljöfrågor, dvs. teman som, åtminstone 1995, kunde betraktas som okontroversiella. I äldre exempel beskrivs också män ofta som modiga och intelligenta medan kvinnor beskrivs som dygdiga och blygsamma. Det tycks emellertid som om vissa av dagens grammatiker strävat efter att bägge könen rent kvantitativt ska vara lika välrepresenterade i sina verk (Ohlander 1995:230–234; jfr Fjeld 2015). Under de senaste decennierna har det, enligt Lewandowski (2014), också skett en förändring vad gäller exemplens innehåll, bl.a. när det kommer till yrkesroller. Ohlander (1995) konstaterar dock avslutningsvis att den bild han målar upp kanske är väl konstruerad, inte minst ur ett feministiskt perspektiv och att grammatikernas exempel inte alltid är representativa med tanke på verkligheten: "Instead, they occasionally seem to indulge in some kind of virtual (...) reality, often amounting to wishful thinking, promoting what is, at a certain time, widely regarded as a good and political correct cause" (1995:234). Exempelen i grammatikböcker kan alltså inte alltid tas som intäkt för hur det ser ut i verkligheten.

Enligt Hageberg (1990:48) speglar återgivningen i bl.a. ordböcker det faktum att män genom historien haft mer makt. Det har lett till att ett manligt perspektiv varit förhärskande i språkvetenskapliga avhandlingar, läroböcker m.m. Vidare har ordboksredaktionerna genom tiderna dominerats av manliga medarbetare (se vidare Hageberg 1990:50–51 för en historisk genomgång av könsfördelningen bland lexikografer vid olika ordboksredaktioner i Norden). Många av de informanter som på olika sätt bidragit med språkligt material till ordboksredaktionerna har också varit män. Det har även varit en övervikt av manliga författare till de verk som ordboksredaktörerna excerperat ur (Hageberg 1990:51–54). Detta senare påstående stöds också av Mattisson (2006) som anlagt ett genusperspektiv på SAOB:s källor.

Hageberg påpekar dock att det visat sig att fler kvinnor på ordboksredaktionerna inte per automatik leder till högre grad av kvinnlig närvaro och jämlikhet i ordböcker (jfr Fjeld 2015:53). Hageberg (1990:59–60) konstaterar att ordboksredaktörer har större makt och möjlighet att styra innehållet i ordböckerna än de kanske är medvetna om. Ordboksförfattare (såväl kvinnliga som manliga) har ett stort ansvar när de definierar, väljer ut autentiska språkexempel och inte minst konstruerar egna språkprov.

Flera av de studier som nämns ovan är 20 år eller äldre och de gäller ofta äldre ordböcker och man hoppas förstås att det ser annorlunda ut i ordböcker av idag. Enligt Epple (2001), vars studie utgår från tvåspråkiga ordböcker, går det att se markanta skillnader i graden av medvetenhet om dessa frågor i äldre och mer

nyutkomna verk (se även Landau 2001). Epples resultat kan dock inte återopas som stöd för att allt är bra, fastän det har blivit bättre. En relativt nylig studie som ger besked om nutida verk är redan nämnda Fjeld (2015), som i en djupgående undersökning jämfört *Bokmålsordboka*, SO (2009) och *Den Danske Ordbog* (DDO). För det första har hon räknat antalet exempelmeningar med pronomen som *hon/han*, *henne/honom* och kommer då fram till att samtliga ordböcker har fler exempel med *han*. För det andra har hon granskat i vilken grad *hon-* och *han-*språkprov uttrycker positiva eller negativa konnotationer, aktiva eller passiva subjekt och explicita könsstereotypa föreställningar. Undersökningen visar i korthet att det råder skillnader mellan verken, men att alla tre ordböcker skulle må bra av att revideras ur jämlikhetssynpunkt (se vidare nedan).

[3] ORDBOKENS ROLL – ATT FÖRÄNDRA ELLER SPEGLA?

En mycket viktig aspekt i detta sammanhang är vad ordboksförfattarna egentligen har för uppgift och vad ordboken ska åstadkomma. Å ena sidan kan man i korthet mena att förändringar av språket kan leda till förändringar av attityder i det omgivande samhället. Språkvårdande resurser som t.ex. ordböcker kan här spela en viktig roll. Exempelvis frågar sig Kram (1998:170) om det inte är en ordboksförfattares uppgift att i sina ordval stödja demokratiska processer. Hon ifrågasätter var gränsen går mellan att föra en tradition vidare och att problematisera sådana attityder som inte längre är önskvärda i det samhälle där ordboken ska fungera (se vidare bl.a. Wojahn 2015 om språkaktivism; se även t.ex. Fjeld 2015:54; Josephson 2018:205–206).

Ordboksförfattarna kan å andra sidan ta fasta på den *typ* av ordbok som de har fått i uppdrag att utarbeta eftersom detta har stora konsekvenser för ordbokens innehåll. Syftet med ordboken kan vara deskriptivt, dvs. att visa på hur språkbruket se ut. I enlighet med Moon (2014) finns det, utöver rent biologiska skillnader mellan kvinnor och män, skillnader i kvinnors och mäns livsföring. Detta reflekteras i ordförrådet, i attityder som kommuniceras via språket – och i ordböckernas innehåll. Ett exempel på en i första hand deskriptiv ordbok är således SO (2009). Men syftet med en ordbok kan också vara normativt, dvs. att styra språkbruket (Svensén 2009:24). SAOL är ett exempel på en sådan ordbok. Enligt Hageberg går det t.ex. inte att låtsas som att nedsättande beteckningar inte existerar, men hon är osäker på hur de bäst ska behandlas i ordböcker. Som möjliga alternativa lösningar anförs bl.a. utelämnning av dylika ord i vissa verk, nyttjande av brukskommentarer och bruk av upplysningar till exempelvis inlärare, så att de kan undvika att använda den aktuella typen av ord (se även Whitcut 1983:141). Det råder dock flytande gränser mellan deskriptiva och normativa

ordböcker och, som Svensén (2009:24) påpekar, uppfattas även deskriptiva ordböcker som normativa (se vidare avsnitt 7).

[4] RÖSTER OM SO (2009):S INNEHÅLL

I samband med uppdateringen av den lexikaliska databasen inför SO (2009) reviderades ordboksinnehållet ur könsperspektiv. Detta framgår exempelvis av Malmgren (2009:19) som påpekar att antalet *hon* i SO:s språkprov numera är ungefär lika stort som antalet *han* (jfr Fjeld 2015 ovan). Ett stort arbete som genomfördes inför 2009 var också att lägga till språkprov till de flera tusen idiom som ingår i ordboken. Även i samband med detta arbete strävade SO-redaktionen efter att uppnå ett jämställt innehåll och att idiomexemplen inte skulle visa upp en stereotyp och förlegad syn på könsroller (Hult 2012:216).

Det arbete som genomfördes uppmärksammas också av Nikula som i sin recension av SO (2009) menar att redaktionen nått en bit på väg. Hon skriver:

Ur genussynvinkel kan konstateras att en del av den gamla surdegen rensats bort bland exemplen. Både i SOB och i NEO fanns det gott om exempel där kvinnan sågs som beroende av mannen om hon över huvud tilläts visa sig (...) men numera är kvinna och man mer jämspelta (...). (Nikula 2010:364)

Denna förbättring påverkar, enligt Nikula, i och för sig inte ordbokens användbarhet i någon större utsträckning, men nu bidrar inte ordboken på samma sätt till spridning av stereotypa uppfattningar om könen.

Även Fjeld (2015:49) noterar att den bild som ges av kvinnor i SO är anorlunda än den bild som ges av kvinnor i exempelvis *Bokmålsordboka* (2005). Bl.a. innehåller SO en stor andel språkprov med ett aktivt *hon*. Det finns emellertid, enligt författaren, (2015:54) fortfarande tydliga spår av könsstereotypa gestaltningar i ordboken.

Det är enkelt att finna språkprov i SO (2009) som, gärna tillsammans med andra exempel med liknande innehåll, kan ifrågasättas. För det första finns exempelvis (under uppslagsorden *revir* och *sluka*) språkproven *köket var hennes revir* och *han praktiskt taget slukade henne med blicken*. För det andra finns det (under *håhå* och *kräk*) exempel som *håhå, vad är det för en karl som inte kan laga bilen* och *han är ett beskedligt kräk som aldrig vågar säga emot sin fru*. Exempel som dessa tyder på att SO-exemplen med kvinnor rör traditionella kvinnosysslor, att kvinnor objektifieras i ordboken, och att ordboken ger tydliga signaler om hur man ska uppträda för att betraktas som en riktigt karl.

Exemplen under *revir* och *sluka* har för övrigt lyfts fram i en radiodebatt mel-

lan litteratur- och genusforskaren Ebba Witt-Brattström och Sven-Göran Malmgren (Sveriges Radio 2016). I debatten förfäktade Witt-Brattström åsikten att SO befäster könsförtryckande strukturer och reproducerar stereotypa könsroller. Malmgren återopade det genustänk som genomsyrat arbetet vid redaktionen inför SO (2009). Vidare påpekade han att det bl.a. finns en rad SO-exempel där också män är aktiva i köket. Malmgren konstaterade avslutningsvis att ”man får nog komma med lite mer än en handfull anekdotiska exempel” för att visa på att ordboken reproducerar stereotypa könsroller. (*Aftonbladet* 22 dec. 2016).

[5] TJEJER OCH KILLAR I SO (2009) – RESULTAT AV EN FRITEXTSÖKNING

Mer systematiska undersökningar av hur flera, semantiskt relaterade ord som *dam*, *herre*, *kvinna*, *man*, *gubbe*, *gymma*, *flicka*, *pojke*, *drottning* och *kung* presenteras i SO (2009) kan utgöra stora och tidskrävande uppgifter. Detta gäller i synnerhet om man inte bara studerar aktuella uppslagsord utan också granskar hur de aktuella orden används i andra ordboksartiklar. Se t.ex. Kram som jämför artiklarna *kvinna* och *man* i en av föregångarna till SO (2009). Se också Agebjörn (2016) som studerar hur huvudbetydelserna hos bl.a. *kvinna*, *man*, *tant*, *farbror*, *gumma* och *gubbe* återges i SO, hur motsvarande ord används i textmaterial som ingår i Språkbanken m.m.

Inom ramen för denna artikel har jag valt att granska samtliga träffar av olika böjningsformer av substantiven *tjej* respektive *kille* i SO (2009):s språkexempel till enskilda huvud- och underbetydelser. Att ta fram detta material är inte svårt; det kan vem som helst göra via en fritextsökning i appversionerna av SO.

Sammanlagt finns det 31 språkprov som innehåller *tjej* och 30 språkprov som innehåller *kille* (se tabell 1 i bilagan till denna text). I vissa fall är det betydelserna ’flicka’ resp. ’pojke’ som är aktuella, andra gånger är det betydelserna ’flickvän’ resp. ’pojkvän’ som gäller.

En genomgång av materialet visar att en hel del SO-exempel kan betraktas som relativt sakliga. I någon mån speglar de också dagens verklighet (jfr Ohlander i avsnitt 2 ovan). Det gäller t.ex. *det är ofta svårt att rekrytera tjejer till de tekniska programmen* respektive *sångerskan ackompanjerades av tre killar på gitarr, synt och percussion*. Det kan dock vara samhällliga förändringar på gång beträffande dessa förhållanden vilka redaktörerna givetvis måste vara uppmärksamma på.

Bland exemplen finns det tjejer och killar som är såväl aktiva som passiva. I ett fall som *han tafsade på en tjej i krogkön och fick en spark i skrevet tillbaka* är det både killen och tjejen som tar initiativ och agerar – även om handlingarna är av olika slag. Vidare beskrivs både tjejer och killar i positiva respektive negativa ordalag (jfr t.ex. det mer tilltalande *en salt tjej* med det mer nedlåtande *en mesig liten kille*). Om beskrivningen *hans tjej var alltid villig* ska betraktas som positiv

eller negativ lämnas till läsaren av denna text att avgöra.

Av materialet framgår att både *tjej-* och *kille-*språkprov innehåller beskrivningar av utseende. Detta är dock vanligare i fall som inkluderar ordet *tjej*. Några exempel ur ordboken är *han hade fått korn på en snygg tjej, en ärtig tjej, en sportig tjej med hästsvans och keps, de nakna tjejerna på herrtidningens mittuppslag, en herrtidning med bilder på tjejer med nakna bröst och tjejerna hade tajta jeans*. Killars utseende beskrivs främst genom adjektivet *snygg*: *en snygg kille, hennes nye kille är en riktig snygging, hon hade en dejt med den snyggaste killen i klassen*.

Bland *tjej-*exemplen finns det också ett flertal språkprov som lyfter fram personliga egenskaper och karaktärsdrag (t.ex. *en salt tjej, en skarp tjej, en tjej på sexton vårar, en torr tjej som aldrig ville gå på diskotek* samt redan nämnda *hans tjej var alltid villig*). Vissa tjejer tycks vara ganska all dagliga (som i *askungesagan om en vanlig tjej som gifte sig med en miljardär*) och andra tycks vara bestämda och viljestarka (*tjejerna har tuffat till sig och vågar säga sin mening*). Det är emellertid betydligt vanligare att karaktärsdrag lyfts fram i *kille-*exemplen (se exempelvis *en kul kille, en trevlig och lättsam kille, en städad kille, en mesig liten kille, en praktisk kille som renoverade huset själv, en teknisk kille som alltid lagar bilen själv, han är en reko kille, han verkar vara en hygglö kille* samt *den killen är helt nollställd*). Exemplen ger som synes prov på flera manliga stereotyper, bl.a. den om den händige mannen, ”hemmafixaren”. I flera språkprov målas det även upp en bild av killar som mer aggressiva, våldsbenägna etc. (se t.ex. *killar i tonåren är ofta utagerande och aggressiva, killarnas snack var späckat med svordomar* och *han spöade upp killen som muckade gräl med honom på puben*).

Bland såväl *tjej-*exemplen som *kille-*exemplen kan man se att framställningen av tjejer och killar utgår ifrån det motsatta könets perspektiv. I fall som bl.a. dessa betraktas tjejer eller händelser som rör tjejer ur ett mansperspektiv: *hans flickvän gjorde slut och nu dejtar han en annan tjej, han råkade sätta sin tjej på jäsning, han pejlade in varenda tjej i lokalen, hans nya tjej, hans tjej är så skön, han började genast stöta på den snyggaste tjejen, killarna skulle gå på diskotek och ragga brudar*. I fall som dessa återges istället män ur ett kvinnoperspektiv: *hon hade en dejt med den snyggaste killen i klassen, hon föredrog intellektuella killar, hon ska träffa sin kille ikväll, hon slog upp när hon träffade en annan kille* och *hon strular med två killar*. Rent frekvensmässigt är det klart vanligare att tjejer i materialet återges ut ett manligt perspektiv än tvärtom.

Avslutningsvis finns det språkprov som är obegripliga ur alla perspektiv. Som redan framgått kopplar Nikula (2015) SO:s språkprov till ordbokens grad av användbarhet. Språkprovet *en kille och två tjejer bildade kvällens första triangel*, som ska stödja en bildlig, erotisk innebörd av uppslagsordet *triangel*, stöder knappast

betydelsebeskrivningen. Snarare fungerar det som distraktor för ordboksanvändaren och bör därför ersättas av något annat.

Sammanfattningsvis rymmer SO (2009) *tjej*- och *kille*-språkprov av mycket olika slag, Genomgången av materialet visar emellertid att *tjej*-exemplen innehåller fler beskrivningar av utseende, inklusive drag av objektifiering. *Kille*-exemplen handlar mer om karaktärsdrag. Många situationer som gäller tjejer beskrivs ur ett manligt perspektiv. Det är heller inte ovanligt med schabloner kring hur män tänker och uppträder och indikationer om hur en ”riktig” karl ska vara (jfr Nikulas 1997 diskussion kring dold och mer försåtlig diskriminering eller smygsexism i avsnitt 2 ovan). Av min undersökning framgår det således att SO (2009) kan sägas upprätthålla eller kanske t.o.m. befästa stereotypa tankegångar om hur kvinnor – men också män – förväntas vara, ha för kunskaper o.s.v. (jfr Kram 1998:159–160; Ishikawa 2013).

Vad finns det då för stöd för denna återgivning av de två substantiven i korpusar med svenskt samtida språk? Om SO:s primära uppgift är att beskriva språkbruket och unga kvinnor och män beskrivs på olika sätt och förekommer i olika sammanhang i textmaterialen, kanske detta ordboksinnehåll är fullt rimligt. För att få klarhet i hur språkbruket faktiskt ser ut har jag med konkordansverket Korp gjort en sökning på lemmogrammen *tjej* respektive *kille* i en delmängd av Språkbankens korpusar. De aktuella korpusarna består av moderna texter av olika slag (tidningstexter, romaner, akademiska texter och bloggtexter) och antalet löpande ord i delmängden är drygt 2 miljarder. Med tanke på alla de texter som finns att tillgå i Språkbanken och de stora korpusar som man numera nyttjar inom lexikografiskt arbete måste delmängden betraktas som relativt begränsad. Ett skäl till avgränsningen är att många typer av sökningar i Korp tyvärr är mycket tidsödande – ur lexikografens perspektiv tar det orimligt lång tid att få fram ett resultat av sin sökning.

Totala antalet förekomster av böjningsformer av *tjej* och *kille* är 354 135 respektive 292 392 i det aktuella textmaterialet. Med tanke på det stora antalet träffar är det för lexikografen närmast omöjligt att titta på hur ordet används i KWIK-konkordanser. Däremot är Korps ordbildsvisning mycket användbar (se vidare Borin et al. 2012). Tack vare ordbilden kan man med lätthet överblicka det syntaktiska sammanhang som ett ord förekommer i. Av ordbilden framgår också ordets typiska subjekt, objekt och adverbial (jfr motsvarande resurser i anslutning till Leksikografisk bokmålskorpus och KorpusDK; se vidare bl.a. Fjeld & Henriksen 2012 och Trap-Jensen 2009 om dessa korpusar). En genomgång av sökresultaten i de svenska korpusarna visar bl.a. att det i hög grad är samma attribut som återkommer bland 10-i-topp-orden i de bägge ordbilderna. Såväl tjejer som killar beskrivs med ord som *ung*, *liten*, *duktig*, *snygg*, *glad* och *fin*. Det är

dock vanligare att bruket av ordet *tjej* sammanfaller med adjektivet *söt*. På samma sätt är det vanligare att *kille* sammanfaller med adjektivet *snäll*. Resultat som dessa behöver beaktas vid revidering av ett svenskt ordboksunderlag.

Utifrån det ovan sagda, borde ett flertal exempel i SO (2009) modifieras eller helt bytas ut i samband revideringen av databasen. För att verkligen kunna ta ställning till enskilda språkexempel som dessa måste ordboksredaktören emellertid kontrollera språkbruket i olika moderna korpusar, studera i vilka sammanhang exemplen förekommer, huruvida de betydelse som orden ska illustrera är försedda med brukskommentarer etc. Om det via en kommentar framgår att en viss användning är nedsättande påverkar det givetvis bedömningen av det aktuella språkprovet.

[6] ANVÄNDARSYNPUNKTER PÅ SO (2009):S INNEHÅLL

Numera är det inte ovanligt att ordboksinnehåll uppmärksammas och debatteras i olika sammanhang på internet. Som exempel kan den diskussion som rått kring språkexemplet *rabid feminist* under uppslagsordet *feminism* i *Oxford Dictionary of English* nämnas (se vidare t.ex. Bolt 2016, Flood 2016 och 2020; se även twitterinläggen under t.ex. #dictionarygate).

Visst innehåll i SO (2009) har också fått stor uppmärksamhet bland ordboksanvändare. Först och främst är det ordboksartikeln *mesig* som stått i fokus. Ordet, med huvudbetydelsen 'rädd och ynklig', har två underbetydelse och den ena av dessa åtföljs av ett, enligt nuvarande redaktion, mycket olyckligt språkprov, *mesig tjejfotboll*. I korthet tycks nyheten om ordboksinnehållet ha spridit sig genom att en privatperson uppmärksammade Svenska Fotbollförbundet (SvFF) på att damfotboll nämndes i ett mindre smickrande sammanhang i SO. SvFF publicerade ett inlägg på damlandslagets officiella Instagram-konto med en bild på ordboksartikeln åtföljt av frågan "Har vi inte kommit längre än så här Svenska Akademien?" (Fagerlund 2018). Inlägget spreds sedan snabbt vidare. Språkprovet uppmärksammades också av flera svenska tidningar och snart hade SO-redaktionen mottagit ett förhållandevis stort antal e-brev, inklusive ett från SvFF. En del mejl var sakligt formulerade, andra mer aggressiva och nedlåtande till sin natur. I ännu några mejl uttrycktes sorg över sakernas tillstånd (se utdrag ur några brev i nedan).

Hur kan detta hända 2018? Vilka signaler tror ni att det ger?! Gör om, gör rätt! (Av Camilla H.)

Det räckte inte med att vi damspelare får de sämsta omklädningsrummen, de sämsta träningsstiderna på de mest sönderspelade leriga planerna ... samt en massa skit bland media och allmän opinion – ni

kände att ni ville trycka till oss lite grammatiskt också? :) (Av Jenny W.)

Om jag söker på ordet mesig så får jag upp "mesig tjejfotboll" som förslag ..., varför inte byta ut mot "Mesig ledamot av "fina" Svenska Akademien" istället? Har vi inte kommit längre 2018? Sorgligt att NI av alla uttrycker er så, pinsamt. Klipp er och skaffa er ett nytt jobb, jag tar gärna ert arbete och lön så kommer det bli ordning och reda. (Av Olle G.)

... För mig som mamma till två fotbollsspelande tjejer så är de allt annat än mesiga. Pinsamt är det!!! Hoppas på snar ändring!! (Av Ulrika F.)

Mycket av materialet i SO (2009) härstammar från föregångarna SOB och NEO och en hel del exempel är därmed från 1980-talet och av naturliga skäl i behov av modernisering. Så är emellertid inte fallet med detta språkprov: det lades in i databasen inför SO (2009). Eventuellt hade redaktionen kunnat åberopa att *mesig tjejfotboll* är en återkommande ordkombination och att den därför försvarar sin plats i en ordbok som ska återge hur modernt språkbruk ser ut, men så är det heller inte. En sökning i Korp visar att ordkombinationen, då språkprovet uppmärksammades, inte förekom en enda gång i de då allmänt tillgängliga korpusarna på mer än 13 miljarder löpande ord.

Så, för att sammanfatta, samtidigt som ett stort antal språkexempel uppdaterades och reviderades ur genussynpunkt inför SO (2009), inarbetades det olämpliga exempel i databasen. Det enda redaktionen nu kunde göra i fallet *mesig tjejfotboll* var att svara skribenterna att modernisering av ordboks-databasen pågår och att SO har till uppgift att spegla det faktiska språkbruket men däremot inte att ge rekommendationer om hur orden ska användas (jfr avsnitt 3 ovan). Det betonades också att redaktionen inte står bakom alla de åsikter eller värderingar som kan komma till uttryck i de exempel som ges. I svaret meddelades också att exemplet var borttaget ur databasen. I och med att SO i svenska.se inte uppdateras löpande är detta emellertid inte synligt för användarna ännu.

Redaktionens svar till SvFF och användarna spred sig snabbt och snart kunde man läsa mer positiva rubriker i bl.a. pressen. Se t.ex. utdraget ur *Dagens Nyheter* i figur 1.



FIGUR 1. Del av nyhetsartikel, publicerad i *Dagens Nyheter* 21 febr. 2018

Just detta exempel illustrerar det faktum att ordböckers innehåll kan granskas och spridas på ett annat sätt än tidigare. Det visar på att även ordböcker och ordboksarbete lyder under den nya mediala logik som tillkomsten av internet och sociala medier har gett upphov till. Är det något i ordboken som inte tillfredsställer användarna hör de av sig på ett annat sätt än förut. Redaktionen är bara en knapptryckning bort och detta har både positiva och negativa sidor. Lexikograferna tar emot synpunkter och kommentarer av olika slag och vissa av dem kan leda till förbättringar av databasen – om det finns stöd för dessa ändringar i de språkliga material som används i samband med det lexikografiska arbetet.

[7] DET PÅGÅENDE OCH FRAMTIDA ARBETET MED SO
– UR KÖNSPERSPEKTIV

Hittills har upplagan från 2009 uppmärksamats. Hur ser det då ut inför nästa upplaga av SO? Det arbete som lades på att utveckla innehållet ur ett genusperspektiv i den första utgåvan har fortgått med sikte på kommande upplaga av ordboken. Det är ett stort arbete som bl.a. innebär att många språkprov revideras eller helt enkelt stryks. Som redan framgått baserades första upplagan av SO (samt föregångarna SOB 1986 och NEO 1995–1996), i första hand på etablerat, redigerat skriftspråk i tidningar och romanmaterial. Sedan arbetet med den första upplagan av SO pågick har tillgången till textmaterial ökat i det närmaste lavinartat och exempelvis Språkbanken innehåller nu betydligt fler delkorpusar av olika slag än tidigare. Inte minst är vardagligt, talspråksnära språk nu tillgängligt på ett helt annat sätt än tidigare och det måste lexikografen förhålla sig till. Det är svårt att hävda att inte texter publicerade på åtminstone vissa sociala

medier ska in i underlaget för den nya upplagan (se vidare Sköldberg & Hannesdóttir 2017).

Som tidigare konstaterats finns det inte, med tanke på Korp, stöd för att lägga in en ordkombination som *mesig tjejfotboll* i databasen. Däremot kan det finnas fog för att lägga in eller bevara andra uppslagsord, betydelsebeskrivningar och ordkombinationer – hur provocerande och sexistiska de än må vara. Som exempel är det, i modern svenska, framför allt kvinnoben som är *välsvarvade*. Vidare är det främst kvinnor som framställs som *neurotiska* och på det sexuella området är det framför allt kvinnor som betecknas som *lättfotade*. Söker man på vissa uppslagsord är det alltså lätt att få intryck av att kvinnor och kvinnors kroppar objektifieras. Kvinnor kan också sägas tilldelas mindre tilltalande psykiska egenskaper. Dessutom skambeläggs kvinnlig sexualitet (se t.ex. Caldas-Coulthard & Moon 2010 om asymmetrier i hur kvinnor och män presenteras med hjälp av bl.a. adjektivattribut i engelsk press. Se även Fjeld & Knudsen 2013 om könsdiskriminering i norska texter som ingår i Leksikografisk bokmålskorpus). I detta sammanhang är det alltså inte lemmaansättningen som utgör problemet (jfr Nikula 2008a ovan), men aktuella ord måste beskrivas på ett lämpligt sätt. De måste exempelvis förse med uppgifter som informerar användarna sin negativa laddning m.m.

Balansgången mellan att stryka illa valda exempel, att ducka för reellt språkbruk och att bli för politisk korrekt inom ramen för det lexikografiska arbetet är svår (jfr Ohlander om val av okontroversiella ämnen i grammatikexemplen ovan). Korpusinnehållet, och de resultat som man får av korpussökningar, måste också diskuteras. Som exempel kan det svenska adjektivet *rabiat* anföras (jfr engelskans *rabid* och språkprovet *rabid feminist* i avsnitt 6 ovan). Enligt SO (2009) betyder ordet *rabiat* 'okontrollerat ursinnig' men det kan också användas rent förstärkande. Som exempel på denna underbetydelse anføres språkprovet *en rabiat nationalist*. Huruvida detta är en lyckad ordkombination kan diskuteras. Svaret beror nog på hur man ser på innebörden i ordet *nationalist* och om man anser att det har en positiv eller negativ klang. En ordbild av ordet, baserad på samma Korp-delmängd som användes i samband med undersökningen av orden *tjej* och *kille* i avsnitt 5, visar följande 15 ord som attribut till de 1878 beläggen på ordet.

feminist, manshatare, motståndare, flatfeministen, kärring, folklighetskult, menskossor, ateist, antisemit, flat-feministen, köttätare, eu-motståndarna, nationalist, eu-motståndare, rasist

Många av substantiven i listan har som synes påtagligt negativ klang. Eventuellt kan man tro att det är korpustexterna från sociala medier som får genomslag i

detta sammanhang och att ordbilden skulle se annorlunda ut om man endast använde redigerade texter med mer vårdat språk. En sökning i samma delmängd korpusar, men där endast texter hämtade från tidningar och skönlitteratur ingår, ger emellertid följande, i långa stycken liknande, ordbild:

feminist, folklighetskult, antisemit, motståndare, feministdävlav, manshatare, klimatalarmist, antisemitism, eu-motståndare, nedskärning, inställning, kritiker, nationalist, kritik, manskvinnor

Om redaktionen strikt rättade sig efter innehållet i olika texter i Språkbanken skulle således *rabiat feminist* kunna läggas in i databasen som språkprov. Det samma gäller ordkombinationer som *rabiat flatfeminist, rabiat kärring* etc. och frågan är hur lyckat det är. En beskrivning av SO:s uppdrag, att spegla faktiskt språkbruk, återfinns i och för sig i ordbokens inledning, men denna beskrivning tar de flesta inte del av. Här kan man därför tänka sig att en lämplig lösning kan bestå i att det nuvarande exemplet *en rabiat nationalist* ersätts av det mer neutrala *en rabiat motståndare*. Också denna ordkombination är förhållandevis vanlig i textmaterialen (se ovan) och exemplet torde inte utgöra en distraktor för den betydelse som det ska illustrera.

I ett fall som *rabiat* krävs det nästan att ordboksartikeln innehåller språkexempel med *rabiat* + substantiv för att typiskt bruk av ordet ska illustreras. Det är dock inte alltid så. Som exempel kan ett annat adjektiv, *känslig*, anföras. Ett sätt att göra framtida ordboksartiklarna i SO mer könsneutrala är, i enlighet med Whitcut (1983:143), att man byter ut pronomen som *han* till *hon*, *han* till *de* etc. i språkproven. Denna typ av utbyten skedde i samband med arbetet inför SO (2009). Ett annat sätt är att använda sig av fler s.k. ”döda” språkprov (Svensén 2009:147). Kännetecknande för dessa är att nominalfraser ofta har neutraliserats till indefinita pronomen och verbens finita former till infinitiver. I artikeln *känslig* i SO (2009) finns det flera språkprov, bl.a. *hon har en känslig mage; du får inte ryta åt henne, hon är så känslig* samt *han är känslig för kritik*. Dessa språkprov ska således stödja betydelsebeskrivningen och ge info om uppslagsordets fraseologi. I just detta fall framställs såväl kvinnan som mannen som svaga i något avseende och man kan tycka att de negativa exemplen i ordboksartikeln i viss mån ”tar ut varandra”. För att uppnå högre grad av könsneutral språkbeskrivning, men ändå bibehålla information om vilka ord adjektivet ifråga tenderar att uppträda med, kan emellertid två av språkproven kortas (jfr *ha känslig mage* och *vara känslig för kritik*). Ett exempel som *du får inte ryta åt henne, hon är så känslig* bör man dock överväga om man inte ska stryka eftersom betydelsen hos ordet *känslig* stöds av andra språkprov och det inte illustrerar någon återkommande ordkombination (jfr resonemangen kring distraherande exempel ovan).

[8] SAMTLIGA DISKRIMINERINGSGRUNDER MÅSTE BEAKTAS VID ORDBOKSARBETE

Genusperspektivet är viktigt, men fler perspektiv måste anläggas vid lexicografiskt arbete och ett normkritiskt perspektiv måste hela tiden vara aktuellt. Det behövs också fler normkritiska studier av hur det ser ut i dagens ordböcker. I nuläget är dessa studier begränsade till sitt antal (se dock Moons 2014 övergripande artikel med anförda referenser om ideologiskt laddade ord i engelska inlärningsordböcker. Se även t.ex. Nikula 2008b om åldersstereotyper i olika svenska ordböcker). Lexikografer behöver granska sina material utifrån samtliga diskrimineringsgrunder (dvs. kön, könsöverskridande identitet eller uttryck, etnisk tillhörighet, religion eller annan trosuppfattning, funktionsnedsättning, sexuell läggning, ålder) (se <https://www.do.se/om-diskriminering/vad-ar-diskriminering/>). Exempel på ord som kan relateras till svenska diskrimineringsgrunder är redan nämnda *tjej* och *kille* men också *brud*, *bitch*, *hora*, *icke-binär*, *cisperson*, *blatte*, *blekansikte*, *hedersvåld*, *adhd*, *funkofobi*, *funktionsvariation*, *miffö*, *vanför*, *bög*, *flata*, *fjolla*, *tant*, *gubbe* och *ålderism*. Några av dessa ord är äldre och upptas redan i många ordböcker medan andra är nyare och kanske inte fullt så etablerade i språket. Ord som dessa studerades förstås extra redan inför i SO (2009) men de behöver återigen granskas ur ett normkritiskt perspektiv.

Redaktörerna måste också, i enlighet med t.ex. Moon (2014), vara uppmärksamma på hur olika samhällsgrupper framställs inne i ordboksartiklarna. Som exempel kan frågan om hur äldre personer skildras nämnas (se t.ex. Moon 2014:86-87 om den negativa bild som ges av äldre i engelska ordböcker under det till synes oladdade ordet *cardigan*). Kritiska genomgångar av ordboksmaterial bör således inkludera studier av vilka uppslagsord som är med (inkl. vilka som saknas). De behöver också gälla vilka ord som används i beskrivningarna och innehållet i de språkprov som anföras. De måste också avse bruk av brukskommentarer av slaget <nedsättande>, <kan uppfattas som stötande> och <mest historiskt> vilka informerar användarna om uppslagsordens stilnivå, värdeladdning m.m. (jfr Hageberg i avsnitt 2 ovan).

För att ytterligare förtydliga ovan nämnda tankegångar i förhållande till innehållet i nästa upplaga av SO kan också substantivet *ras* nämnas. Detta ord har kommit att ifrågasättas i vetenskapliga sammanhang. Idén om människoraser är inte förenlig med modern, samtida vetenskap och av detta skäl har SO-redaktionen kommit fram till att antalet användningar av ordet bör begränsas så långt det bara går i nästa upplaga av ordboken. Samtidigt kan det inte helt tas bort. Ordet som sådant samt en rad sammansättningar med *ras* används fortfarande i modernt språk, t.ex. *rasbiologi*, *rasdiskriminering*, *rashets* och *rasmotsättning*. Dessa

får därmed vara kvar i SO, antingen som egna uppslagsord eller som morfologiska språkprov i andra artiklar. Vid behov behöver de dock förses med brukskommentarer av olika slag, t.ex. <ngt åld.>. Att helt bortse från dem och stryka dem ur ordboken vore dock att ge efter för politisk korrekthet i för stor grad (jfr Ohlander i avsnitt 2 ovan). I SO (2009) används ordet *ras* emellertid också inne i andra artiklar (se t.ex. *eskimå* och *mörkhyad*) och där kan det många gånger utelämnas eller ersättas av exempelvis *etnisk tillhörighet*.

Utöver diskrimineringsgrunderna behöver redaktörerna också granska andra kontroversiella teman som kroppsform, klass, politisk ståndpunkt, stad vs. landsbygd och sexualitet. Exempelvis kan det diskuteras hur övervikt och fetma framställs i ordböcker. En genomgång av SO (2009) visar t.ex. att till uppslagsord som *trakassera*, *hånfull* och *håna* hör språkprov som *den tjocka pojken trakasserades av kamraterna*, *den fete mannen omgavs av en hånfull skara barn* och *barnen hånade honom för hans fetma*. Betydelsen hos ett ord som *stilig* illustreras istället av ett exempel som *hon var en stilig kvinna, slank och rakryggad*. Frågan är vad språkprov som dessa ger för signaler. Kan man säga att de helt enkelt ligger i linje med den syn på övervikt som är förhärskade i samhället av idag? Eller reproducerar och förstärker de denna syn?

En för lexikografen komplicerande omständighet är att antalet ord i ordboks-databaser är mycket stort, och att bruket av ett visst ord snabbt kan förändras över tid. Vissa laddade ord kan desutom återtas, reclaimas, (se Edlund et al. 2007:209–210 och Wojahn 2015:129–130 om laddningen hos beteckningar som *bög* och *flata*). Alternativa ord som *funkofobi* och *funktionsvariation* lanseras men det är inte säkert att de etableras. Om de får större spridning är det inte självklart vilken eller vilka betydelser de då får och vilken laddning de har bland olika användargrupper.

[9] SLUTORD

Ordböcker har anklagats för att ge en skev bild av kvinnor och män, kvinnligt och manligt. Dels har det riktats kritik mot att det ordförråd som är förknippat med kvinnor och kvinnors liv, såsom beteckningar på traditionella kvinnoyrslor, inte är lika representerat i ordböckerna som motsvarande manliga ordförråd. Dels har det riktats kritik mot hur personer tillhörande det kvinnliga könet framställs överlag.

Den databas som ligger till grund för SO, dvs. *Svensk ordbok utgiven av Svenska Akademien*, har väldigt många förtjänster men betydande delar av innehållet i databasen är från 1980-talet och behöver revideras. En hel del har förändrats sedan SO (2009) utarbetades (se vidare Sköldberg & Hannesdóttir 2017). Detta

gäller inte minst det omgivande samhället, men det gäller också förutsättningarna för ordboksarbete, tillgången till korpusar och avståndet mellan ordboksredaktörerna och ordboksanvändarna.

Redaktionen för SO behöver arbeta vidare med den granskning och revidering ur genussynpunkt som skedde i databasen inför den första upplagan av verket. En nyanserad och varierad bild av både kvinnor och män, flickor och pojkar måste eftersträvas. Nikula (2008a:310) menar att vägen till en ordbok där könen är jämställda på alla nivåer tar sin början vid valet av uppslagsord. Olika ordförord bör täckas in i en ordbok som gör anspråk på att beskriva svenskt samtida språkbruk. Men ordboksredaktörerna måste också se hur de aktuella uppslagsorden definieras, om kontroversiella ord används i definitioner, vad språkproven ger för signaler och om brukskommentarer används på ett adekvat och rättvisande sätt.

Behovet av kritiskt lexikografiska studier är stort. Genom mer systematiska granskningar kan den smygsexism som Nikula (1997) diskuterar uppenbaras. Nuvarande SO-redaktion består av såväl kvinnor som män i olika åldrar, men det utgör, som Hageberg (1990) konstaterar, ingen garanti för att alla könsstereotyper tvättas bort. Desto viktigare är att det råder medvetenhet inom redaktionen kring normkritiska frågeställningar så att exempelvis inte nya, mindre lyckade exempel av slaget *mesig tjeifotboll* läggs in i databasen under revideringsarbetets gång (jfr avsnitt 6 ovan).

Balansgången mellan att stryka eller modifiera illa valda exempel och att bli för politisk korrekt är svår. Whitcut (1983:143–144) konstaterar att lexikograferna kan gå för långt i sin iver att radera könsstereotyper och att exemplet därigenom blir helt onaturliga ur brukssynpunkt (jfr även Ohlander om val av okontroversiella ämnen i grammatikexemplen). En viktig förutsättning för att ordboksarbetet ska bli så bra som möjligt är tillgången till större, balanserade och språkligt sett varierade korpusar. Utöver detta behövs välfungerande korpusverktyg som snabbt kan hjälpa lexikograferna att se återkommande mönster i stora material. Är bruket tillräckligt starkt i den korpus som används kan detta stödja ordboksförfattarna att på ett så adekvat sätt som möjligt ge sig i kast med kontroversiella ord och ordförbindelser (jfr *rabiät feminist*). Utbudet av såväl korpusar som adekvata korpusverktyg torde dock variera inom bl.a. Norden. I Sverige finns det stora korpusar med varierat språkbruk i bl.a. Språkbanken, men tillgången till senare textmaterial och snabba verktyg är tyvärr inte tillfredsställande i dagsläget.

En hel del torde trots allt ha hänt i svenska ordböcker de senaste decennierna och man borde kunna skönja förändringar till de bättre (jfr t.ex. Nikula 2010 och Fjeld 2015 om gamla surdegar i föregångare till SO). Det vore dock intressant att,

i analogi med Ohlander (1995) och Lewandowski (2014), anlägga ett historiskt perspektiv på svenska ordböcker och exempelvis se hur språkproven ser ut i Dalins enspråkiga ordbok från 1850-talet. Det kunde ge nyttiga perspektiv på dagens exempelmaningar.

Det räcker emellertid inte att anlägga ett genuskritiskt perspektiv under det lexikografiska arbetet. Alla diskrimineringsgrunder måste beaktas (dvs. kön, könsöverskridande identitet eller uttryck, etnisk tillhörighet, religion eller annan trosuppfattning, funktionsnedsättning, sexuell läggning, ålder). Ett normkritiskt perspektiv bör hela tiden vara aktuellt i samband med ordboks-författandet. Lexikograferna behöver också vara uppmärksamma på hur andra kontroversiella ämnen, t.ex. de idéer som rör vår tids kroppsfixering, behandlas i de verk de arbetar med. Förhoppningsvis kan det leda till mer inkluderande ordböcker i framtiden.

REFERENSER

- Aftonbladet* 2016. Genuskritik mot Svensk ordbok. Publicerat: 22 december 2016. Tillgängligt via <https://www.aftonbladet.se/nyheter/a/5K7gK/genuskritik-mot-svensk-ordbok>.
- Arimbi, Diah, A. & Deny A. Kwary. 2016. Linguistic Turn and Gendering Language in the Cambridge Advanced Learner's Dictionary. I *English Language teaching* 9, 166–174.
- Bokmålsordboka* 2005. 3. reviderade utgåva. Oslo: Kunnskapsforlaget.
- Bolt, Adam. 2016. Feminists attack Oxford Dictionary of English for 'reinforcing sexist stereotypes'. I *The Telegraph* 23 januari 2016. Tillgängligt via <https://www.telegraph.co.uk/news/uknews/12117574/Feminists-attack-Oxford-Dictionary-of-English-for-reinforcing-sexist-stereotypes.html>
- Borin, Lars, Markus Forsberg & Johan Roxendal. 2012. Korp – the corpus infrastructure of Språkbanken. I *Proceedings of LREC 2012*, 474–478. Istanbul: ELRA.
- Caldas-Coulthard, Carmen & Rosamund Moon. 2010. 'Curvy, hunky, kinky': Using corpora as tool for critical analysis. I *Discourse & Society* 21(2), 99–133.
- Edlund, Ann-Catrine, Eva Erson & Karin Milles. 2007. *Språk och kön*. Stockholm: Norstedts.
- Epple, Barbara. 2000. Sexismus in Wörterbüchern. I Ulrich Heid, Stefan Evert, Egbert Lehmann & Christian Rohrer (red.), *Proceedings of the 9th EURALEX International Congress*, 739–754. Stuttgart.

- Fagerlund, Frida. 2018. Svenska Akademien: ”Mesig tjeifotboll”. Publicerat i *Aftonbladet* 19 febr. 2018. Tillgängligt via <https://www.aftonbladet.se/sportbladet/fotboll/a/J1l3X8/svenska-akademien-mesig-tjeifotboll>.
- Fjeld, Ruth Vatvedt. 2001. Ingenting å snakke om – kvinneord og ordbøker. I *Språknytt* 4, 10–13.
- Fjeld, Ruth Vatvedt. 2015. Om ordbokseksempel og stereotypisering av kjønn i noen nordiska ordbøker. I Caroline Sandström, Ilse Cantell, Eija-Riitta Grönros, Pirkko Nuolijärvi & Eivor Sommardahl (red.), *Perspektiv på lexikografi, grammatik och språkpolitik i Norden*, 35–65. Helsingfors.
- Fjeld, Ruth Vatvedt & Petter Henriksen. 2012. The BRO-project, a bridge in the wild, Norwegian linguistic landscape. I Ruth Vatvedt Fjeld & Julie Matilde Torjusen (red.), *Proceedings of the 15th EURALEX International Congress*, 936–946. Oslo: Institutt for lingvistiske og nordiske studier, Universitetet i Oslo.
- Fjeld, Ruth Vatvedt & Rune Lain Knudsen. 2013. Balance, Depth, Synergies – Assembling a Corpus for Documenting Ideal and Reality in Lexis. I Deny A. Kwary et al. (red.), *Lexicography and Dictionaries in the Information Age. Selected papers from the 8th ASIALEX International Conference*, 360–361. Surabaya: Airlangga University Press.
- Flood, Alison. 2016. Sexism row prompts Oxford Dictionaries to review language used in definitions. I *The Guardian*, 25 januari 2016. Tillgängligt via <https://www.theguardian.com/books/2016/jan/25/oxford-dictionary-review-sexist-language-rabid-feminist-gender>.
- Flood, Alison. 2020. No more ‘nagging wives’: how Oxford Dictionaries is cleaning up sexist language. I *The Guardian*, 6 mars 2020. Tillgängligt via <https://www.theguardian.com/books/2020/mar/06/no-more-nagging-wives-how-oxford-dictionaries-is-cleaning-up-sexist-language>.
- Josephson, Olle. 2018. *Språkpolitik*. Stockholm: Morfem.
- Hageberg, Arnbjörg. 1990. Kvinner og men og ordbøker. *Norsk Lingvistisk Tidsskrift* 8, 48–67.
- Hult, Ann-Kristin. 2010. Kort och gott. Om idiomens språkprov i Svensk ordbok utgiven av Svenska Akademien (2009). I Kristina Nikula & Harry Lönnroth (red.), *Nordiska Studier i Lexikografi* 10, 209–222. Tammerfors.
- Ishikawa, Yuka. 2013. Gender stereotypes seen in dictionary descriptions published in Japan. I Deny A. Kwary et al. (red.), *Lexicography and Dictionaries in*

- the Information Age. Selected papers from the 8th ASIALEX International Conference*, 222–227. Surabaya: Airlangga University Press.
- Korp, *Språkbankens konkordansverktyg*. <https://spraakbanken.gu.se/korp/>, hämtat augusti 2019.
- KorpusDK. <https://ordnet.dk/korpusdk>, hämtat augusti 2019.
- Kram, Kristina. 1998. ”... blott rent undantagsvis om kvinnor.” Om representation av kvinnan i moderna ordböcker. I *Språk och stil* 8, 159–171.
- Landau, Sidney I. 2001. *Dictionaries. The art and craft of lexicography*. (Second edition). Cambridge: Cambridge University Press.
- Leksikografisk bokmålskorpus*. <https://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/>, hämtat augusti 2019.
- Lewandowski, Marcin. 2014. Gender stereotyping in EFL grammar textbooks. A diachronic approach. I *Linguistik online* 68, 83–99. Tillgänglig via <https://bop.unibe.ch/linguistik-online/article/view/1635/2763>.
- Malmgren, Sven-Göran. 2009. *LEDA-Nyt* 47, 14–20.
- Mattisson, Anki. 2006. Genusperspektiv på SAOB:s källor. I *LexicoNordica* 13, 55–68.
- Moon, Rosamund. 2014. Meanings, Ideologies, and Learners’ Dictionaries. I Andrea Abel, Chiara Vettori & Natascia Ralli (red.), *Proceedings of the 16th EURALEX International Congress*, 85–105. Bolzano.
- NEO = Nationalencyklopedins ordbok* 1995–1996. Band 1–3. Höganäs: Bra böcker.
- Nikula, Kristina. 1997. Språkprov i könsperspektiv. I Rune Ingo, et al. (red.), *Fackspråk och översättningsteori. VAKKI-symposium XVII*, 194–206. Vasa: Vasa universitet.
- Nikula Kristina. 2008a. Kvinnligt och manligt. Illustrationerna i Bonniers svensk ordbok I Kristinn Jóhannesson et al. (red.), *Nog ordat? Festskrift till Sven-Göran Malmgren den 25 april 2008*, 304–311. Göteborg.
- Nikula Kristina. 2008b. Pensionär – aldrig i livet. De äldre i ordböckerna. I Ásta Svavarsdóttir et al. (red.), *Nordiske studier i leksikografi* 9, 337–351. Reykjavík.
- Nikula, Kristina. 2010. Svensk ordbok – en guldgruva för språkintresserade. I *LexicoNordica* 17, 351–375.

Ohlander, Sölve. 1995. "A wise Man Controls His Passions" – Of men and women, morals and mentality in four hundred years of English grammar examples. I Gunnel Melchers & Beatrice Warren (red.), *Studies in Anglistics*, 217–236. Stockholm: Almqvist & Wiksell.

Pilke, Nina. 2009. Hon arbetar som extrabiträde – han blev chef för avdelningen. Kvinnlig och manlig yrkesbild i ljuset av exempelmeningar i en tvåspråkig cd-romordbok. I Mona Enell-Nilsson & Niina Nissilä (red.), *Language and Power. VAKKI Symposium XXIX*, 332–342. Vasa, Vasa universitet.

SAOB = *Svenska Akademiens ordbok*, 1898–. Tillgänglig via <https://svenska.se/>.

SAOL = *Svenska Akademiens ordlista*, 2015. Fjortonde upplagan. Tillgänglig via <https://svenska.se/>.

Sköldberg, Emma & Anna Helga Hannesdóttir. 2017. Svenska ord – men vilka? Om uppslagsorden i Svensk ordbok utgiven av Svenska Akademien. I Emma Sköldberg et al. (red.), *Svenskans beskrivning 35*. Göteborg: Göteborgs universitet, 329–340.

Språkbanken. <https://spraakbanken.gu.se/>, hämtat augusti 2019.

SO = *Svensk ordbok utgiven av Svenska Akademien*, 2009. Tillgänglig via <https://svenska.se/>.

SOB = *Svensk ordbok* 1986. Solna: Esselte studium.

Svensén, Bo. 2009. *A handbook in lexicography. The theory and practice of dictionary-making*. Cambridge.

Sveriges Radio 2016. Svensk ordbok kritiseras för stereotypa könsroller. Sändes 22 december 2016. Tillgängligt via <https://sverigesradio.se/sida/artikel.aspx?programid=83&artikel=6591853> (augusti 2019).

Trap-Jensen, Lars. 2009. Access to multiple lexical resources at a stroke: integrating dictionary, corpus and Wordnet data. I Sylviane Granger & Magali Paquot (red.), *eLexicography in the 21st century. New challenges, new applications. Proceedings of eLex 2009*, 295–305. Louvain-la-Neuve.

Whitcut, Janet. 1983. Sexism in dictionaries. I Reinhard R. K. Hartmann (red.), *LEXeter '83. International Conference on Lexicography at Exeter*. (Lexicographica. Series Maior 1.), 141–144. Berlin: De Gruyter.

Wojahn, Daniel. 2015. *Språkaktivism. Diskussioner om feministiska språkförändringar*

i Sverige från 1960-talet till 2015. (Skrifter utgivna av institutionen för nordiska språk vid Uppsala universitet 92.) Uppsala.

BILAGA

Språkprov i SO 2009 som ska stödja huvudbetydelse och underbetydelse och som innehåller någon böjningsform av substantiven *tjej* eller *kille*. Aktuella uppslagsord indikeras med hjälp av fetstil.

Tjej	Kille
askungesagan om en vanlig tjej som gifte sig med en miljardär	hon hade en dejt med den snyggaste killen i klassen
en herrtidning med bilder på tjejer med nakna bröst	hon föredrog intellektuella killar
hans flickvän gjorde slut och nu dejt han en annan tjej	han verkar vara en hygglo kille
flukta på tjejer	hon ska träffa sin kille i kväll
tjejer är bara förnamn på hans problem	tjejer och killer
han råkade sätta sin tjej på jäsning	en kul kille
en sportig tjej med hästsvans och keps	en trevlig och lätt kille
han kladdar på tjejerna när han dansar	en mesig liten kille
han hade fått korn på en snygg tjej	den killen är helt nollställd
en läskigt snygg tjej	sångerskan ackompanjerades av tre killar på gitarr, synt och percussion
de nakna tjejerna på herrtidningens mittuppslag	en praktisk kille som renoverade huset själv
han pejlade in varenda tjej i lokalen	en stor kille på en racer körde om honom

det är ofta svårt att rekrytera tjejer till de tekniska programmen	killarna skulle gå på diskotek och ragga brudar
ordet "tjej" är lånat från romani	han är en reko kille
en salt tjej	ranta efter killar
en skarp tjej	en del killar tycker att det är skämmigt att gå till ungdomsmottagningen
hans tjej är så skön	" Skärpning nu, killar!" ropade tränaren
en snygg tjej	hon slog upp när hon träffade en annan kille
han började genast stöta på den snyggaste tjejen	en snygg kille
han tafsade på en tjej i krogkön och fick en spark i skrevet tillbaka	hennes nye kille är en riktig snygging
tjejerna hade tajta jeans	han spöade upp killen som muckade gräl med honom på puben
killar och tjejer	hon strular med två killar
den snyggaste tjejen i klassen	en städad kille
tjejerna tar över det ena chefsjobbet efter det andra	en kille med sug i blicken
hans nya tjej	killarnas snack var späckat med svordomar
en torr tjej som aldrig ville gå på diskotek	killarna har börjat tagga upp inför kvallet
en kille och två tjejer bildade kvällens första triangel	en teknisk kille som alltid lagar bilen själv
tjejerna har tuffat till sig och vågar säga sin mening	en kille och två tjejer bildade kvällens första triangel

hans tjej var alltid **villig**

killar i tonåren är ofta **utagerande** och
aggressiva

en tjej på sexton **vårar**

en kille i gänget som alltid försöker spela
översittare

en **ärtig** tjej

KONTAKT

Emma Sköldberg

Institutionen för svenska språket, Göteborgs universitet

emma.skoeldberg@svenska.gu.se

SKÆVHEDER OG LIGESTILLING: I ORDBOGEN, I KORPUS OG I SAMFUNDET

LARS TRAP-JENSEN

RESUMÉ

Sproget får vi i arv fra tidligere generationer. Det ændrer sig kun langsomt og er derfor fuldt af vidnesbyrd om ældre tiders opfattelser og syn på verden. Det gælder ikke mindst identitetspolitiske spørgsmål som forholdet mellem kønnene eller synet på de fremmede. I denne artikel ser jeg på hvordan *Den Danske Ordbog* forsøger at navigere mellem videnskabelighed, deskriptiv forpligtelse og et stigende krav om ligestilling og inklusion når de semantiske beskrivelser revideres og nye ord optages. Redegørelsen omfatter både ordbogen og dens korpusmæssige beskrivelsesgrundlag, og hvor det er relevant, sammenlignes resultaterne med de tilsvarende for norske forhold.

[1] INDLEDNING

Ordbøgernes definitionspraksis har ændret sig gennem de seneste 100 år. Det hænger givetvis sammen med at beskrivelsesgrundlaget i samme periode har ændret sig. For omkring 100 år siden blev den deskriptive tradition fastlagt i Norden med nationale opslagsværker som *Ordbog over det danske Sprog* (ODS) i Danmark og lidt senere *Norsk Riksmålsordbok* (NRO) i Norge. Sidstnævnte er i dag omarbejdet til en moderne ordbog og udgivet under navnet *Det Norske Akademis ordbok* (NAOB). Traditionen for deskriptive ordbøger er grundlagt på systematisk og videnskabelig beskrivelse af ordforrådet på empirisk grundlag. Til grund lå dengang store samlinger af sproglige excerpter registreret på kartotekskort, og metoden sikrede at leksikograferne ikke alene konsulterede deres egen sprogforfølelse, men kunne støtte sig til faktisk forekommende sprogbrug. Alligevel var antallet af excerpter for hver enkelt leksikalsk enhed, forstået som et lemma i én af dets betydninger, af overskuelig størrelse, og for lavfrekvente lemmers vedkommende undertiden nok så begrænset. Nationalordbøgernes definitionspraksis må forstås på denne baggrund.

Min erfaring med brug af disse ordbøger siger mig at redaktørerne ofte udformede deres definitioner så de i videst muligt omfang dækkede samtlige de belæg som de fandt i excerpterne. En tilsvarende observation er gjort af Nilstun (2012: 130), der ser NRO's definitioner som tegn på "svak abstraksjon", hvormed

menes at definitionerne blev udformet så de var skræddersyede til konkrete enkeltbelæg. Udviklingen af sproglige korpuser op gennem det 20. århundrede medførte blandt andet at det empiriske beskrivelsesgrundlag voksede eksponentielt fra 1960'erne og fremefter. Dermed voksede også antallet af konkrete belæg som redaktørerne fandt for de leksikalske enheder, hvilket umuliggjorde en praksis der omfattede enkeltforekomster i materialet. I stedet er tendensen gået i retning af at beskrive den centrale og typiske sprogbrug. To eksempler kan illustrere forskellen i praksis.

Øjenlaag, et.

1) hos mennesket og visse (højere) dyr: hver af de to tynde, bevægelige hud-folder, -klapper, der udgaar fra henh. øverste og nederste øjenhulerand, og som kan trækkes beskyttende hen over øjeæblets forflade; laag (2.3)

FIGUR 1: ODS: Udsnit af artiklen Øjenlaag

Figur 1 giver et indtryk af hvordan en definition i ODS gerne beskriver detaljeret og med maksimal præcision. Midlerne er indskrænkninger og præciseringer i form af indskudte sætninger eller fraser, forbehold i parenteser og synonymer eller supplerende forklaringer efter komma eller semikolon. Detaljerighed og udtømmende forklaring har højeste prioritet. Prisen for semantisk præcision er ofte læsbarhed: Forklaringerne kan være syntaktisk komplicerede at afkode, og der kan optræde sjældne eller vanskelige ord der stiller store krav til brugeren (*øjnehulerand*, *forflade*). Ordbøgerne henvendte sig snarere til et veluddannet mindretal end til skoleelever og den almene befolkning.

De store korpuser der er vokset frem siden slutningen af 1900-tallet, har efterhånden gjort det umuligt for en enkelt redaktør at overskue det samlede empiriske materiale, og dermed gøres en udtømmende beskrivelse af alle forekomster nærmest umulig. Inden for leksikografien er tendensen derfor dels gået i retning af mere automatisk præprocessering hvor computere automatisk behandler teksterne og finder karakteristiske mønstre i korpus, dels i at resultatet opdeles i semantiske klynger på grundlag af typisk sprogbrug. Tilsvarende bliver leksikografens opgave i højere grad at beskrive disse centrale og typiske mønstre og i mindre grad at redegøre for grænsetilfælde og perifere forekomster (jf. Trap-Jensen 2014: 53f.).

Figur 2 fra Den Danske Ordbog (DDO) viser at det centrale betydningsindhold beskrives som det første. I eksemplet er der desuden en efterhængt tillægsoplysning der viser at det netop er det typiske: "især" antyder at der er andre muligheder.

øjenlåg substantiv, intetkøn
bevægelig hudfold der kan dække øjet, især den øverste hudfold

FIGUR 2: DDO: udsnit af artiklen *øjenlåg*

Det er en helt almindelig definitionsmåde i moderne ordbøger og benyttes også i fx NAOB:

øyelokk substantiv
hudfold over (og under) [øveeplet](#)

FIGUR 3: NAOB: udsnit af artiklen *øyelokk*

Tanken er at man ved at beskrive det typiske nemt og hurtigt fremkalder en forståelse hos brugeren, og at denne derefter er i stand til at ekstrapolere sin forståelse fra typiske til mere perifere forekomster, i dette tilfælde eksempelvis hvis brugeren har mødt *øjenlåg* brugt om et dyr.

[2] SAMFUNDET, SPROGET OG ORDBØGERNE

Imidlertid er også denne praksis i de senere år blevet udfordret af en ny tendens. Ordbøger mødes i stigende omfang med et krav om at sprogbeskrivelsen lever op til ideelle fordringer om ligestilling og sproglig inklusion. Efter nogle gruppers mening kan ordbøgerne ikke nøjes med at beskrive ordene og sproget sådan som de afspejles i et korpus, men bør i højere grad beskrive dem sådan som de burde bruges. Leksikografen kan efter denne forestilling ikke nøjes med at iagttage og beskrive, men har et moralsk ansvar for at handle aktivistisk og rette op på uønskede skævheder i sproget, skævheder som skyldes et efterslæb fra en vildfarende fortid som vi bør lægge bag os (se fx diskussioner heraf i Russell 2012, Nossem 2018 samt Moon 2014 nedenfor).

Leksikografisk er problemstillingen interessant og relevant, for selvom leksikografer ynder at opfatte sig selv som dokumenterende og deskriptive, er leksikografi for det første ikke nogen eksakt videnskab, men rummer mange elementer af subjektivt skøn, for det andet spiller sproget og ordbøgernes beskrivelse af det en vigtig rolle når det gælder normdannelse i samfundet – uanset hvilken opfattelse ordbøgerne har af sig selv. Ifølge Moon (2014) er det beskrivelsen af ikke-denotativ betydning og især det ideologisk ladede ordforråd der volder problemer:

there is a particular dilemma facing lexicographers attempting to deal with contested and ideologically loaded words: to balance a description of what data suggests about meaning with how in a post-

modern inclusive society, the relevant concept “ought” to be regarded and represented.

(Moon 2014: 89)

Metodisk er det dog vigtigt at skelne mellem hvilke normer der præger samfundet, fx opfattelsen af mænds og kvinders roller og forholdet mellem kønnene, og hvordan normerne afspejles sprogligt, fx i form af ordforråd og sproglige forbindelser, og endelig det operationaliserede sprog i form af tekster som kan studeres empirisk af leksikografer og føre til beskrivelser i eksempelvis ordbøger. Man må følgelig skelne mellem 1) skævheder i samfundet, 2) skævheder i sproget, 3) skævheder i korpus og 4) skævhed i ordbøgerne. Der er naturligvis en vis sammenhæng mellem dem, idet samfundsnormer sætter sig sproglige spor som kan beskrives i ordbøger på grundlag af korpustekster, men forholdet mellem dem er ikke nødvendigvis enkelt og omsætningen ikke én-til-én.

Inspirationen til denne artikel kommer i høj grad fra Fjeld (2015), som påpeger kønsskævheder og stereotyper i nordiske ordbøger, heriblandt den ordbog jeg selv har været involveret i, DDO. Eftersom redaktionen i den seneste periode har haft fokus på revidering af artikelbestanden, har resultaterne i Fjeld (2015) været med til at anvise områder hvor ordbogen med fordel kan revideres, og det er især overvejelserne i den forbindelse der præsenteres her. Jeg vil derfor undersøge forholdet mellem korpuser og ordbøger og sammenligne norske og danske forhold. For norsk ser jeg på Leksikografisk bokmålskorpus, LBK, (Knudsen & Fjeld 2013) og NAOB, som i sin transformation fra NRO til NAOB har gjort brug af LBK, mens det danske materiale beskriver DDO og denne ordbogs underliggende korpus (det redaktionsinterne korpus *Bakspejlet* i korpusværktøjet CoREST, herefter omtalt som CoREST).

[3] KORPUS SOM ET SPEJL AF SPROGET – IDEAL OG ILLUSION

En ordbogs forankring i et korpus tjener til at sikre at beskrivelsen ikke begrænses til det sprog som – tilfældigvis – kendes og bruges af redaktørerne selv, men at det beskrevne sprog repræsenteres af en mangfoldighed af genrer og stillejer på tværs af generationer, geografi og sociale grupperinger. Korpusforankringen bliver derved garant for en systematisk og objektiv ordbeskrivelse, noget der i sagens natur er vigtigt for en ordbog der gerne vil forstås og forstå sig selv som videnskabelig.

I tilknytning hertil kan man stille sig to spørgsmål: I hvilken udstrækning giver det overhovedet mening at opfatte korpus som et spejl af sproget? Og hvad stiller man op hvis sproget og/eller korpus viser sig at indeholde træk der er modstrid med hvordan vi ønsker os at verden ser ud?

Udviklingen af korpusser er i høj grad relevant for svaret på det første spørgsmål. I DDO's periode har beskrivelsesgrundlaget udviklet sig enormt når det gælder kvantitet: Den trykte ordbogs korpus fra midten af 1990'erne var på ca. 40 mio. løbende ord og var statisk (Norling-Christensen & Asmussen 1998: 225). Tanken var at den trykte DDO skulle være et øjebliksbillede af sproget sådan som det var indfanget i korpusteksterne. Til grund for DDO på nettet ligger i dag et korpus der rummer over 1 milliard løbende ord og desuden er dynamisk. Det vokser måned for måned i takt med at nye tekster tilføjes, og på den måde kan nye ord og sproglig forandring opdages og registreres nærmest mens de finder sted. Prisen for kvantitet og dynamik kommer i form af repræsentativitet. Til den trykte ordbogs korpus blev der gjort meget ud af at indsamle tekster fra en lang række genrer og kanaler; foruden det professionelle sprog fra aviser, fag- og skønlitteratur også mere privat sprog i form af breve og dagbøger – samt op imod 20 % talesprog (jf. Asmussen & Norling-Christensen 1998: 228). I det moderne korpus udgør journalistiske tekster en uforholdsmæssig stor del på trods af et udbredt ønske om at inkludere flere teksttyper. Korpussammensætningen er ikke resultatet af en beslutning, men en simpel konsekvens af praktiske forhold. Det er ganske enkelt uforholdsmæssigt mere resursekrævende at få løbende adgang til andre teksttyper end journalistiske, især af ophavsretlige grunde.

Det er derfor i dag åbenlyst mere problematisk at anskue korpus som et spejl af sproget, og dermed undermineres naturligvis også en del af ordbogens videnskabelige forankring. Hvis et ord eller en betydning savnes i ordbogen, kan opdagelse i ordbogen ikke så let afgøres ved at konsultere forekomsttal i korpus. Det kunne jo tænkes at tilhøre en del af sproget som ikke er tilstrækkelig repræsenteret i korpusteksterne. I sin fulde konsekvens rummer ræsonnementet et alvorligt problem, for derved kan korpusevidens jo altid afvises med at korpus ikke er tilstrækkelig balanceret. Og hvor efterlader det videnskabeligheden?

Det andet spørgsmål ovenfor er principielt uafhængigt af det første, men ikke af den grund enkelt at adskille fra det i praksis. Hvis en ordbog viser en sproglig virkelighed som man er utilfreds med, skyldes det da manglende repræsentativitet i ordbogens beskrivelsesgrundlag, altså i korpus, skyldes det at sproget selv er ufuldkomment og fordomsfuldt, eller skyldes det fordomme hos den eller de redaktører der ved deres udvælgelse og beskrivelse er ansvarlige? Selvom det kan være vanskeligt at sige noget om de tre størrelses respektive bidrag til eventuelle skævheder, kan det være interessant at sammenligne korpusserne og de ordbogsbeskrivelser der baserer sig på dem. Jeg har derfor udvalgt en række kønsspecifikke ord og set på deres repræsentation i korpusser og ordbøger.

[4] UNDERSØGELSE AF KORPUSSE OG ORDBØGER

I tabel 1 vises en oversigt over ord der er udvalgt med henblik på at belyse sammenhængen mellem kønsskævheder i ordbøgernes beskrivelser og korpusrepræsentation, jf. oversigten for ordbøger i Fjeld (2015: 44)¹.

	CoREST	DDO	LB	NAOB
han	6.208.586	12.920	878.232	27.162
hans	1.293.143	3.148	115.025	15.590
ham	804.153	2.044	157.109	6.858
hun	2.278.888	7.150	486.466	14.565
hendes/hennes	422.456	1.061	60.931	1.371/2.449
hende/henne	332.338	1.551	109.979	1.850/2.336
mand/mann	702.736	2.368	51.762	2.742/4.118=6860
kvinde/kvinne	440.275	1.779	35.334	425/3.624=4049
dreng/gutt	118.976	469	11.181	80/937=1017
pige/jente	156.420	718	12.035	187/774=961

TABEL 1: udvalgte ordforekomster i ordbøger og korpusser

På samme måde som det er gjort i Fjeld (2015), kan man få et indtryk af skævheden i resurserne ved at se på kvotienterne for hvert kønnet ordpar, idet kvotienten 1 er udtryk for fuldstændig ligelig repræsentation, mens en kvotient over 1 er udtryk for dominans af hankønsord og en kvotient under 1 for dominans af hunkønsord. For nemheds skyld er formerne han/ham/hans og hun/hende/hendes slået sammen under ét, til han henholdsvis hun, og dreng/pige dækker også over gutt/jente i det norske materiale (og tilsvarende for mann/kvinne). Kvotienterne er sammenfattet i tabel 2.

	CoREST	DDO	LB	NAOB
kvotient han/hun	2,7	1,9	1,7	2,2
kvotient mand/kvinde	1,6	1,3	1,5	1,7
kvotient dreng/pige	0,7	0,7	0,9	1,1

TABEL 2: kvotienter for 3 ordpar i korpusser og ordbøger

Kvotienterne giver et omtrentligt billede af kønsskævhederne i materialet, men

[1] Der kan tælles på forskellige måder. For at få sammenlignelige tal har jeg her taget udgangspunkt i NAOB's fritekstfunktion, dvs. optællingerne er strengbaserede. Dog er der søgt på både ubestemt form singularis og pluralis for substantivernes vedkommende og ikke taget hensyn til store og små bogstaver (med de risici for skævheder som det indebærer, fx indgår også forekomster af propriet *Hans* i søgeresultatet for *hans*). Da NAOB er baseret på *Norsk Riksmålsordbok* og derfor rummer et stort antal ældre citater, har jeg for denne ordbogs vedkommende både søgt på gældende bokmålsformer (*mann, menn, kvinne, kvinner, hennes*) og på ældre riksmålsformer (*mand, mænd, kvinde, kvinder, hendes*). Optællingen er foretaget i august 2019.

må dog tages med visse forbehold. At *mand* dominerer for parret MAND/KVINDE kan ikke udelukkende tolkes som en kønsskævhed fordi alene det forhold at *mand* er et mere polysemt ord end *kvinde* gør at der må forventes relativt flere forekomster af *mand*. Eksempelvis dækker en del af forekomsterne betydningen ‘mandlig part i et ægteskab’, mens den tilsvarende betydning for den kvindelige part ofte er dækket af andre opslagsord end *kvinde* (fx *kone*, *hustru*, *viv*. Selvom NAOB har registreret denne betydning også for *kvinne*, er det næppe den mest udbredte betegnelse). Som nævnt i note 1) medfører den strengbaserede søgning desuden at også resultater for propriet *Hans* er talt med i resultatet for *hans*.

Uanset disse forbehold er det overordnede indtryk af tallene i de to tabeller klart at de maskuline former dominerer. Eneste undtagelse er DRENG/PIGE (herunder GUTT/JENTE), hvor den feminine form udviser svag dominans i begge resurser for dansk og i korpusmaterialet for norsk. Det er derudover betegnende at skævheden i det danske materiale er større i korpus end i ordbogen, mens det for norsk forholder sig omvendt. En mulig forklaring på det første kunne være at redaktørerne her har været bevidste om kønsproblematikken og derfor stiltiende har rettet en smule op på ubalancen (stiltiende fordi det ikke tidligere var et eksplicit indsatsområde)², mens det sidste vel især lader sig forklare ved at NAOB i udgangspunktet var en historisk ordbog med mange ældre citater og derved i et vist omfang afspejler en tid hvor forholdene mellem kønnene var anderledes. Dette må forventes at slå igennem, også selvom et stort revisionsarbejde har fundet sted.

Ellers er den mest slående forskel i materialet kvotienten HAN/HUN i henholdsvis det danske og norske korpus. Teoretisk set kan forskellen skyldes at de to sprog simpelt hen er forskellige med hensyn til brug af personligt pronomen i tredje person. Men i to så nærtbeslægtede sprog og samfund som det norske og det danske, er det næppe sandsynligt at det er hele, om overhovedet en del af forklaringen. Mere sandsynligt er det at forskellen i de to korpussers sammensætning slår igennem. Her er den mest oplagte forskel at Bokmålskorpus indeholder en større andel af sagprosa og skønlitteratur, mens de journalistiske tekster kun udgør ca. 10 %. Sidstnævnte er til gengæld overrepræsenteret i CoREST, hvor det udgør den dominerende del, sådan som det ofte er tilfældet med store korpuser på over 1 mia. løbende ord. Det er en interessant tanke hvis et bedre balanceret korpus i sig selv fører til en mere ligelig kønsrepræsentation. Om det virkelig er tilfældet, kræver dog yderligere undersøgelser.

[2] Også denne tolkning må tages med forbehold. DDO's korpus har ændret sig meget over tid, og sammensætningen af tekster i den trykte ordbogs korpus var væsentlig anderledes end det aktuelle korpus, der domineres af journalistiske tekster, jf. diskussionen nedenfor.

[5] REVISION AF DDO'S BETYDNINGSBESKRIVELSER

DDO's redaktion har inden for de senere år fået et stigende antal henvendelser fra brugere som er utilfredse med ordbogens semantiske beskrivelser i en række tilfælde. De opfordrer til at ændre forklaringerne så de bliver mere inkluderende og tidssvarende. Det betyder givetvis noget at mange af DDO's definitioner efterhånden er skrevet for over 20 år siden. Verden har ikke stået stille i den periode, og vi har i dag ikke det samme syn på fx seksuelle minoriteter og kønsrelationer som i 1990'erne – det gælder formentlig også for ordbogsredaktører. Hvad der tidligere var prototypisk, opfattes i dag af mange som stereotypisk – fordi det unødigt ekskluderer mindretalsgrupper der i forvejen har det svært. I redaktionen har vi derfor i den seneste periode især haft fokus på kønsladete ord og udtryk med henblik på at revidere en række ord så de fremtræder mere kønsneutrale.

Lad os se på et eksempel der viser problemet:

ægtepar substantiv intetkøn
 par bestående af en mand og en kvinde som er gift med hinanden

FIGUR 4: DDO: udsnit af artiklen *ægtepar*, før revision

Der er ikke tvivl om at DDO's artikel *ægtepar* beskriver det der forekommer statistisk mest hyppigt. Ifølge tal fra Danmarks Statistik³ består danske *ægtepar* i 2019 for 99,7 %'s vedkommende af en mand og en kvinde, og 98,5 % af de ægteskaber der blev indgået i 2018, var mellem en mand og en kvinde⁴. Alligevel adskiller NAOB sig på et afgørende punkt fra DDO med følgende forklaring:

ektepar substantiv
 par som har inngået ekteskab; gift par

FIGUR 5: NAOB: udsnit af artiklen *ektepar*

I NAOB's forklaring nævnes ikke kønnet på de personer der indgår ægteskab. Man kan hævde at DDO's forklaring er unødigt specificerende, og at der ingen grund er til at beskrive ægtefolkenes køn når man derved ekskluderer en hel gruppe personer – uanset om det statistisk set er typisk. Omvendt kan man kritisere NAOB for at underspecificere: Kun ved at følge linket til *par* kan brugeren få at vide at det par der refereres til, består af to personer.

I den reviderede version (november 2019) lyder forklaringen af *ægtepar* i DDO: "par bestående af to personer som er gift med hinanden". Et andet eksempel på

[3] <https://www.dst.dk/da/Statistik/emner/befolkning-og-valg/husstande-familier-boern/familier>

[4] <https://www.dst.dk/da/Statistik/emner/befolkning-og-valg/vielser-og-skilsmisser/vielser>

før- og efterversioner ses i figurene 6a og 6b.

svigersøn substantiv fælleskøn
mand som er gift med ens datter

FIGUR 6A: DDO: udsnit af artiklen *svigersøn* (før revision)

svigersøn substantiv fælleskøn
mand som er gift med ens datter eller søn

FIGUR 6B: DDO: udsnit af artiklen *svigersøn* (efter revision)

Eksemplet i figur 6 viser at den reviderede artikel også dækker situationer med ægtefolk af samme køn, og dette er desuden i artiklen illustreret med et citat.

Eksempler som disse er for så vidt uproblematisk som de uden videre kan ændres, med henvisning til at samfundets love og sprogsamfundets opfattelse har ændret sig. De hører med andre ord til en gruppe ord som det er helt rimeligt at revidere fordi samfundet har ændret sig. Den leksikografiske udfordring ved disse ord består mest i at finde frem til netop de rigtige kandidater og derefter foretage det fornødne.

[6] GRÆNSER FOR INKLUSION?

Mere problematisk er det hvis brugerne ønsker sig ændringer der ikke er forenelige med beskrivelsesgrundlaget, enten ved at kræve noget fjernet som er veldokumenteret i korpus, eller ved at kræve noget tilføjet som der ikke er sprogligt belæg for. Her bliver man som leksikograf udfordret og kan føle sig splittet mellem videnskabelige principper og personlige værdier.

To henvendelser til DDO's redaktion har haft denne karakter, og begge medførte heftige offentlige debatter om sproget og ordbogens rolle. Den første sag var foranlediget af en opdatering hvor ordet *grønlænderstiv* blev optaget i DDO. Det blev hurtigt opfanget af danske journalister, og det varede ikke længe før grønlandske brugere tog afstand fra ordet og krævede det fjernet fra ordbogen. Historien fik sit eget liv i Grønland, hvor det af mange blev udlagt som et udslag af dansk racisme. Sagen nåede op på øverste politiske niveau da den grønlandske kulturminister sendte et brev til sin danske kollega hvori hun tog kraftigt afstand fra ordet og optagelsen i ordbogen, som hun kaldte en blåstempling af en racistisk betegnelse⁵.

En lignende debat udspillede sig ved en anden opdatering af ordbogen, hvor det var ordet *hudfarvet* der kom i centrum. En gruppe brugere på et socialt medie

[5] <https://naalakkersuisut.gl/da/Naalakkersuisut/Nyheder/2017/11/161117-Groenlaenderstiv-er-nedladende-og-generaliserende>. Se desuden en kritik af DDO's beskrivelse i Farø og Jensen 2018: 219)

iværksatte en kampagne for at få ændret ordbogens forklaring, og sagen blev grundigt dækket af den danske presse. Det der faldt brugerne for brystet, var at der i DDO kun var angivet forklaringen ”af farve som huden hos gruppen af hvide mennesker; meget lys og afdæmpet orange”, men ikke betydningen ”af samme farve som ens hud”. Det anså gruppen som et udtryk for ”strukturel racisme”⁶ og krævede sidstnævnte betydningsforklaring tilføjet.

I første omgang førte brugerkravene og presseomtalen ikke til ændringer, med den begrundelse at vi ikke kunne finde nogen entydige belæg for beslutningen i vores korpusgrundlag. Men sagerne gjorde indtryk og påmindede os om vores samfundsrolle og ansvar. Efter ihærdig søgning i kilder uden for ordbogens korpus lykkedes det også at finde nogle eksempler på den nye brug, og det – sammenholdt med det faktum at mange korpusbelæg jo ikke lader sig afkode med sikkerhed – førte til at den nye betydning faktisk blev tilføjet ved en opdatering i november 2019. Redaktionsinternt var der ikke fuldstændig enighed om beslutningen. Uenigheden har især drejet sig om hvor meget eller hvor lidt der skal til for at ordbogen fraviger sin korpusforankring: Skal man alligevel tage noget med hvis der (overhovedet) findes belæg på det uden for korpus, eller hvis man selv mener at kende ordet eller betydningen? Og kan man omvendt tillade sig at undertrykke ord og betydninger der findes i korpus?

[7] IDEELLE FORDRINGER – I SPROGET OG I ORDBØGERNE

Hvis der er et redaktionelt ønske om at rette op på en skæv kønsrepræsentation i ordbøgerne, må man først spørge sig hvordan det ideelt set bedst gøres? Bør man ved udvælgelse af citater til ordbogen tilstræbe en ligelig kønsfordeling, fx i fordelingen af ord som de undersøgte HAN/HUN, MAND/KVINDE OG DRENG/PIGE, eller bør fordelingen snarere afspejle sproget, dvs. også hvis dette udviser en kønsskævhed, eller korpus? Uden at skrive det eksplicit synes Fjeld (2015) at mene at en ligelig fordeling er det ideelle⁷. Jeg er tilbøjelig til at dele synspunktet, men mest ud fra en antagelse om at fordelingen også er nogenlunde ligelig i sproget. Jeg er dog ikke bekendt med undersøgelser som dokumenterer om dette faktisk er tilfældet – og ved nærmere eftertanke kan jeg godt finde på argumenter for det modsatte. Selvom det er rigtigt at en ordbog er et vindue til den kultur den er lavet i (Fjeld 2015: 40), er en ordbogs opgave først og fremmest at beskrive sproget, ikke samfundet eller kulturen selv. En vigtig pointe i denne sammenhæng er dog at uanset om idealfordringen er beskrivelse af sproget, er det i prak-

[6] <https://www.berlingske.dk/kommentarer/den-danske-ordbogs-definition-af-hudfarvet-er-racistisk>

[7] Jf. Fjeld (2015): ”Kvotienten i Oxford English Dictionary fra 1987 er 3,75 og skiller seg dermed svært negativt ut i forhold til de nordiske ordbøkene” (p. 44).

sis sproget i korpus der beskrives. Korpus er jo leksikografens praktiske operationalisering af den mere luftige og abstrakte størrelse som et sprog er, og som sådan nødvendig for en systematisk sproglig beskrivelse og dokumentation.

En lignende ideel fordring kan stilles til ordforrådet selv, idet det også rummer vidnesbyrd om kulturelle værdier: Mange erhvervsbetegnelser vidner om en stereotyp opfattelse af kønnenes rollefordeling i arbejdslivet: *skraldemand, flyttemand, landmand, husbestyrerinde, lærerinde, brodøse, husholderske, syerske*. Ud fra et samfundsmæssigt ligestillingsperspektiv kunne de kønsladete ord med fordel erstattes af neutrale betegnelser, men det er næppe ordbøgernes opgave at gøre det så længe de oprindelige ord er i brug i sprogsamfundet. Ordbøgernes rolle er primært dokumenterende og beskrivende.

Det betyder ikke at ordbøgerne skal forholde sig passive og afvente at sprogsamfundet ændrer opfattelse. Redaktionen kan aktivt gøre noget for at være en del af udviklingen ved at sørge for at der i korpus også findes tekster der bruger andre betegnelser og værdier end de traditionelle (fx tekster om og af forskellige minoritetsgrupper eller kønspolitiske tekster), ved at optage alternative betegnelser og redegøre kontrastivt for forskelle i valør og konnotation mellem ordene og ved målrettet at revidere eksisterende artikler så tidligere tiders stereotype forestillinger ikke får lov at leve videre længere end højst nødvendigt.

[8] KONKRETE REVIDERINGSOPGAVER

Ud over de eksempler der allerede er nævnt ovenfor, findes der en lang række af ordpar som kan gennemgås med henblik på kontrastiv beskrivelse: *kassedame/kasseassistent, tillidsmand/tillidsrepræsentant, husholderske/husholder, sangerinde/sanger, violinistinde/violinist, portnerkone/portner* osv. I disse ordpar er det kønsforskellene der er i fokus, men det er nemt at udvide listen med mange andre ord og ordpar fra andre områder: seksuelle og etniske minoriteter, religion, sygdom og andre værdiladete områder: *transseksuel/transkønnet, sigøjner/roma, muhammedaner/muslim, mongolisme/Downs syndrom, åndssvag/udviklingshæmmet* osv. Det kan ses at problematikken vedrørende kønsskævhed er nært beslægtet med diskussionen af politisk korrekthed. Begge dele handler om at vores normer er under forandring, og det er ikke ligegyldigt hvilke ord vi bruger. Ord fremkalder mentale billeder hos sprogbrugerne, hvilket på den ene side er kommunikativt hensigtsmæssigt fordi vi derved har kognitivt lettere ved at processere dem, men på den anden side er uheldigt hvis billederne er belastede af forældede konnotationer, fordomme og stereotype forestillinger. I den situation overvejer vi i redaktionen om det er nødvendigt at udvide beskrivelsesapparatet hvis det skal være muligt at vejlede ordentligt om sprogbrugen. Det eksisterende inventar af

sprogbrugsmarkeringer har især gjort brug af “nedsættende” og “uformelt”. Betegnelsen “uformelt” er dog hverken tilstrækkelig stærk eller præcis til at dække diskriminerende sprogbrug når markeringen samtidig også bruges om mere hverdags sproglige, talesprogsprægede og ufarlige ord som *bælgmørk*, *grovæde*, *hundefryse* og *hængerøv*. Betegnelsen “nedsættende” er mest velegnet når både afsender og modtager er klar over ordets negative klang: *borgerdyr*, *flæskebjerg*, *kulturpave* og *nasserøv* er eksempler fra DDO. Men når normer er under forandring, er der ofte netop ikke enighed om ordenes valør, og det bliver da en vigtig opgave for ordbogen at vejlede om ordenes virkning, herunder om de kan virke stødende eller krænkende på modtageren, uanset hvilken intention afsenderen måtte have haft. Det er især vigtigt i forbindelse med brug af ordbogen til produktion. DDO’s redaktion har derfor besluttet at indføre betegnelsen “kan virke stødende” for at gøre denne forskel klar. Ved særlig problematiske ord eller ord-felter indføres der også mulighed for en udvidet kommentar hvor redaktørerne kan redegøre for en historisk udvikling (beslægtet med de ”stilrutor” som findes i *Svensk ordbok*, fx ved ordet *neger*) eller behandle nærtbeslægtede ord sammen og udrede forskelle i betydning og sprogbrug mellem dem.

Et andet indsatsområde er unødigt kønnede opslagsord. De minder om de tidligere nævnte ord med overspecificerende definitioner, men med den forskel at opslagsordet selv indeholder et kønnet ord. I figur 7 vises et eksempel.

narkobaron substantiv, fælleskøn
mand der er kommet til rigdom og magt ved narkohandel i stor stil (ofte nedsættende)

FIGUR 7A: DDO: udsnit af artiklen *narkobaron*

Baron er som bekendt en titel der kun gives til mænd, og det forhold kan derfor skygge for at bruge sammensætninger med *-baron* om andet end mænd. Desuden gør de sociale roller i samfundet at virkelighedens narkotikaverden er domineret af mandlige kriminelle (med nogle få undtagelser som Griselda “Queen of Cocain” Blanco), hvorfor beretninger om narkobaroner altovervejende handler om mænd. Alligevel har NAOB valgt en anden løsning:

narkobaron substantiv
NEDSETTENDE person som tjener seg rik på storsalg av narkotika

FIGUR 7B: NAOB: udsnit af artiklen *narkobaron*

Der er næppe forskel i sproglig evidens mellem norsk og dansk i dette tilfælde, og det mulige alternativ, at optage opslagsordet *narkobaronesse*, er ikke reelt da det ord slet ikke forekommer i hverken LBK eller i CoREST. Selvom DDO måske er tættere på at beskrive det statistisk mest sandsynlige scenarie, er det ikke

nødvendigvis den bedste løsning fordi den cementerer en traditionel kønsopfattelse. Det kan være bedre at gøre som NAOB og underspecificere for at lade definitionen være kønsneutral så vidt det er muligt.

På tilsvarende vis kan andre opslagsord med kønnede sidsteled gennemgås, navnlig sammensætninger med *-mand*. Nogle sammensætninger er symmetriske på den måde at der findes både kvindelige og kønsneutrale pendants (*idrætsmand, idrætskvinde, idrætsudøver; tillidsmand, tillidskvinde, tillidsrepræsentant*), og i de tilfælde er en kønnet definition som regel på sin plads. Men der er også sammensætninger der helt savner kvindelige og/eller neutrale modsvarigheder, og her bør man se på om en revision er påkrævet. Det gælder ord som *handelsmand, havemand, pistolmand, postmand* og *pølsemand*, ord der alle er defineret med "mand" i definitionen i den nuværende DDO. En eventuel revision kan bestå i at definere mindre specifikt eller ved at undersøge om en sammensætning med *-kvinde* eller en neutral betegnelse er berettiget som opslagsord. Det må bero på et skøn i hvert enkelt tilfælde om revision er påkrævet, og hvilken løsning der i givet fald fungerer bedst. Med til revisionen hører naturligvis også et kritisk blik på de valgte eksempelsætninger.

På tilsvarende måde kan ukønnede opslagsord med kønnede definitioner tjekkes for at undersøge om forklaringen er overspecificerende. Det gælder ord som *bejler, beundrer, bonde, forfører, frier, hjerteknuser, homo, kleppert, kontormus, mannequin, montør, muskelbundet og tilbeder*, der alle i DDO er defineret med formuleringer som "mand" eller "person, især mand" eller tilsvarende med "kvinde". Ofte kan man med fordel bruge en mere kønsneutral formulering og supplere med citater der viser en ikke-stereotyp brug.

I revideringsarbejdet må man heller ikke være blind for egne fordomme. En brugerhenvendelse kan tjene som illustration. Brugeren var utilfreds med DDO's forklaring på *voldtægtsforbryder*, der i DDO var defineret som "person (mand) der har begået voldtægt". Umiddelbart kan man synes at brugeren måske var lidt for sart, for det er vel næsten altid mænd der begår voldtægt? Nej, helt så enkel (og for mig var det overraskende) er virkeligheden tilsyneladende ikke. Ifølge en artikel, ganske vist om amerikanske forhold, udgør mænd næsten 40 % af ofrene for sexforbrydelser⁸. Det var også overraskende for mig at der står kvinder bag 70 til 80 % af sexforbrydelser begået mod mænd (i tilfælde hvor overgrebet ikke involverer penetration). Hvis man alene ser på fremstillingen i korpus, er der mindre tvivl. Her er 99 % af de voldtægtsforbrydere der omtales, mænd; kun ganske få eksempler på andet ud af ca. 1000 er det lykkedes mig at finde. Måske er sexforbrydelser mod mænd slet ikke så sjældne, men til gengæld så meget mere

[8] <https://www.theatlantic.com/science/archive/2016/11/the-understudied-female-sexual-predator/503492/>

tabubelagte at de ikke omtales? Resultatet blev i dette tilfælde at artiklen *voldtægtsforbryder* nu er revideret så der kun står ”person” – i øvrigt i lighed med forklaringen i NAOB.

[9] KONKLUSION

Køns- og andre skævheder er svære at undgå helt i ordbøger fordi sproget er præget af inertie; ord og udtryk nedarves fra ældre tider, og undertiden bærer de tidligere tiders normer og syn på verden med sig. Talemåder og faste udtryk er kendte eksempler (*være herre i eget hus, mande sig op, have hår på brystet*), men også ordene selv afslører fortidens syn (*drengepige* og *tøsedreng*). Ud fra et moderne kønspolitisk perspektiv kan det fx være rimeligt at have et kønsneutralt personligt pronomener, men det betyder ikke at det er nemt at indføre det. I Sverige ser det ud til at være ved at lykkes med *hen*, men hverken i Danmark eller Norge ser det ud til at være nært forestående. Og sådan er det med mange ord i ordforrådet. Hvorfor har vi kun en *havemand* eller en *garderobedame*? Ordene afspejler ældre tiders samfund og rollefordeling, og det tager tid at ændre dem.

Omvendt er der netop grund til at være særlig opmærksom på disse ord så stereotype forestillinger ikke bliver hængende i sproget længere end højst nødvendigt. Redaktionen bør have blik for kønsbalancen og minoritetssprog ved sammensætning af korpustekster, ved lemmasektion, ved udformning af definitioner og ved udvalg af citater og andre sproglige eksempler. Behandlingen af ideologisk ladede ord har peget på at en udtømmende semantisk beskrivelse næppe er mulig uafhængigt af de værdier der præger det omgivende samfund på et givet tidspunkt. Selv ved tilsyneladende neutrale ord kan der være konnotationer der kræver beskrivelse i form af fx sprogbrugsmarkører eller kollokationer.

Personlig tror jeg at fuldstændig kvantitativ egalitet betyder mindre for ligestillingen end forklaringer og ikke mindst indholdet i eksempelsætningerne. Her bør redaktionen til gengæld skærpe blikket og forsøge at undgå de værste stereotyper. På DDO er arbejdet med det begyndt, ikke mindst takket være Ruth V. Fjelds artikel fra 2015.

LITTERATUR

CoREST. <https://korpus.dsl.dk/corest>, set august 2019.

Den Danske Ordbog. <https://ordnet.dk/ddo>, set august 2019.

Det Norske Akademis ordbok. <https://www.naob.no>, set august 2019.

Farø, Ken & Lene Vinther Jensen. 2018. Den Danske Ordbog på nettet – en undersøgelse af version 3.0. I Sturla Berg-Olsen & Emma Sköldberg (red.), *LexicoNordica · 2015 – Fackspråk i nordiska ordböcker*. Nordisk förening for lexikografi, 215-232.

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

Knudsen, Rune Lain & Ruth Vatvedt Fjeld. 2013. LBK2013. A Lexicographic Corpus for Norwegian. I Deny A. Kwary, Nur Wulan & Lilla Musyahda (red.), *ASIALEX: Lexicography and Dictionaries in the Information Age*, Bali, Indonesia 20-22 August 2013, 361-362.

Leksikografisk bokmålskorpus. <https://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/lbk/>, set august 2019.

Moon, Rosamund. 2014. Meanings, Ideologies, and Learners' Dictionaries. I Andrea Abel, Chiara Vettori & Natascia Ralli (red.), *Proceedings of the XVI EURALEX International Congress: The User in Focus*. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism, 85-105.

Nilstun, Carina. 2012. Fra dype betydningshierarkier til flatere struktur – et innblikk i revisjonen av Norsk Riksmålsordbok. I Henrik Lorentzen & Emma Sköldberg (red.), *LexicoNordica 19 · 2012 – Betydelsesbeskrivningar i nordiska ordböcker*. Nordisk förening for lexikografi, 123-140.

Norling-Christensen, Ole & Jørg Asmussen. 1998. The Corpus of The Danish Dictionary. I *Lexikos. Afrilex Series 8*, 223-242.

Nossem, Eva. 2018. Queering Lexicography – Balancing Power Relations in Dictionaries. I Brian James Baer & Klaus Kaindl (red.), *Queering Translation, Translating the Queer. Theory, Practice, Activism*. Routledge, New York, 172-187.

Russell, Lindsay Rose. 2012. This is What a Dictionary Looks Like: The Lexicographical Contributions of Feminist Dictionaries. I *International Journal of Lexicography*, 25(1). Oxford University Press, 1-29.

Svensk ordbok utgiven av Svenska Akademien 2009. <https://svenska.se>, set august 2019.

Trap-Jensen, Lars. 2014. Leksikografisk tradition og fornyelse: tre revolutioner på 100 år? I Ruth Vatvedt Fjeld & Marit Hovdenak (red.), *Nordiske Studier i Leksikografi 12, Rapport fra Konferanse om leksikografi i Norden*, Oslo 13.-16. august 2013, 42-68.

KONTAKT

Lars Trap-Jensen
Det Danske Sprog- og Litteraturselskab
ltj@dsl.dk

FRÅ ORDBOK TIL KORPUS – OG TILBAKE?

BJØRNETJENESTE OG FORFORDELE

LARS S. VIKØR

SAMANDRAG

Denne artikkelen drøftar forholdet mellom deskriptivitet og normativitet i norske ordbøker på grunnlag av to ord som har fått nye tydingar i moderne tid i forhold til dei tradisjonelle normene: *bjørneteneste* og *forfordele*. Bruken av desse orda i moderne norsk skriftspråk blir undersøkt på grunnlag av søk i to tekstkorpus. Det viser seg at *forfordele* blir brukt med heilt motsette tydingar i vanleg skriftspråk. Det blir antyda at ordbøkene bør revidere tydings- og bruksbeskrivinga av dette ordet i samsvar med det mønsteret bruken viser.

[1] INNLEIING

Utgangspunktet for denne vesle undersøkinga er at ordbøker pr. definisjon er både normative og deskriptive, mens eit tekstkorpus pr. definisjon berre er deskriptivt. Visse typar av ordbøker kan rett nok seiast å vere reint normative, men dei skal eg ikkje kome inn på her. Eg tenker på dei store allmennordbøkene som inneheld både formopplysningar og definisjonar. Formopplysningane er da prinsipielt normative – definisjonane prinsipielt deskriptive. Det er dette siste eg skal ta utgangspunkt i no.

Startpoenget er da at om vi først konsulterer den *prinsipielt deskriptive* definisjonen til eit oppslagsord i ordboka, og så går til det *reelt deskriptive* korpuset og studerer bruken av ordet slik den framtrer der, så bør vi finne ein stor grad av samsvar. Dersom vi ikkje gjer det, og altså finn ut at bruken av ordet ikkje er i samsvar med definisjonen i ordboka, kan vi reagere på ulike måtar, om vi altså er vanlege ”brukarar” av ordboka og korpuset: Vi kan slå fast at mange bruker ordet feil, ut ifrå ei overtyding om at det er ordboka sin definisjon som er «rett». Eller vi kan dra i tvil kva som er rett og sende inn eit spørsmål til Språkrådet eller Språkteigen. Eller vi kan akseptere begge bruksmåtene og trekke på skuldrene: ”Jaja, språket utviklar seg jo.” Den siste reaksjonsmåten er nok den sjeldnaste, mest begrensa til lingvistar.

Er ein leksikograf, og ansvarleg for redigering eller oppdatering av ei ordbok, står ein i ein annan situasjon. Det er dette eg har markert med ”og tilbake” med spørsmålsteikn i tittelen. Skal innsikta frå korpuset reflekterast i ein revisjon av

ordboksartikkelen? Eller uttrykt deskriptivt: Blir det gjort? Eg vil eksemplifisere denne situasjonen ved å sjå nærare på to ord som aktualiserer dette dilemmaet.

I innføringsboka *Ord og ordbøker* er orda *bjørneteneste* og *forfordele* nemnt som døme på ord som i bruken har fått den stikk motsette tydinga av den opphavlege, samtidig som den opphavlege tydinga også lever vidare i beste velgåande (Fjeld og Vikør 2008: 103). Under arbeidet med det seminarinnlegget som ligg til grunn for denne artikkelen (våren 2018), søkte eg på desse orda i Leksikografisk Bokmålskorpus (heretter: LBK, jf. Knudsen og Fjeld 2013) og samanlikna med definisjonane eg fann i dei store allmennordbøkene for bokmål: *Bokmålsordboka* (BOB), *Norsk Riksmålsordbok* (NRO) og *NAOB*, og til jamføring også i dei tilsvarande korpusa og ordbøkene for nynorsk (spesifisert nedanfor). Korleis er forholdet mellom ordbøkene og korpusa her? Det som følgjer i denne artikkelen, baserer seg på desse søkeresultata frå 2018. Sidan bokmål er hovudfokuset i denne artikkelen, bruker eg *bjørnetjeneste* som referanseform for dette ordet.

[2] BJØRNETJENESTE

Bjørnetjeneste kan eg gjere meg fort ferdig med. BOB definerer det slik: «velment, men uheldig hjelp som blir mest til skade» (med forklaring av opphavet i parentes: «etter La Fontaines fabel om bjørnen som skulle jage bort en flue av sin herres nese, men som i stedet knuste hodet hans»). NRO har ein tilsvarande definisjon med same opphavsforklaringa, og det same gjeld begge nynorskordbøkene (*Nynorskordboka* og *Norsk Ordbok*). Den nyaste ordboka, *NAOB*, har derimot to definisjonar: "1 velment, men klossete hjelp som blir mest til skade. 2 NYERE BRUK, ANSETT SOM UHELDIG stor tjeneste". Her er det fire sitatbelegg på tyding 1, ingen på tyding 2.

I LBK er det 52 belegg på ordet, og ein gjennomgang gir eit heilt eintydig bilde, såpass at eg ikkje treng eksemplifisere og drøfte det vidare: Alle belegga følgjer den tradisjonelle definisjonen. Den dominerer altså totalt blant dei skrivande menneska som er representert i LBK; kanskje alle har høyrte La Fontaines fabel, for dei som kjenner den, vil nok ha vanskeleg for å bruke *bjørneteneste* som noko positivt. Her vil eg altså gi ordbøkene "godkjent" i måten å framstille dette ordet på – ut ifrå vanleg skriftleg standardspråkleg bruk.

[3] FORFORDELE

Resten av denne artikkelen skal handle om verbet *forfordele*, som er meir intrikat på fleire måtar. Eg startar med ordboksdefinisjonane av dette ordet, og går så inn i LBK, der ordet har 37 belegg. Det er få nok til at eg kan analysere belegga eitt for eitt på den begrensa plassen eg har her, og mange nok til at vi kan skimte eit mønster i dei ulike tydingane og bruksmåtene.

[3.1] *Verbet sine tre tydingar*

Først altså definisjonane i dei tre sentrale bokmålsordbøkene:

Norsk Riksmålsordbok: «**1**) ved deling gi (en) mindre enn det som han har krav på; gi (en) mindre enn andre. **2**) fam., gi (en) for meget»

Bokmålsordboka: «gi en mindre enn en har rett på ved fordeling»

NAOB: «**1** ved fordeling gi (noen) mindre enn det vedkommende har krav på; gi (noen) mindre enn andre. **2** UHELDIG, MUNTLLIG gi (noen) for mye». (Her er det fire sitat på tyding 1 av Hamsun, Olaf Bull, Jostein Gaarder og Bror Hagemann, ingen på tyding 2.)

Sjølv om LBK skal vere utgangspunktet for denne artikkelen, kjem eg også til å dra inn nynorskmateriale til jamføring. Derfor siterer eg også dei to sentrale nynorskordbøkene:

Nynorskordboka: «gje (ein) mindre enn ein har krav på ved fordeling; fordele urettvist; misbyte»

Norsk Ordbok: «**1** **a**) (ved deling) gje (ein) mindre enn det han har krav på; gje (ein) mindre enn andre; misbyta. **b**) (sj) gje (ein) for mykje (i høve til andre). **2**) fordela (noko) ulikt, urettferdig; misbyta». (Det blir gitt eitt sitatdøme frå Syn og Segn 1967 på tyding 1a, ikkje noko på 1b, og eitt frå Dag og Tid 1975 på tyding 2.)

Her har vi altså tre tydingar, som vi kan kategorisere slik: 1) ei *negativ*: 'gi nokon for lite (i forhold til andre)'; 2) ei *positiv*: 'gi nokon for mykje (i forhold til andre)'; 3) ei *nøytral*: 'skeivfordele (på begge måtar)'. Vi merkar oss at den negative tydinga blir rekna som den mest tradisjonelle og vanlege, altså umarkerte, mens den positive blir markert som «fam(iliær)», «uheldig, muntlig» eller «sj(eldsynt)», eller rett og slett utelaten. Vi merkar oss òg at den nøytrale tydinga berre er representert i nynorskordbøkene.

Det er ikkje alltid så enkelt å bestemme kva for ein tydingskategori kvart belegg høyrer inn under, for det kan trengast mykje kontekst for å gjere det klart. Eg refererer no eit døme på kvar kategori der det går nokså tydeleg fram av sjølve sitatet. Først den negative tydinga:

En insolvent skyldner som begunstiger enkelte fordringshavere, forfordeler de øvrige.

Så den positive:

Det er en metode som tillater å forfordele det kjønnnet det er for lite av i en rekrutteringsprosess.

Dette er den enklaste typen å kategorisere – når det i den umiddelbare konteksten blir antyda ei motsetning til den forfordelte parten, gjennom verbet «be-gunstige» vs. tilstandsbeskrivinga «det er for lite av» i døma ovanfor.

Eit døme på den nøytrale bruken 'skeivfordele' har vi her:

Utposteringen av skytterne følger også prinsippet om at ingen skal forfordeles på noen måte og alle har lik sjanse til å felle viltet.

Her er det openbert at «forfordeles» omfattar både dei som får ein for god posi-sjon og dei som får ein for dårleg, jf. òg følgjande døme:

Gilligan hevder at Kohlberg i sin vurdering av svarene forfordeler de to respondentene.

Det er openbert at når det berre er snakk om to partar, og begge blir forfordelte, må det nødvendigvis bety at den eine får for god behandling og den andre for dårleg.

Det er ikkje alltid at tydinga av verbet går så tydeleg fram som i desse sitata. Følgjande døme, som ikkje er frå LBK, kunne tolkast både positivt og negativt:

Ny språklov styrker norsk, men forfordeler bokmålsbrukere

Men reelt er nok denne også heilt eintydig: I ingressen til den påfølgjande teksta står det at lova (eller lovutkastet, som det er snakk om her) «favoriserer nynorsk på bekostning av flertallsmålformen», altså bokmålet. Og sidan dette står på nettsidene til Riksmålsforbundet, vil nok overskrifta som er sitert ovanfor, uan-sett vere heilt eintydig for lesaren.

Den følgjande setninga (frå LKB) er det derimot knapt mogleg å tolke i isola-sjon:

Med årets øyafestival som det fremste skrekkeksempelet hevdes det her at kvinner for-fordeles.

Her er det ikkje anna å gjere enn å leite rundt i konteksten for å finne ut om den som hevdar dette (usynleggjort gjennom passivforma «hevdes»), meiner at kvin-nene får for stort rom eller for lite. I dette tilfellet er det klart at «forfordeles» skal bety «favoriseres», altså heilt motsett Riksmålsforbundets praksis ovanfor. Det er ein feministisk artikkelforfattar som kritiserer ein kommentator som tar

opp emnet «booking av kvinnelige band». Det er denne kommentatoren som hevdar at «kvinner forfordes», altså får for stort rom. I artikkelforfattaren sine auge framstår dermed kommentatoren som ein representant for tradisjonelle, altså kvinnediskriminerande, haldningar i musikkmiljøet. Artikkelforfattaren kontrar: «Hvorfor skal man betvile øyafestivalens bookingansvarlige Claes Olsen når han sier: ‘Vi booker band basert på kvalitet, ikke fordi de er enten gutter eller jenter.’»

Nokre fleire dømme på «forfordele» brukt negativt vs. positivt vs. nøytralt:

Negativ tyding:

Grunnen til at faren ønsket å forfordele Esther, var at han hadde mer tiltro til sønnen og misbilliget datterens livsførsel.

Å appellere til toleranse kan derfor bli å anmode om å godta at grupper forfordeles.

Selv om et land er et demokrati, forhindrer ikke det at landets minoriteter blir forfordelt.

I Sør-Afrika forfordelte apartheidregimet det svarte flertallet og begunstiget det hvite mindretallets vannforbruk.

Dommer Lewis F. Powel jr. erklærte, på vegne av flertallet, at saksøkerne ikke hadde klare bevis for at det var en forfordelt gruppe av fattige elever i dette skoledistriktet, og at Grunnloven ikke nevner utdanning som en grunnleggende rettighet.

Og du forfordelte virkelig ikke, du var raus.

Positiv tyding:

Det må foreligge garantier for at polets egne flasker i hyllene ikke blir forfordelt når importmonopolet av vin og brennevin faller.

Det har vært en utbredt oppfatning blant norske leverandørbedrifter at britiske OSO har forfordelt nasjonale leverandører.

De to viktigste endringene for konsesjonssystemet var at nasjonalitet ikke kunne brukes som et kriterium så lenge søkeren hørte til EØS-området, og at norske myndigheter ikke kunne forfordele søkere som ville bruke norske varer, tjenester og arbeidskraft.

Det betyr at en forsømt sektor må forfordeles i en tid fremover.

Det ville ha vært svært uheldig for oss å ha sittende på oss at vi forfordelte atlantakan- didaturet fremfor de øvrige.

Vi har en valgdekning som forfordeler visse partier og visse ledere, etter kriterier som først og fremst er fokusert mot å maksimere seertall.

Nøytral tyding:

Fornyingen av 329 konsesjoner uten anbudsrunde var til hinder for konkurransen og kontrollen med at ingen tilbydere forfordeles.

Store summer forfordeles også via Norge 2005, som eies av Kulturdepartementet.

[3.2] *Objektets semantiske roller*

Både subjekt og objekt kan ha ulike semantiske roller. Dei tre viktigaste rollene er *agens*, *patiens* og *recipiens* (Faarlund, Lie og Vannebo 1997: 47-49). I aktive setningar er subjektet oftast *agens*, det direkte objektet *patiens* eller *recipiens*, og det indirekte objektet *recipiens* (når ei setning har både direkte og indirekte objekt). Da står *recipiens* for den personen eller parten (som ikkje treng vere ein person) som får eller ikkje får den aktuelle fordelten eller ressursen som er emnet for setninga. I alle dei døma eg har sitert ovanfor, er objektet *recipiens*.

Patiens står derimot for den storleiken som blir fordelt på dei aktuelle partane. Eg gir eit døme på dette:

Vi kjenner dem [hersketeknikkane] alle sammen ... fra de lett synlige som mobbing/erting og latterliggjøring til de mer subtile som å sette kvinnene til å lage vaktlister, holde tilbake informasjon og forfordele viktige og interessante arbeidsoppgaver.

I passivsetningar er subjektet *patiens* eller *recipiens*, mens *agens*rolla blir tildelt eit preposisjonsuttrykk med *av* eller *sløyfa* heilt. Døme:

Maten blir forfordelt av partene i konflikten.

Derimot ville eiendom være omfordelt, men fortsatt være forfordelt.

Kjærligheten er urettferdig og forfordelt.

I alle desse døma er subjektet *patiens*.

Ein kan òg la recipiensrolla bli uttrykt i eit preposisjonsuttrykk, typisk med *til*:

I dag forfordeles ikke fisken til direkte anvendelse for kystbefolkningen, men til de som skal tjene penger på ressursene.

Fordelinga av semantiske roller til objektet (og til subjektet i passivsetningar) går på tvers av skilja mellom dei tre tydingane av verbet, men det ser ut til å vere eit hovudmønster: Den positive og den negative tydinga av verbet involverer for det meste recipiens-objekt, mens verbet i dei fleste av døma med patiens-objekt har den nøytrale tydinga 'skeivfordele'. Men vi har unntak: I det siste dømet, der subjektet i passivsetninga, fisken, er patiens, ser vi at det er den positive tydinga av verbet det handlar om. Og i dei to døma på den nøytrale tydinga lenger oppe som begynner på *Utposteringen* og *Gilligan*, er det recipiens-rolla som er inne i bildet. Det kan, som antyda før, skje når eit kollektiv av «recipients» omfattar partar som både kjem (eller kan kome) ut på den positive og den negative sida ved skeivfordelinga.

Ein meir spesiell situasjon får vi når verbet blir brukt i presens partisipp og dermed transformert til eit adjektiv. Da fell dei semantiske rollene bort, og det blir berre fokusert på sjølve handlinga. «Forfordelende» blir da ein eigenskap ved substantivet og ikkje ei handling i og for seg; det er tvert imot (verbal)substantivet som uttrykker handlinga. Men i konteksten kan objektet likevel bli spesifisert:

De utbetalinger som B. R. foretok ... må anses som en illojal og forfordelende handling i forhold til de to eneste kreditorer som ikke fikk oppgjør.

Eit problematisk eksempel til slutt:

Ulikhetene er systematisk forfordelt, og i stor grad blir de overført fra en generasjon til en annen.

Dette gjeld sosiale forskjellar i utdanningssystemet, der elevar med lågt utdanna foreldre og minoritetsspråklege elevar systematisk kjem dårlegare ut enn dei meir ressurssterke. Det logiske innhaldet i setninga er for så vidt klart nok i samanhengen, men reint syntaktisk kan ein freistast til å meine at setninga bit seg sjølv i halen – i og med at «ulikhetene» allereie er innebygd i verbet *forfordele*, og ingen patiens (som må vere ekstern i forhold til verbet) ut over det er nemnt.

[3.3] *Belegga*

No er tida inne til å telje belegg. Med 37 belegg er det klart at vi må ha alle moglege atterhald til dei tala vi får fram, men ein hypotese kan dei i alle fall gi grunnlag for. Eg set opp dei tre tydingane av verbet først:

Negativ tyding (ein får for lite): 15

Nøytral tyding (skeivfordeling): 12

Positiv tyding (ein får for mykje): 10

Dette er språkbruk som for det meste er yngre enn 2000, i eit par tilfelle går belegga tilbake til 1980-talet. Ei tilsvarande teljing av dei semantiske rollene til objektet gir dette resultatet:

Recipiens: 27

Patiens: 8

Ikkje noko objekt (presens partisipp av verbet): 2

Kryssar ein så kriteria, finn ein at seks av patiens-objekta knyter seg til den nøytrale tydinga, eitt kvar til dei to andre. Dei to døma på bruk av presens partisipp knyter seg til den negative tydinga.

Ser ein på kjeldene for belegga, finn ein at det stort sett dreiar seg om sakprosa i bokform, nokre få tidsskrift, nokre lærebøker, og ein del juridiske tekster. Nesten ingen aviser, men tre belegg er frå Bergens Tidende i 1995. Tidsmessig ligg dei fleste belegga mellom 2000 og 2012, men nokre er frå midt på 1980-talet og 1990-talet. Det er ikkje mogleg å finne noko mønster for fordelinga av dei ymse tydingane på tidsrom og kjeldetype.

Eg har også tatt ein kik på det nynorske tekstkorpuset som er brukt i Norsk Ordbok. Det har eit noko anna kjeldegrunnlag, med stor vekt på aviser, nesten alt frå 2000-talet. Det er ikkje så overraskande når det gjeld eit ord som knyter seg mest til den moderne språkbruken og mindre til den tradisjonelle på nynorsk. Dei avisene det er mest av, er Dag og Tid, Firda, Hallingdølen og Sogns Avis. Belegg frå tidsskrift og bøker er meir spreidde. Også her er fleire belegg tvilsame eller vanskelege å tolke og dermed å avgrense, så tala baserer seg på skjønn, men det store bildet blir nok likevel nokolunde dekkande. Eg nøyer meg her med å sjå på tydingsfordelinga; dei semantiske rollene er mindre interessante, for recipiens-objekta dominerer her òg.

Eg finn ca. 40 belegg som er relevante og kategoriserbare i dette materialet. Fordeler vi dei på dei tre tydingane av verbet, ser vi dette mønsteret (igjen med det atterhaldet at tala er omtrentlege og ikkje presise):

Positiv tyding: 20

Negativ tyding: 12

Nøytral: 8

Altså ei klar overvekt på den positive tydinga, samtidig som den negative også er godt representert.

Vil ein prøve å forklare denne skilnaden mellom bokmåls- og nynorskmateriale, så er det neppe særleg fruktbart å peike på målforma. Bruken spriker så mykje i begge målformene at det må vere eit trekk ved norsk språkbruk heva over skilnaden mellom bokmål og nynorsk. Derimot kunne ein kanskje sjå nærare på sjangeren kjeldene høyrer til: Det verkar ikkje så overraskande at eit utval belegg frå aviser frå etter 2000 gir større utslag for den positive tydinga enn eit utval sakprosattekster i bokform, der ein kan tenke seg at profesjonelle språkvaskarar har vore meir inne i bildet enn i avisredaksjonane. Ei tilsvarande granskning basert på eit aviskorpus på bokmål kunne kanskje ha kasta lys over det. Denne artikkelen har eit format som ikkje tillèt ei slik utviding, men det står alle fritt å supplere den undersøkinga eg har gjort.

[3.4] *Og kva skal leksikografane gjere?*

No er tida inne til å kaste eit blick på ordbøkene att. Spørsmålet er om ikkje dei dataa eg legg fram her, kvalitetssikra med supplerande undersøkingar, burde verke tilbake på ordbøkene, slik at begge tydingane blir greidd betre ut, og at kvalifiseringar som «familiært» og «muntlig» blir modifisert eller tatt ut. For «forfordele» i tydinga «gi for mykje» er openbert ein del av vanleg skriftspråksbruk i dag. Om det er «uheldig», må det vere opp til kvar enkelt å synse om.

Eit par av nynorskbelegga er tatt frå drøftingar av sjølve ordet og tydinga. I Hallingdølen skriv Dag Wollebæk i 2009:

I Hallingdølen tysdag 12.5 går det fram at bustyraren for Elkjøp Nesbyen AS meiner at Elkjøp Norge og Nes Prestegjelds Sparebank vart forfordelt i forkant av konkursen i Elkjøp. Og då meinte bustyraren at desse hadde fått for mykje samanlikna med andre kreditorar. Om det er Hallingdølen eller bustyraren som har gått i «forfordelingsgrøfta» veit eg ikkje. Men å bli forfordelt tyder faktisk å få for lite. Svært mange trur det er motsett. Derfor eit råd til alle journalistar - og alle andre skrive- og taleføre. Ikkje bruk ordet forfordele; ein risikerer anten å bli misforstått - eller ein brukar det feil.

Og nokre år før, i 2000, skriv Johan Brox i Dag og Tid:

... ordet forfordelt ... tyder diskriminert, men dei fleste skriv og les ordet i motsett tyding. Når korkje avsendar eller mottakar veit kva den andre legg i orda, er dei ubrukelege for alle. Til dei eventuelt får eit nytt innhald som alle er samde om.

Denne konklusjonen finn vi att i utgangspunktet for denne artikkelen, Fjeld og Vikør 2008: 103:

Da vi lånte det tyske *vervorteilen* inn i norsk, betydde det at den som ble *forfordelt*, fikk mindre enn han/hun skulle ha i en delingsprosess. I dag mener mange at den forfordelte får mer enn det som rimelig er, og med den betydningen kan ordet fungere greit kommunikativt dersom begge parter aksepterer "misforståelsen". Da har ordet rett og slett fått en annen betydning. Verre er det dersom språkbrukerne er uenige om betydningen, slik at en aldri kan vite hva en annen språkbruker mener med ordet. Slik er nok situasjonen for *forfordele* i norsk i dag, og i et slikt tilfelle kan en bli nødt til å advare mot å bruke ordet i det hele tatt uten presisering.

[3.5] Opphav og nærskyldte språk

Ordbøkene (og Fjeld og Vikør) gir opp tysk *vervorteilen* som opphav til *forfordele*. Dette ordet finst ikkje i moderne tyske ordbøker, men det er beskrevet i Grimms historiske *Deutsches Wörterbuch* (der stava *vervortheilen*) med ei negativ tyding som er klart sterkare enn den moderne norske, noko i retning av 'bedra', 'skade' o.l. *Ordbog over det danske Sprog* plasserer dette ordet i det "eldre nyhøgtyske" språkstadiet.

Ein sjekk i nederlandske standardordbøker viser det same mønsteret: *vervoordelen*, som det heiter der, er der belagt frå 15- og særleg 1600-talet, med den same tydinga som i tysk. I den store historiske *Woordenboek der Nederlandsche Taal* har dette ordet (stava *vervoordeelen* etter rettskrivinga fram til 1947) fått ein relativt lang artikkel, men der har ordet i den positive tydinga ('hjelpe') også fått ein liten artikkel der det går fram at det i denne tydinga berre er belagt som dialektord frå tiåra rundt 1900, og da tydeleg inspirert av substantivet *voordeel* 'fordel'. I moderne skriftspråk er dette verbet likevel utdøyd og erstatta av to ord som i motsetning til vårt *forfordele* er fullstendig eintydige: *bevoordelen* står for den positive tydinga og *benadelen* for den negative. Preposisjonen *na* er det motsette av *voor*, altså 'etter', og *nadeel* er det vi kallar *ulempe*. For nederlendingane finst altså ikkje dette problemet lenger.

Svenskane ser også ut til å ha unngått det; *förfördela* blir eintydig definert med den negative tydinga i alle ordbøker som eg har konsultert. Men *bruken* i svensk

har eg ikkje undersøkt.

I dansk liknar stoda meir på den norske: *forfordele* får den negative tydinga i alle ordbøker, bortsett frå ei, og den er viktig: Seksbandsverket *Den danske ordbog* fører opp begge tydingane: den tradisjonelle negative som tyding 1, den positive som tyding 2, og der blir det opplyst at den er belagt frå 1964 av og blir mest brukt blant yngre menneske. "Denne brug regnes af mange for ukorrekt", melder ordboka.

Både i dansk og i norsk er det vel rimeleg å tenke seg at denne positive tydinga har utvikla seg i språkbruken på grunnlag av substantivet *fordel*, slik vi såg eit lite utslag av i nederlandsk: Ein lite gjennomsiiktig tradisjonell bruk blir omtolka på grunnlag av eit notidig språkbruksmønster som er umiddelbart forståeleg for nye generasjonar. Dermed ser vi ein parallell til det som har skjedd med *bjørneteneste* i moderne norsk talemål.

[4] DESKRIPTIVITET VS. NORMATIVITET

Utgangspunktet for denne drøftinga er forholdet mellom deskriptivitet og normativitet i ordbøkene. Eg skreiv at i dei store allmennordbøkene er formopplysningane (rettskriving og bøying) prinsipielt normative, mens definisjonane og andre tydings- og bruksopplysningar må vere prinsipielt deskriptive.

Grensa mellom deskriptivitet og normativitet er likevel ikkje hoggen i stein (jf. Fjeld og Vikør 2008: 138). Det er fordi den faktiske språkbruken aldri foregår i eit normativt tomrom. Når NAOB supplerer ein definisjon med markeringar som «NYERE BRUK; ANSETT SOM UHELDIG» (*bjørnetjeneste* tyding 2) og «UHELDIG, MUNTLLIG» (*forfordele* tyding 2), kan det reknast som deskriptivt i seg sjølv: Leksikografen beskriv faktisk eksisterande haldningar i språksamfunnet. Når Norsk Ordbok under *forfordela* tyding 2b skriv «sj(eldsynt)» kan det i enda høgare grad reknast som deskriptivt. Men slike kvalifiseringar, som mest baserer seg på leksikografens oppfatning av situasjonen i språksamfunnet, er risikable. Korpuset kan stille leksikografen i skammekroken: Det er som vist ovanfor klart at *forfordele* i dag er vanleg både i den positive og den negative tydinga i den skriftlege språkbruken; bruksmarkeringane i NAOB og i Norsk Ordbok (og NROs «fam(iliært)») er openbert rett og slett feil i dag. (For Norsk Ordbok har vi den orsakinga at artikkelen *forfordela* stod i eit hefte som kom ut i 1989, men det er slett ikkje sikkert at opplysninga «sj(eldsynt)» stemte da heller; jf. opplysninga i *Den danske ordbog* ovanfor.) Opplysninga «NYERE BRUK» under *bjørnetjeneste* tyding 2 i NAOB er meir «objektiv», og truleg rimeleg enkel å dokumentere. «ANSETT SOM UHELDIG» er ei normativ utsegn i deskriptiv forkledning, det går ikkje fram kven som «anser» dette. Men det at denne tydinga manglar i korpuset, er ein sterk indikasjon på at kvalifiseringa reint faktisk stemmer, og brukarens intuitive

kunnskap om språksamfunnet vil nok stø opp under denne konklusjonen; i alle fall gjeld dette litt eldre språkbrukarar, og i alle fall inntil vidare.

LITTERATUR

Faarlund, Jan Terje, Svein Lie & Kjell Ivar Vannebo. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.

Fjeld, Ruth Vatvedt & Lars S. Vikør. 2008. *Ord og ordbøker. Innføring i leksikografi og leksikologi*. Kristiansand: Høyskoleforlaget.

Knudsen, Rune Lain & Ruth Vatvedt Fjeld. 2013. LBK2013: A balanced; annotated national corpus for Norwegian Bokmål. I *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013*; May 22-24; 2013; Oslo; Norway. NEALT Proceedings Series 19.

ORDBØKER:

Bokmålsordboka. Boye Wangensteen. 2006. Oslo: Universitetsforlaget / Kunnskapsforlaget (1. utg. 1986 ved Marit I. Landrø og Boye Wangensteen).

Den Danske Ordbog. Ebba Hjorth o.a. 2003-05. København: Det Danske Sprog- og Litteraturselskab.

Deutsches Wörterbuch. I-XXXIII. Jacob og Wilhelm Grimm o.a. 1838-1962.

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet I-XII. Alf Hellevik o.a. 1966-2016. Oslo: Det Norske Samlaget.

Norsk Riksmålsordbok I-IV. Trygve Knudsen og Alf Sommerfelt 1937-57. Oslo: Aschehoug.

Nynorskordboka. Marit Hovdenak o.a. 2006. Oslo: Det Norske Samlaget (1. utg. 1986).

Ordbog over det danske Sprog I-XXVIII. Verner Dahlerup o.a. 1918-56. København: Det Danske Sprog- og Litteraturselskab.

Woordenboek der Nederlandsche Taal I-XXIX. Matthias de Vries o.a. 1882-1998. Den Haag: Martinus Nijhoff.

ELEKTRONISKE KJELDER:

Leksikografisk bokmålskorpus - LBK. <http://www.hf.uio.no/iln/tjenester/kunnskap/samlinger/bokmal/veiledningkorpus/>

NAOB: *Det Norske Akademis Ordbok*. <https://www.naob.no>

Norsk Ordboks Nynorskkorpus. http://no2014.uib.no/korpuset/conc_enkelt-sok.htm

KONTAKT

Lars Vikør

Professor emeritus, Institutt for lingvistiske og nordiske studier,
Universitetet i Oslo

l.s.vikor@iln.uio.no

DET UMULIGE ER MULIG

URD VINDENES OG HANS-OLAV ENGER

SAMMENDRAG

Artikkelen tar for seg spørsmålet om defektivitet i grammatiske mønstre er en absolutt egenskap i seg sjøl. Nærmere bestemt undersøker vi adjektiver som angivelig mangler nøytrumsform og adjektiver som angivelig er begrensa til predikativ eller attributiv posisjon. Ved hjelp av data fra Leksikografisk bokmålskorpus argumenterer vi for at det er mer fornuftig å forklare slike defekte mønstre som konvensjonelt *uvanlige* snarere enn grammatisk *umulige*.

[1] INTRODUKSJON

I grammatikken er det vanlig å skille mellom mønstre som er *uvanlige*, og mønstre som rett og slett er *umulige*. Adjektivet i ytringa *Ruth er en glad jublant* er absolutt mulig å bruke, men for mange vil det virke umulig i en setning som *Et glad bursdagsbarn. Et lilla bursdagsbarn*, derimot, vil bare regnes som *uvanlig*. Noen lingvister vil si at nøytrumsforma mangler i bøyingsparadigmet til *glad*, og at defektivitet er en absolutt egenskap. I denne artikkelen vil vi drøfte om det faktisk er slik, eller om det i praksis heller dreier seg om gradualitet. For å belyse spørsmålet vil vi se på forekomsten og frekvensen av morfologiske og syntaktiske mønstre for adjektiver i faktisk bruk. Når vi velger å konsentrere oss om adjektivene, er det delvis fordi de spiller en viktig rolle i Ruth Fjelds vitenskapelige produksjon (se f.eks. Fjeld 1995, 1998, 2001, 2005), delvis fordi de faktisk byr på mange gode illustrasjoner på problemet.

Artikkelen har to hoveddeler. Den første delen er en teoretisk bakgrunn der vi stiller et spørsmål. Vi vil kontrastere to perspektiver som vi slagordspreget kan kalle «mulig mot umulig» og «vanlig mot uvanlig».

Så går vi over til en mer empirisk del, der vi prøver å se nærmere på hva vi kan finne ut om spørsmålet ved hjelp av korpus, særlig Leksikografisk bokmålskorpus (Knudsen & Fjeld 2013).

[2] TEORETISK BAKGRUNN

[2.1] Generelt

A priori virker det rimelig å skille mellom det som er mulig, men tilfeldigvis ikke

forekommer i praksis, og det som prinsipielt – av en eller annen grunn – ikke er mulig. Dette har vært vanlig i både strukturalistisk og generativ tenking, og det er gjenspeilt i ulike. For eksempel sier Enger & Kristoffersen (2000: 118) som følger:

Syntaks er ikke begrenset til studiet av eksisterende setninger i norsk. Syntaksen handler like mye om å finne ut hva som er mulige setninger i norsk. Setningen *Tore André Flo lagde et vakkert mål i fotball-VM i Frankrike* har nok vært ytret mange ganger; setningen *Min avdøde grandtante Hildur lagde et vakkert mål i fotball-VM i Frankrike* har sannsynligvis aldri vært ytret (før nå). I en viss forstand er altså den førstnevnte setningen eksisterende, den andre ikke, men syntaktikere flest vil si at det ikke er så interessant; det interessante er at begge setningene er mulige [...] På samme vis vil de fleste morfologer mene at vi ikke bør innskrenke oss til å se på skillet mellom eksisterende og ikke-eksisterende ord. Vi må også studere skillet mellom mulige og umulige ord.

Å betone skillet mulig – umulig framfor for eksempel skillet vanlig – uvanlig, det gir god mening både innafor et strukturalistisk og et generativt perspektiv. Og som sagt, det er langt fra urimelig å tenke på denne måten.

Samtidig kan motsetningsparet «mulig versus umulig» nyanseres videre i flere ulike retninger. I en del generativ lingvistikk har en skilt mellom grammatikalitet og akseptabilitet. Det u/grammatiske er da det som den formelle grammatikken genererer, eventuelt ikke genererer, og det skillet er skarpt og et enten/eller. Informanter kan vurdere akseptabilitet, og de har graderbare intuisjoner. Denne tenkemåten har det vært uttrykt en del skepsis til i seinere år (f.eks. Dąbrowska 2012, Sampson 2017). En kan også lure på om grammatikalitet kan løsrives fullstendig fra akseptabilitet, men slike spørsmål vil fort føre til en svært omfattende diskusjon. Vi har et litt annet utgangspunkt, idet vi snarere ønsker å betone konvensjonalitet og bruk.¹

Om en framfor å betone språkets kreative aspekter, slik generativ grammatikk har tendert mot, velger å betone språkbruk og språkets konvensjonalitet, som en del kognitive lingvister har betont, så vil 'vanlig' eller 'konvensjonell' rykke oppover på prioriteringslista. Bruksbaserte teorier vil gjerne betone konvensjonalitet, blant annet for å få fram den sosiale dimensjonen ved språkssystemet (her er det ei likhet med europeisk strukturalisme) og for å få fram at bruken former systemet.

Konvensjonalitet er graduel. Å trekke inn konvensjonalitet og bruk betyr

[1] Takk til fagfellene for grundige og nyttige kommentarer, både her og mange andre steder.

ikke at alt er mulig, snarere at det er grader av u/mulighet. Grammatiske vurderinger framstår etter dette synet snarere som graduelle og probabilistiske.

Svanlund (2004: 134) setter ord på dette synet når han hevder at «ordens egenskaper i hög grad är ... probabilistiska snarare än absoluta och detta gäller även grammatiska egenskaper».

[2.2] Skandinavisk adjektivmorfologi

Innafor bøyingsmorfologien er det vanlig å regne med at det er en prinsipielt viktig forskjell på det at en celle i paradigmet av en eller annen grunn er 'lukket', slik at paradigmet er defekt, og at cellen i prinsippet kan fylles, men i praksis sjelden blir det. Så da blir et spørsmål hva skandinaviske grammatikere regner som utelukket, og vi konsentrerer oss her om adjektivene.

I svensk grammatisk tradisjon er det vanlig å anta at adjektivet *lat* har et defekt paradigme. En kjent artikkel heter faktisk «Varför barnet inte kan vara latt» (Pettersson 1990), og Löwenadler (2010) behandler slike restriksjoner på produktivitet på mange svenske adjektiver (ikke bare *lat*) utførlig i et OT-inspirert perspektiv.

Svenska akademien grammatik bind 2 (Teleman et al. 1999, heretter SAG), s. 214 skiller mellom «Adjektiv utan obestämd form neutrum singularis», som er omtalt i § 56, og «Oböjliga adjektiv», som altså har ei nøytrumsform som er identisk med felleskjønnsforma, og er omtalt i § 57. Om adjektiver uten nøytrumsform heter det som følger (SAG, bind 2, s. 214):

Av vissa adjektiv bildas inte eller bara sällan neutrumformer. Detta gäller särskilt följande typer av adjektiv:

a) Flerstaviga adjektiv med slutbetoning och stamslut på lång vokal plus [d]: *timid, stupid, morbid, perfid, hybrid, rigid*. Undantag är *solid*, som har neutrumformen *solitt* [...] Någon gång kan en sådan neutrumform uppträda vid något av de andra adjektiven [...] men i allmänhet uppfattas den inte som språkriktig...

b) Adjektiven *rädd* och *fadd*. Någon gång förekommer neutrumformer på [-t] [...]

c) Tre enstaviga adjektiv som slutar på -t, nämligen *lat, flat* och det obscena *kåt*, samt tre enstaviga adjektiv som slutar på -d, nämligen *pryd, snöd* och *vred*. Vid de tre som slutar på -t är neutrumformen helt omöjlig [...] Vid dem som slutar på -d torde somliga språkbrukare marginellt kunna tänka sig neutrumform

d) Flerstaviga adjektiv (med slutbetoning) som slutar på lång vokal, t.ex. [...] *disträ, jalu*. Marginellt kan dessa adjektiv ansluta sig till de oböjliga adjektiven och uppträda i oböjd form [...] Inte heller

kry torde kunna stå i neutrum. Benägenheten att acceptera oböjd form som neutrumform varierar mellan olika språkbrukare och för olika adjektiv.

Et slående trekk ved dette tekstutdraget er ordlegginga «inte eller bara sällan». Ut fra ideen om at noe er enten mulig eller umulig, kan vi si at SAG iallfall iblant plasserer seg trygt i midten.

Vi har ikke funnet noen norsk grammatikk som er like utførlige som SAG med hensyn til adjektiver som ikke kan forekomme i nøytrum, og som altså skulle være defekte med hensyn til nøytrum. Norsk referansegrammatikk (Faarlund et al. 1997, heretter NRG) har for eksempel ikke noen liste som svarer til denne. De opererer rett nok med ubøyelige adjektiver, men ikke med adjektiver som er defekte på dette punktet.

Ubøyelighet er i utgangspunktet noe annet enn defektivitet. Et defekt leksem skal etter vanlig definisjon mangle noe; det skal være tomme «celler» i paradigmet, slik det for eksempel kan være for flertall av *ull* eller entall av *opptøyer*, for mange av oss. Et ubøyelig adjektiv har alle cellene fylt opp, men alle «fangene» ser like ut.

Noen ubøyelige adjektiver er åpenbart ikke defekte; et eksempel kan være *bra*, som kan brukes både framfor nøytrumsord og ord i flertall. Ved en del andre ubøyelige adjektiver er det likevel større usikkerhet. Albright (2003, 2009) antyder at usikkerhet om en prosess gjelder, diakront sett kan gi opphav til genuin defektivitet, og Sims-Williams & Enger (i kjømda) mener dette kanskje kan gjelde også for en del av de norske adjektivene vi ser på her. Det er for eksempel ganske sjelden adjektivet *lat* blir kombinert med et nøytrumsord (ettersom de fleste substantiver som betegner folk, ikke er nøytrer), og en kan bli usikker på det en sjelden hører.

Det kan altså være en viss sammenheng mellom defektivitet og ubøyelighet, og da er det ikke så urimelig at en iblant kan komme litt i tvil om det er det ene eller andre som er ment i den norske litteraturen. Dette viser seg ikke minst i en del ordbøker, som har en tendens til å insistere på at alle celler for et adjektiv må være fylt. Da fylles de cellene der noen ville se defektivitet, opp av former som ser ut som grunnforma, litt som i SAGs punkt d) i sitatet ovafor. (*Et gøy barn skal for eksempel være greit, ifølge Bokmålsordboka*).

Om norske grammatikker ikke er like utførlige som SAG, gir de likevel en del opplysninger om ubøyelighet og defektivitet. Næs (1972: 219, 220) skriver at «[bøyning] med bare to former (uten -t i n. sg.) har [...] Noen enkelte ord: kåt, lat, glad, [...], redd, lærd, solid, absurd, fremmed» og han nevner også «Adj. uten bøyning: [...] Noen enkelte ord: bra, tro, slu, kry, edru, sanndru». Om nasjonalitetsadjektiver som *fransk, engelsk*, sier Aasen (1965: 159) at «man dog helst undgaar

at bruke dem i Intetkjøn». Enger (1993: 51) skriver følgende:

Nå er det et spørsmål i hvilken grad adjektiver som *glad* i det hele tatt forekommer i nøytrum; jeg kjenner iallfall et visst ubehag ved konstruksjoner som *et glad barn*. Det fins flere adjektiver som i praksis ser ut til å ha defekte paradigmer [...] Men dette problemet skal vi [...] se bort fra i denne artikkelen.

Johansson & Torkildsen (2005) skriver at adjektiver som *lat* er sterkt markert som felleskjønn, både fonologisk og semantisk, og den semantiske grunnen ser de i at disse adjektivene er «adjektiv för känslor och tankeprocesser». Indre, ofte skjulte tilstander gjør, mener de, nøytrum av visse adjektiver (som *lat*, *redd*, *glad*) mindre akseptable. I undersøkelsen som Johansson & Torkildsen (2005) utførte, brukte de svensktalende forsøkspersonene signifikant lengre tid på å akseptere at former som *ett latt barn* fins, sammenlikna med *ett fint barn*. For norsk blei *et lat barn* og *et kry barn* vurdert (og akseptert) langsommere enn *et latt barn*. De blei altså akseptert av mange, men under tvil. *Et latt barn* blei ikke akseptert av majoriteten, men det var betydelig individuell variasjon.

Sjøl om NRG ikke regner med noen defekte adjektiver mht. genuskongruens, så gjør de det når det gjelder attributiv/predikativ (s. 393):

enkelte adjektiver står helst ikke som foranstilt adledd:

tilstanden var ille	?en ille tilstand
selskapet var gøy	?et gøy(t) selskap
dei vart kvitt han	*ein kvitt mann
huset er verdt ein million	*eit verdt hus

Det fins også adjektiver som helst ikke står predikativt. Det gjelder bl.a. epistemiske adjektiver [...] og adjektiver i partitiv bruk [...]:

eventuell konkurs	*konkursen er eventuell
forrige søker	*søkeren var førre
Østre Toten	*Toten er østre

Vi merker oss ordlegginga med «helst ikke», som kan minne om SAGs formulering ovafor – og om Aasens «dog helst ikke» og Engers «et visst ubehag». Det er verdt å merke seg at NRG faktisk stjernemerker noen nøytrumsformer og setter spørretegn framfor andre – i samme avsnittet, innledet av «helst ikke». Det kan

tolkes som at grensa mellom «absolutt umulig» og «bare litt uvanlig» er vanskelig å trekke her, og det støtter vel våre antakelser i 2.1.

Flere hint om defektive adjektiver finner vi hos Beito (1986: 216), som skriver om «fåforma adjektiv» at «[somme] adjektiv vert mest berre nytta i upersonlege ordlag», og nevner som eksempel (*det var*) *farsynt, kveldvart, ufredt*.

Også ved gradbøying har det vært bemerket at det er «ofte vanskelig å vite om et adjektiv kan gradbøyes eller ikke, og om det i så fall skal få morfologisk eller perifrastisk gradbøying» (Fjeld & Vindenes 2013, se også Fjeld 1995 og 1998, og Spilling 2012, Spilling & Haugen 2013).

Dette åpner for et interessant spørsmål, i tråd med diskusjonen om konvensjonalitet versus grammatikalitet: Er defektivitet en absolutt egenskap i praksis? Hvordan ser det ut i et korpus – er det umulige mulig? Og omvendt: Er alt som er mulig, like mulig, eller kan det være slik at «alle dyr er mulige, men noen er muligere enn andre»?

[3] HVA SIER EMPIRIEN?

[3.1] Metode

For å belyse spørsmålet om adjektiver har defekte paradigmer, og om defektivitet framstår som en absolutt egenskap, drøfter vi eksempler og frekvenser fra Leksikografisk bokmålskorpus (LBK).

Vi ser først på et syntaktisk mønster, nemlig begrensning til enten attributiv eller predikativ bruk (jf. avsnitt 2.2). For å finne relevante treff har vi tatt for oss adjektivene som er nevnt i NRG, og søkt på henholdsvis «artikkel + ADJ + substantiv» (for adjektiver som skulle være begrensa til predikativ: *ille, gøy, kvitt* og *verdt*), og «være/bli (lemma) + ADJ + sluttegn» (for adjektiver som skulle være begrensa til attributiv: *eventuell, forrige*). I tillegg har vi gjort ‘omvendte’ søk for å sammenlikne frekvenser, f.eks. sammenlikning av predikativt *ille* med attributivt *ille*.

Det andre mønsteret vi undersøker, er de angivelig defekte bøyingsparadigmene for adjektiver, nærmere bestemt manglende nøytrumsform. Innafor denne kategorien ser vi både på en del fonologisk sett like adjektiver som kanskje mangler nøytrumsform, og på noen som av andre grunner ikke skal forekomme i nøytrum. For å finne relevante treff har vi søkt på ubestemt artikkel «et»/«en»/«ei» og det aktuelle adjektivet, med mulighet for inntil to tilleggsord mellom. Vi har også søkt på den tenkte nøytrumsforma for det aktuelle adjektivet (som *gladt, latt* og *rutt*). Vi har manuelt luka ut irrelevante treff, som «et ille valgt eksempel», hvor *ille* fungerer som adverb, og «gjøre en ille til mote», hvor *en* ikke er artikkel, men pronomen.

[3.2] *Adjektiver som skulle være begrensa til enten attributiv eller predikativ**Ille*

Adjektivet *ille* er et av adjektivene NRG regner som defektivt med hensyn til attributiv posisjon. Søkestrengen «et ille» får 4 treff i LBK, men bare to av disse er attributive og relevante for oss:

- (1) For mange barn er lutefisk et ille barndomsminne (LBK, SA11MH0603.43)
- (2) et ille eksempel (LBK, SA11Ro1235.69)

LBK gir også 10 treff på «en ille» (men ingen på «ei ille»), og kun tre av dem er relevante. Alle tre er fra samme tekst, og har den samme oppbygginga med det samme substantivet:

- (3) æ hadde en ille oppvekst (LBK, SK01ArIn01.5573)

Et enkelt søk på *ille* i predikativ posisjon gir 247 relevante treff, så de tre eksemplene på attributiv posisjon er ikke overveldende mange, men nok til å vise at det ikke er *umulig* for adjektivet *ille* å forekomme i dette mønsteret, iallfall ikke for alle brukere av norsk. Sjøl om *ille* forekommer veldig mye oftere som predikativ, står det som regel til pronomenet *det*, som i «Det er ille når plantene begynner å brekke» (LBK, TV03Bo1015.311). Hele 145 av 247 treff på predikativt *ille* har *det* som subjekt. Med andre ord inngår *ille* svært ofte i dette nærmest leksikaliserte uttrykket, og det virker ikke spesielt vanlig som modifierende ledd til substantiver verken attributivt (*en ille oppvekst*) eller predikativt (*oppveksten var ille*).²

Gøy

Adjektivet *gøy* er fonologisk sett likt adjektivet *drøy*, og en skulle kanskje forvente at de oppførte seg likt med hensyn til syntaktisk posisjon og bøyning, men slik er det ikke. Søk på «et gøy» og «et gøyt» får ingen treff i LBK, men her kommer også den problematiske – og potensielt ‘ikke-eksisterende’ – nøytrumsforma inn. Likevel er det mulig å finne eksempler på «gøyt», men da som predikativ (seks treff). «Ei gøy» får ingen treff, men «en gøy» får tre treff:

- (4) det er en gøy bok (LBK, SK01KrBe01.2223)
- (5) Det var en gøy samling av mennesker (LBK, UN00Ungd01.29994)

[2] Tradisjonelt er jo *ille* regnet som adverb, ikke adjektiv. Vi følger NRG her, for fullstendighetens skyld. Et mer semantisk argument for den tradisjonelle adverbklassifiseringa er likevel at pronomenet i *Det var ille* i regelen ikke viser til entiteter, men til prosesser i vid forstand (altså ikke for eksempel *Det bordet var ille*, men snarere *Det var ille at Trump ble valgt*).

- (6) Jeg kan en gøy en (LBK, TV03BB1005.332)

Eksempelene over er kanskje marginale, men mulige.

Kvitt og verdt

Det fins ingen forekomster av adjektivene *kvitt* og *verdt* i attributiv og foranstilt posisjon i LBK. Disse adjektivene tar obligatorisk nominalfrase som utfylling: *kvitt problemet*, *verdt mange penger* (se NRG: 403), som gjør at de vanskeligere kan stå foranstilt. Derimot er det ikke noe uvanlig at disse adjektivene forekommer attributivt og etterstilt: *Et friskt pust verdt sin vekt i gull*.

Epistemiske adjektiver

Ifølge NRG står epistemiske adjektiver «helst ikke» predikativt. Eksempler på slike adjektiver er *sann*, *virkelig* og *reell*, og disse er nok for de fleste uproblematisk i predikativ posisjon – men betydninga blir da en annen (jf. f.eks. *en sann prøvelse* med *historia er sann*). Vi har ikke klart å få relevante treff på epistemiske adjektiver i predikativ posisjon, som f.eks. «... er eventuell» eller «... er potensiell» i LBK. Dette har med semantiske begrensninger å gjøre: En del adjektiver får klassifiserende betydning i attributiv posisjon, men ikke i predikativ (se Haugen 2013, kap. 5).

NRG bruker som nevnt i avsnitt 2.2 eksempelet *en eventuell konkurs* vs. **konkursen er eventuell*. Den tenkte ytringa **konkursen er eventuell* fungerer dårlig fordi *eventuell* står i predikativ posisjon, men det kan også tenkes at den bestemte forma av substantivet, sammen med et epistemisk adjektiv, gjør at den oppfattes enda mer marginal. Et eksempel som *?en konkurs (som) er eventuell* høres heller ikke bra ut, men er kanskje noe bedre enn varianten med bestemt form. Det samme gjelder eksempler som *en mulig løsning* og *løsningen er mulig*. Dersom vi konstruerer et motstykke med predikativ, og beholder den ubestemte formen, blir ikke resultatet så dårlig som med bestemt form: *en løsning som er mulig, er* Et utvida søk etter eksempler på epistemiske adjektiver i predikativ posisjon utafør LBK ga oss følgende resultat:

- (7) ikke se bort ifra at han er potensiell som glassknuser (vgd.no)
 (8) Lokalitet er potensiell som restaurerings-biotop for gammel granskog i området (rapport på follolandbruk.no)

Eksempelene over viser at *potensiell* blir brukt i predikativ posisjon, sjøl om det fortsatt må regnes som marginalt.

[3.3] Adjektiver som har fonologisk motivasjon for avvik mht. nøytrumsbruk

Enstava adjektiver som ender på lang a

Denne gruppa adjektiver har oftest ikke noen egen nøytrumsform på *-t*. Spørsmålet er om denne begrensninga gjør nøytrumsbruk umulig. Korpussøk i LBK gir 208 treff på «et bra» (mot ca. 702 «en bra»), 5 treff på «et glad» (mot ca. 87 «en glad»), og 2 treff på «et sta» (mot 22 «en sta»). Skilnaden mellom *bra* på den ene sida, og *glad* og *sta* på den andre, høver jo godt med ideen om at animathet (og «indre egenskaper», jf. avsnitt 2.2) er relevant for frekvensfordelinga mellom nøytrum og utrum, men *bra* er uansett mye mer frekvent enn de to andre. Eksempler på *sta* og *glad* med nøytrum er:

- (9) med et glad og rolig sinn (LBK, SA01MoTo01.6038)
- (10) som en indisk tempeldanserinne nylig sa i et glad utbrudd (LBK, SA04LaLR01.584)
- (11) Vi er et sta og et hardt folk (LBK, TV02UM9401.5848)

En sammenlikning av adjektivet *glad* med det semantisk relativt like *blid* viser at fordelinga nøytrum – utrum er annerledes når de fonologiske begrensningene ikke er til stede: For *glad* var den 5 mot 87, mens for *blid* er den 23 mot 100. Kanskje påvirker likevel semantikken også frekvensfordelinga her til en viss grad (jf. Johansson & Torkildsen 2005 og avsnitt 2.2), for *glad* brukes gjerne om en indre følelse (*en glad forsker*), mens *blid* brukes om et ytre uttrykk (*et blidt åsyn*). Flesteparten av eksemplene på «et blidt» i LBK gjenspeiler dette: «et blidt oppsyn», «et blidt ytre», «et blidt smil», osv. Videre kan en forsker være både glad og snill, mens et forslag bare kan være snilt, ikke glad.

Hvis vi sammenlikner det ‘vanlige’ adjektivet *trassig* med det semantisk sett ganske like *sta*, ser vi at det er en forskjell i fordelinga av utrum og nøytrum også her: «En trassig» gir 39 treff, og «et trassig» gir 26 treff, mens «en sta» som nevnt gir 22 treff, og «et sta» 2 treff i LBK. Her ser vi altså en tydelig skeivfordeling: For *trassig* er fordelinga hankjønn mot intetkjønn 3 mot 2, mens for *sta* er den 11 mot 1. Det tyder på at fonologisk motivasjon virker inn, men ikke deterministisk.

Her må termen «fonologisk motivasjon» presiseres. Det hadde ikke vært vanskelig å uttale ei adjektivform *statt*, jamfør det homonyme stedsnavnet *Stadt*. For den som er usikker på nøytrum av *ru* (se nedafor), er det neppe fonotaksen som legger hindringer i vegen; heldigvis kan vi uttale navnet *Ruth*. Poenget er at det er en tendens i norsk morfologi til at enstava adjektiver som ender på trykksterk vokal, ikke så gjerne får /t/ lagt til i nøytrum. Denne tendensen kan sammenlignes med tendensen til at adjektiver som brukes om høyere animater, ikke så gjerne får /t/ lagt til i nøytrum. Denne sistnevnte tendensen gir god mening,

fordi substantiver som betegner høyere animater, sjelden er nøytrere. Den førstnevnte tendensen gir derimot knapt noen mening, fonologisk sett, men for morfologien spiller det liten rolle at den ene tendensen er mye mindre velmotivert enn den andre; den førstnevnte morfologiske tendensen tar hensyn til fonologiske egenskaper, den sistnevnte til semantiske – og det er det.³

Enstava adjektiver som ender på annen lang vokal

Ei annen gruppe adjektiver der vi kan tale om «fonologisk motivasjon», nettopp i denne betydningen, for at -t ikke blir lagt til, er enstava adjektiver på -u og -y, som *sky*, *kry*, *slu* og *ru*. I LBK fikk vi ingen treff på «et kry», men til gjengjeld var det kun 3 relevante treff på «en kry»: «en slags kry spankulering», «en kry hane» og «en kry mann». Dette adjektivet er altså ikke så frekvent i utgangspunktet.⁴ Adjektivet *sky* er nok mer vanlig, og frekvensfordelinga ser slik ut: 10 «et sky» mot 25 «en sky» (hvor «sky» er markert som adjektiv).

(12) med et sky blikk mot naboene (LBK, SK01StGu02.3807)

(13) en tilsynelatende sky mann (LBK, SK01HaEs02.3218)

Med andre ord er det langt fra umulig for *sky* å brukes med nøytrum, men forma *skytt* har vi ingen eksempler på.

Adjektivet *slu* er heller ikke fryktelig frekvent i attributiv posisjon i LBK, og fordelinga mellom nøytrum og utrum ser slik ut: «et slu» gir 4 treff, mot 55 treff på «en slu». For *ru* er fordelinga nokså lik: 1 «et ru» mot 37 «en ru».

(14) Powers sendte meg enda et slu smil (LBK, SK01HoAn01.1609)

(15) en liten slu kontorstrateg (LBK, SK04LoOi02.754)

(16) Kroppen er som et ru sandpapir (LBK, SK01AgSu01.2948)

(17) Den første akkorden har en ru overflate (LBK, AV03Un0203.157)

Det som derimot er verre å finne eksempler på, er tenkte nøytrumsformer som *rutt* eller *slutt* (eller *skytt*, *krytt*, *statt*, *bratt*, *gladt*). Fjeld (2017) skriver i *Skriv godt og riktig på jobben* at om en skal bruke adjektiver i denne kategorien attributivt, så bør det hete «et bra intervju» og «et slu dyr», ikke «et slutt dyr». Når Ruth Fjeld tar opp dette spørsmålet i ei språkkritighetshåndbok, er det rimelig å anta at dette reflekterer at språkbrukerne kan være usikre på hvordan nøytrum faktisk skal dannes. Og likevel har vi sett at nøytrumsbruk ikke er helt umulig eller

[3] Fordi termen 'motivasjon' kan antyde funksjonalitet, hadde en annen term kanskje vært mer tjenlig, men vi holder oss her til nokså vanlig terminologi.

[4] Et søk i N-gram-tjenesten til Nasjonalbiblioteket (2015) viser at populariteten til *kry* var på topp i 1919, og stort sett har vært dalende siden den gang.

helt blokkert. Det tyder igjen på at fonologisk motivasjon virker inn, men ikke deterministisk.

Andre enstava adjektiver med lang rotvokal

Enkelte enstava adjektiver med lang rotvokal mangler tilsynelatende nøytrumsform, f.eks. *lat* og *kåt*. Det sistnevnte adjektivet er ikke spesielt frekvent i LBK, og fordelinga er 0 treff på «et kåt» mot 19 treff på «en kåt». Dette kan nok være både semantisk og fonologisk motivert. Om vi beveger oss ut av LBK for å leite videre etter nøytrumskontekster, er det ikke vanskelig å finne eksempler som følgende:

- (18) «Innovasjonsindeksen» er et kåt frieri til et utvalg av de største bedriftene. (shifter.no)

Adjektivet *lat* er mer frekvent, men hva sier LBK om genusfordelings? «Et lat» gir ikke mer enn 1 treff, mens «en lat» gir 40.

- (19) fordi jeg selv er et lat og distre menneske (LBK, UN04USN01.45131)

Om vi sammenlikner *lat* med det liknende adjektivet *doven*, ser vi en tydelig skeivfordeling: Mens fordelinga felleskjønn mot intetkjønn for *lat* er 40 mot 1, er den cirka 3 mot 1 for *doven* (52 «en doven» og 17 «et dovent»). Igjen kan vi konstatere at fonologien påvirker genusfrekvensen, men heller ikke i denne kategorien virker den deterministisk inn.

Genusfrekvens for adjektiver med normal bøyning

Vi har i dette avsnittet sett på adjektiver med angivelig manglende eller svært marginal nøytrumsform. For å få et anslag på hvor frekvente nøytrumsformer av adjektiver er generelt, har vi gjort en opptelling av bøyingsformer i LBK for noen adjektiver med 'normalt' bøyingsmønster:

adjektiv	nøytrum	utrum
flat	458 (30%)	1093 (70%)
stolt	116 (9%)	1206 (91%)
fin	725 (23%)	2488 (77%)
snill	22 (8%)	251 (92%)
høy	1010 (37%)	1740 (63%)

TABELL 1: Fordeling av genusformer for noen adjektiver med 'normalt' bøyingsparadigme (attributiv posisjon).

Som vi ser i tabellen, er ikke nøytrumsformer veldig frekvente i utgangspunktet.

Når det aktuelle adjektivet attpåtil uttrykker en egenskap som er en følelse eller indre egenskap (som *snill* og *stolt* – men ikke *høy*), er det ikke ofte det kongruerer med nøytrumsord. Som nevnt i 2.2 kan slikt føre til usikkerhet, som igjen i ytterste konsekvens kan føre til defektivitet.

[4] OPPSUMMERING

Leksikografisk bokmålskorpus viser at det umulige i en del tilfeller er mulig, men at det kanskje ikke er så frekvent. I en del tilfeller kan det se ut som semantisk og fonologisk informasjon «konspirerer» mot nøytrumsformer. For adjektivene *glad*, *sta*, *gra*, *kry*, gjelder det for eksempel at de ikke bare har vokaler som skulle tilsi ingen *t*, men også at de har semantikk som skulle tale mot at nøytrumsforma fins særlig ofte. Når det gjelder begrensning av adjektiver til enten predikativ eller attributiv posisjon, viser det seg at det også her ofte er snakk om uvanlige, men ikke umulige mønstre. Adjektiv som *ille* og *gøy* kan for eksempel ikke sies å være helt utelukka fra attributiv bruk, sjøl om det er lavfrekvent.

Det kan altså se ut som det er mer gagnlig å se på motsetningen mellom regelrette og defekte adjektiver som graduell snarere enn absolutt. Det utelukker ikke at enkelte former faktisk er absolutt utelukket, men det gjør det enklere å forholde seg til de øvrige.

Bruksbaserte teorier om grammatikk legger vekt på gradvishet både i synkron og diakron språklig variasjon (se f.eks. Traugott & Trousdale 2010). Termen *gradvishet* refererer blant annet til de semantiske og morfosyntaktiske kopleingene mellom ulike grammatiske konstruksjoner (f.eks. likheten mellom adjektivet *god* og andre adjektiv *fin*, *gul*, *rar*, som kan føre til den analogiske nydanninga *godere*). Denne sida ved begrepet ‘gradvishet’ kan også knyttes til frekvens: Grammatiske mønstre har ulik typefrekvens og ulik bruksfrekvens, som kan endre seg over tid. Om språkbrukeren plukker opp signaler om fonologisk motivasjon for et mønster (f.eks. at kortstava adjektiver med lang rotvokal ikke skal ha egen nøytrumsform), vil de forsøke å følge den konvensjonen (det føles rart eller umulig for mange å si *et glad barn*). Samtidig har disse adjektivene flere likheter og kopleinger til andre adjektiv som ofte forekommer i nøytrumsbruk, og disse likhetene åpner opp for ny bruk basert på analogi.

Grammatiske mønstre som er på randen, eller som ikke passer helt inn i en bestemt kategori, er ofte på randen nettopp på grunn av den lave bruksfrekvensen. Eventuell økning i bruksfrekvensen til slike angivelig umulige mønstre avhenger av hvordan språkbrukerne tolker de semantiske og fonologiske motivasjonene bak begrensningene. Korpusbaserte metoder er nyttige for å peke på slike frekvensforskjeller som kan si noe om hvor sterke slike motivasjoner er. For *kvitt* og *verd*t opplever språkbrukerne kanskje en sterkere begrensning på

attributiv plassering enn de gjør for *gøy* og *ille*, på grunn av den obligatoriske nominale utfyllinga. Umulige mønstre er til sjuende og sist konvensjoner, og konvensjoner kan brytes om ikke motivasjonen for å holde på dem er høy nok.

REFERANSER

- Albright, A. 2003. A quantitative study of Spanish paradigm gaps. I G. Garding & M. Tsujimura (red.), *WCCFL 22 Proceedings*. Somerville, MA: Cascadilla Press, 1–14.
- Albright, A. 2009. Lexical and morphological conditioning of paradigm gaps. I C. Rice (red.), *When nothing wins: Modeling ungrammaticality in OT*. Sheffield: Equinox Publishing, 117–164.
- Beito, O.T. 1986. *Nynorsk grammatikk*. 2. utg. Oslo: Samlaget.
- Enger, H.-O. & K. Kristoffersen. 2000. *Innføring i norsk grammatikk*. Oslo: Cappelen akademisk forlag.
- Enger, H.-O. 1993. Noen adjektiver i nøytrum. *Norskraft* 78, 50 – 61.
- Dąbrowska, E. 2012. Different speakers, different grammars: Individual differences in native language attainment. I *Linguistic Approaches to Bilingualism* 2:3, 219–253.
- Fjeld, R. V. 1995. Noen vanskelige adjektiver i leksikografisk perspektiv. I *Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995*. Nordiske Studier i Leksikografi 3, 149–157.
- Fjeld, R. V. 1998. *Rimelig ut fra sakens art. Om tolkning av ubestemte adjektiv i regelgivende språk*. Doktorgradsavhandling, Universitetet i Oslo.
- Fjeld, R. V. 2001. Interpretation of Indefinite Adjectives in Legislative Language. I Mayer, F. (red.), *Language for Special Purposes: Perspectives for the New Millennium*. Tübingen: Narr, 643–650.
- Fjeld, R. V. 2005. The Lexical Semantics of Vague Adjectives in Normative Texts. I Bhatia, V., J. Engberg, M. Gotti & D. Heller (red.), *Vagueness in Normative Texts*. Bern: Peter Lang, 157–172.
- Fjeld, R. V. 2017. *Skriv godt og riktig på jobben*. Oslo: Universitetsforlaget.
- Fjeld, R. V. & U. Vindenes. 2013. *Vanskelig – vanskeligere – vanskeligst? Om gradbøying som leksikografisk informasjonstype*. Møte om norsk språk 15, Oslo.

- Faarlund, J. T., K. I. Vannebo & S. Lie. 1997. *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Haugen, T. A. 2013. *Polyvalent Adjectives in Norwegian: Aspects of their Semantics and Complementation Patterns*. Doktorgradsavhandling, Universitetet i Oslo.
- Johansson, C. & J. Torkildsen. 2005. Reaktionstider för genuskongruens: hur processeras genus? I S. Lie, G. Nedrelied & H. Omdal (red.), *MONS 10: utvalde artiklar frå det tiande Møte om norsk språk i Kristiansand 2003*. Kristiansand: Høyskoleforlaget, 157–166.
- Knudsen, R. L. & R. V. Fjeld. 2013. LBK2013: A Balanced, Annotated National Corpus for Norwegian Bokmål. I *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013*. NEALT Proceedings Series 19 / Linköping Electronic Conference Proceedings 88, 12–20.
- LBK = *Leksikografisk bokmålskorpus*. <https://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/>
- Nasjonalbiblioteket. 2015. NB N-gram. http://www.nb.no/sp_tjenester/beta/ngram_1/
- Löwenadler, J. 2010. Restrictions on productivity: defectiveness in Swedish adjective paradigms. I *Morphology* 20(1), 71–107.
- NRG = Faarlund et al. 1997
- Næs, O. 1972. *Norsk grammatikk. Elementære strukturer og syntaks*. 3. utgave. Oslo: Fabritius Forlag.
- Petterson, T. 1990. Varför barnet inte kan vara latt. I Andersson, E. & M. Sundman, *Svenskans beskrivning 17*. Åbo: Åbo Akademis förlag, 293–302.
- SAG = Teleman et al. 1999
- Sampson, G. 2017. *The Linguistics Delusion*. Sheffield: Equinox Publishing.
- Sims-Williams, H. & H.-O. Enger (i kjømda). The loss of inflection as grammar complication. *Diachronica*.
- Spilling, E. F. 2012. Gradbøying i norsk: en korpusbasert undersøkelse av talespråk. MA-avhandling, UiO.
- Spilling, E. F. & T. A. Haugen. 2013. Gradbøying i norsk: En bruksbasert tilnærming. I *Maal og Minne*, 105(2), 1–40.

- Svanlund, J. 2004. Varför är sex så knepigt? I U. Hellsten & C. Grünbaum (red.), *Språkets vård och värden – en festskrift till Catharina Grünbaum*. Stockholm: Norstedts Ordbok, 134–145.
- Teleman, U., S. Hellberg & E. Andersson. 1999. *Svenska Akademiens grammatik*. Svenska Akademien.
- Traugott, E.C. & G. Trousdale. 2010. Gradience, gradualness and grammaticalization: How do they intersect? I E.C. Traugott & G. Trousdale (red.), *Gradience, gradualness and grammaticalization*. Amsterdam: John Benjamins, 19–44.
- Aasen, I. 1965 [1864]. *Norsk Grammatik*. Oslo: Universitetsforlaget.

KONTAKT

Urd Vindenes

Institutt for språk og litteratur, Universitetet i Sørøst-Norge

urd.vindenes@usn.no

Hans-Olav Enger

Institutt for lingvistiske og nordiske studier, Universitetet , Universitetet i Oslo

h.o.enger@iln.uio.no

UNDERLIGT NOG/KLOKT NOG – EN GANSKA NY TYP AV SATSADVERB I SVENSKA OCH NORSKA

SVEN-GÖRAN MALMGREN

SAMMANFATTNING

Syftet med studien är att demonstrera hur nya och allt större korpusar kan ge oss djupare insikter i svensk och norsk språkhistoria. Detta illustreras av den ganska nya satsadverbtypen *underligt/klokt nog*, t.ex. *underligt nog var ingen hemma*. För att hitta ett förstabelägg för konstruktionstypen år 2000 fanns bara den stora historiska svenska ordboken (SAOB) att tillgå; där hittades ett belägg från 1839. År 2014 var den svenska Litteraturbanken, där stora delar av den klassiska svenska litteraturen finns digitaliserad, redan väl utbyggd, och ett belägg från 1833 hittades. I dag, år 2020, är så gott som alla viktiga svenska tidningar från 1700-talet och framåt digitalt tillgängliga via Kungliga biblioteket, och nu kan konstruktionstypen spåras tillbaka till åtminstone 1805. Som framgår av studien är utvecklingen i Norge och norskan, såväl vad gäller digitala textmaterial som den aktuella konstruktions-typen, i allt väsentligt parallell.

[1] INLEDNING

I början av 2000-talet ledde jag det Riksbanksfinansierade forskningsprojektet *Det svenska ordförrådets utveckling 1800–2000* (ORDAT-projektet, se Malmgren 2000). En av projektets många delstudier gällde utvecklingen av det svenska formordssystemet, särskilt adverbena och bland dem inte minst satsadverben (Malmgren 2002). Jag kunde konstatera att det svenska adverbsystemet verkade ha genomgått en kraftig utvidgning under de senaste 200 åren, och fann också i litteraturen belägg för att en sådan utveckling hade paralleller i nära nog alla de västeuropeiska språken (Swan 1991, Ramat & Ricca 1998).

Arbetet i ORDAT-projektet byggde empiriskt på sin tids elektroniska och icke-elektroniska material: SAOB/OSA (dvs. SAOB i maskinläsbar form), nyordsböcker och andra ordböcker, samt korpusar som mest täckte det moderna språket men som också gjorde vissa nedslag i äldre skeden. I dag, nästan 20 år senare, är läget helt förändrat: den svenska Litteraturbanken växer ständigt och omfattar numera en stor del av den klassiska svenska skönlitteraturen, och under de senaste åren har en annan imponerande resurs blivit tillgänglig, Kungliga bibliotekets

(KB:s) digitaliserade samling av nästan alla svenska dagstidningar genom mer än 200 år. I Norge tycks man snarast ligga ännu längre fram än i Sverige, särskilt med Nasjonalbibliotekets sajt bokhylla.no.

Ett blygsamt försök att med hjälp av Litteraturbanken bygga vidare på en mindre del av min undersökning från 2002 gjorde jag 2014, innan KB:s tidningsmaterial ännu var tillgängligt (Malmgren 2014). Här ska jag försöka komma ytterligare ett steg längre, denna gång främst med hjälp av KB-materialet och, för en försiktig jämförelse mellan svenska och norska, Nasjonalbibliotekets motsvarande tidningsmaterial. Huvudsyftet är att illustrera vilka möjligheter utvecklingen av de digitala resurserna öppnar för språkhistorikerna, men några preliminära språkliga resultat kommer också att presenteras.

[2] PRESENTATION AV *UNDERLIGT/KLOKT NOG*-KONSTRUKTIONEN. DESS FÖREKOMST I ÄLDRE SVENSKA KORPUSAR

Artikeln tar liksom Malmgren 2014 sikte på en undergrupp bland de svenska (och nu även norska) satsadverb, här exemplifierade med *underligt nog/nok* och *klokt nog/nok*, med följande svenska exempel som lätt kan överföras till norska:¹

- (1) Underligt nog var ingen hemma
(=Ingen var hemma, vilket var underligt)
- (2) Klokt nog svarade han inte
(=Han svarade inte, vilket var klokt)

Jag betraktar *nog/nok* som ett suffix,² varför skenbara fraser som *klokt nog* kan ses som likvärdiga med ettords-adverb som *lyckligtvis*, *tyvärr* m.fl. Det är därför motiverat att kalla *underligt nog/nok* etc. *adverb*. Att dessa adverb (som adverbial) är bestämningar till hela satser antyds av parafaserna inom parentes.³

Adverb bildade med hjälp av *nog/nok* (jag använder i fortsättningen termen *underligt/klokt nog(/nok)*-konstruktioner) möjliggör ett kondenserat uttrycksätt som lämpar sig väl t.ex. för tidningsprosa, och användningen av dem har nära nog exploderat i svenskan och norskan (och säkert även danskan) under de senaste 100–200 åren (förmodligen något tidigare i svenskan än i norskan).

Vilka adverb (ursprungligen adjektiv) kan ingå i *underligt/klokt nog*-konstruktioner? Huvudsakligen två typer, dels adverb/adjektiv som innebär någon

[1] Här tas enbart bokmål upp, inte nynorska. Eftersom studien delvis hänför sig till tidigt 1800-tal, är termen *norska* problematisk. Ibland används termen *dansk-norska*, något som jag dock avstått från här. De texter som ingår i det norska materialet är i varje fall tryckta i Norge och har i allmänhet norska författare.

[2] Bl.a. är det omöjligt att infoga något ord mellan adverbet och *nog* (**klokt inte nog*). Frågan om suffixstatus för *nog/nok* har emellertid ingen betydelse för den följande framställningen.

[3] Inte heller detta är viktigt för fortsättningen, och jag avstår därför från ytterligare argumentation.

sorts värdering av den situation, händelse e.d. som omtalas i satsen (*underligt, beklagligt, märkvärdigt, sällsamt*),⁴ dels adverb/adjektiv som i princip är personbeskrivande och som innehåller ett element av värdering (*klokt, elakt, generöst, förmätet, hyggligt*), Den första gruppen kallas i fortsättningen *situationsvärderande*, den andra *personvärderande* (*adverb*).

Vi ska nu se vilka möjligheter det har funnits att studera framväxten av *underligt/klokt nog(/nok)*-konstruktionen i tillgängliga språkliga källor och material, dels 2002, dels 2014, dels slutligen 2020.

Vi går alltså först tillbaka till 2002. För att få en uppfattning om ursprunget för *underligt/klokt nog*-konstruktionen fanns vid denna tidpunkt i första hand SAOB/OSA att tillgå, och därtill en korpus över tidigare 1800-talssvenska som dock var för liten för att ge något resultat i det här fallet. Men i SAOB hittades ett belägg från 1836 och ett från 1839. Belägget från 1836 bedömde jag som en smula osäkert, men belägget från 1839 (av författaren Almqvist) var otvetydigt:

- (3) Bland alla sina tusen omsorger hade han, skam nog [ung. 'genant nog'], glömt bort kofodret.

Sökningar i SAOB:s arkiv gav inget ytterligare resultat. Däremot hittades många belägg från senare delen av 1800-talet, bl.a. i den Strindbergskorpus som redan nu var tillgänglig, och givetvis enorma mängder från 1900-talet.

På grundval av dessa magra data kunde man möjligen formulera en försiktig förmodan att *underligt/klokt nog*-konstruktionen uppstod i svenskan i början av 1800-talet, och senast på 1830-talet. Längre gick det knappast att komma.

Det kan nämnas, att *Ordbog over det danske sprog*, ODS, ger ett belägg på motsvarande danska konstruktion från mitten av 1800-talet:

- (4) I hendes Saloner samledes ikke, hvad man sædvanligt – naragtigt nok – kalder Eliten af Selskabet.

Intressant nog nämns inte *underligt/klokt nog*-konstruktionen i Westerns norska riksmålsgrammatik från 1921 (Swan 1991). Swan drar den kanske lite djärva slutsatsen att detta sätt att bilda satsadverb ännu inte var produktivt i norskan i början av 1900-talet.

Den motsvarande engelska konstruktionen, ofta med optionellt *enough* som i exemplet *strangely (enough), he didn't come*, är belagd åtminstone sedan slutet av

[4] Ett fåtal svenska satsadverb utgörs enbart av adjektivets neutrumform (dvs. *nog*-konstruktionen behövs inte, och är vanligen inte möjlig): *bestämt, eventuellt, faktiskt, följdligt* (eller *följdenligt nog*), *omöjligt* (dock bara i kontexter med *kan*), *sannolikt, självfallet, självklart, skenbart, säkert och uppenbart*. Ett betydligt vanligare satsadverbbyggande suffix har traditionellt varit *-en* (*troligen, antagligen* etc.). Suffixet *-vis* (*lyckligtvis* etc.) och dess förhållande till *nog* behandlas i korthet i Malmgren 2014.

1700-talet – i OED finns ett exempel från 1783. Man kan då åtminstone tänka tanken att den svenska och danska *underligt/klokt nog*-konstruktionen möjligen uppstått genom påverkan från engelskan, kanske via översättningslitteraturen.⁵

Drygt 10 år efter 2002, 2014, var Litteraturbanken redan i snabb tillväxt. Där påträffades inte så få belägg från 1830-talet av olika skönlitterära författare, däribland ett par som var något äldre än SAOB:s förstabelägg:

- (5) [...] alla dessa tusende strunt-resonemanger åt alla håll som, ömkligt nog, bära namn af bildning [...].
(Almqvist, Dialog om sättet att sluta stycken, 1834)
- (6) Besynnerligt nog blef föret med hvarje mil bättre.
(A. Tömeros, Brev och dagboksanteckningar III (brev 1833))

Det sista citatet var vid denna tidpunkt det äldsta belägg på *underligt/klokt nog*-konstruktionen som med säkerhet kunde påvisas. Både Almqvist och Törneros var representerade i Litteraturbanken med texter från före 1830, men i dessa texter hittades inga otvetydiga belägg på den nya typen av satsadverb. Två ställen i tidigare brev och dagböcker av Törneros förtjänar dock att citeras:

- (7) [...] att han tappert nog slagit sig igenom prestexamen [...]
- (8) [...] Det är märkvärdigt nog, att jag [...]

I det första citatet kan *tappert nog* tolkas både som satsadverb(ial) ("det var tappert av honom") och sättsadverb(ial) ("han slog sig tappert igenom"), men den senare tolkningen förefaller troligast. En poäng är emellertid att det är fullt möjligt att *underligt/klokt nog*-konstruktionen uppkommit via sådana dubbeltydiga fraser (som kunde kallas *brofraser*). I det andra citatet är *märkvärdigt nog* självfallet inte satsadverb(ial), men också denna konstruktionstyp kan ha främjat uppkomsten av *underligt/klokt nog*-konstruktionen (genom en sorts satsförkortning).

Tack vare Litteraturbanken gick det alltså att komma ett litet steg vidare 2014. För det första ökade sannolikheten för att *underligt/klokt nog*-konstruktionen uppstått i början av 1800-talet, kanske på 1810-talet eller 1820-talet. Och för det andra verkade det mycket troligt att den nya konstruktionen var i ganska allmänt bruk på den svenska parnassen redan på 1830-talet. (Se vidare Malmgren 2014.)

[5] Konstruktionen finns också i holländska, med ett förstabelägg från mitten av 1800-talet, däremot inte i tyska och knappast alls i de romanska och slaviska språken (med visst undantag för franska), enligt Ramat & Ricca 1998.

[3] KONSTRUKTIONEN I NYARE SVENSKA KORPUSAR

Nu är det 2020 och digitaliseringen av äldre texter har nått nya höjder. KB:s tidningskorpus framstår närmast som ett ymnighetshorn när det gäller tidiga *underligt/klokt nog*-belägg, som strax ska påvisas. Jag gick (2019) först igenom en lista med några tusen svenska adjektiv och antecknade dem som kunde tänkas ingå i *underligt/klokt nog*-uttryck. Sedan sökte jag på dessa uttryck (*barnligt nog*, *begripligt nog* etc.) och markerade dem som hade förstabelägg tidigare än 1840. Detta resulterade i listan i tabell 1 på nästa sida.⁶

Vad som genast faller i ögonen är här de två 1700-talsbeläggen. Det ska sägas, att båda uppträder i poesi, vilket möjligen kan göra dem mindre vittnesgilla. Men i och för sig förefaller åtminstone det första belägget otvetydigt:

- (9) De Gåfwor, dem af Dig [Gud] vi lånt
Ha wi, beklagligt nog! ej nyttjat som wi böre
(Carlsronas wekoblads 1776)
- (10) En Gång förmätet [ung. självöverskattande] nog, jag wille bjuda til
At upför Wishets Berget klifwa
(Nytt och Gammalt 1784)

Här kan *förmätet nog* tolkas både som sättsadverbial och satsadverbial. På 1700-talet kunde *nog* ha betydelsen 'mycket', och en parafras skulle kunna vara "En gång ville jag på ett mycket förmätet sätt bjuda till ..." (vid sidan av "jag ville bjuda till ..., vilket var förmätet"). Samma dubbeltydighet uppträder ofta hos de personvärderande adverbena (jfr exempel (7) med *tappert nog* ovan), men inte ofta bland de situationsvärderande – *beklagligt nog*, *underligt nog* etc. kan knappast tänkas som sättsadverbial. Med viss tvekan förkastas därför det sista 1700-talsbelägget ovan, medan det första godtas – men också med tvekan. Man kan nämligen notera en liten detalj, utropstecknet efter *nog*. Det kan möjligen tolkas så att inskottet *beklagligt nog* inte är en helt integrerad del av satsen, utan ett s.k. *annex* för att använda en term som lanseras i Svenska Akademiens grammatik, SAG.

I övrigt lägger man märke till en visserligen ganska svag tendens i Tabell 1: de situationsvärderande adverbena uppträder något tidigare än de personvärderande. Vi noterar å ena sidan *beklagligt nog* 1776 (visserligen alltså i poesi och kanske som *annex*), *sällsamt nog* 1805, *besynnerligt nog* 1808, *underligt nog* 1810, *olyckligt nog* 1811, *otroligt nog* 1817, *rysligt nog* 1817 och *märkligt nog* 1824, och å andra sidan *förnuftigt nog* 1812, *otacksamt nog* 1812, *naivt nog* 1825, *klokt nog* 1826

[6] Återigen: KB:s material omfattar bara tidningstexter. Det är fullt möjligt att det går att hitta tidigare belägg i andra typer av texter (dock knappast i Litteraturbanken).

och *barnsligt nog* 1829. (*Löjligt nog* är ett gränsfall som kan föras till båda grupperna, men i första hand till den situationsvärderande gruppen.) Något tillspetsat skulle man kanske kunna säga, att de personvärderande adverbena av *underligt/klokt nog*-typ bryter igenom på 1810-talet eller 1820-talet (och på bred front på 1830-talet), medan de situationsvärderande etablerar sig på allvar redan under 1800-talets första decennium.

Satsdverb av <i>underligt/klokt nog</i> -typ	Årtal för första belägg
barnsligt nog	1829
begripligt nog	1837
beklagligt nog	1776 (?!)
besynnerligt nog	1808
billigt nog (=rimligt nog)	1831
blygsamt nog	1835
elakt nog	1831
fegt nog	1830
frikostigt nog	1837
fräckt nog	1831
förargligt nog	1835
förmädet nog	1784 (?!)
förnuftigt nog	1812
förskräckligt nog	1832
förunderligt nog	1827
klokt nog	1826
löjligt nog	1806
märkligt nog	1824
märkvärdigt nog	1827
naivt nog	1825
naturligt nog	1836
oblygt nog	1829
oförlåtligt nog	1834
olyckligt nog	1811
otacksamt nog	1812
otroligt nog	1817
ovanligt nog	1831
rysligt nog	1817
sällsamt nog	1805
underligt nog	1810
ömkligt nog	1834

TABELL 1: Satsdverb av *klokt nog*-typ med förstabelägg före 1840

[4] KONSTRUKTIONEN I NYARE NORSKA KORPUSAR

Om man söker i Nasjonalbibliotekets korpus över norska tidningar från decennierna före 1850 (www.bokhylla.no), får man också en ganska riklig skörd. Av naturliga skäl kan jag inte gå ner på samma detaljnivå som när det gällde svenska texter, men med det förbehållet är det äldsta påträffade belägget på *underligt/klokt nok*-konstruktionen i de norska texterna följande:

- (11) Underligt nok contrastere de forhen omtalte Berømmelser over Prinds Christian Frederik med hvad der siges følgende St. 30
(Det Norske Nationalblad 15/2 1817)

Andra tidiga *underligt/klokt nok*-adverb i de norska tidningarna är *selsomt nok* (1820), *forunderligt nok* (1825) och *utroligt nok* (1836). Liksom i det svenska materialet verkar de situationsvärderande adverbena uppträda något tidigare i de norska texterna än de personvärderande. Dock finns ett belägg på *naivt nok* från 1830, medan *klokt nok* faktiskt inte är belagt en enda gång under 1800-talet.

Också i det norska materialet finns flera exempel på att *nok*-frasen kanske snarare är annex än (integrerat) satsadverbial (jfr (9) ovan); observera ordföljden:

- (12) [...] men, selsomt nok! den lange, ensomme Forvaring hade næsten bragt henne til at glemme hendes Sprog
(Morgenbladet 1825)

Det ser alltså ut som om den tidiga utvecklingen av *underligt/klokt nok*-konstruktionen i norska i stort sett är parallell med utvecklingen i svenska. I svenska verkar det finnas aningen tidigare förstabelägg, även om man inte räknar med posesibelägget från 1776. Om man bortser från det belägget, är det tidigaste engelska belägget fortfarande äldre än det äldsta svenska (och norska), och dessutom bör man kunna hitta betydligt äldre engelska belägg om man gör sökningar i större engelska 1700-talskorpusar. Det finns därför ingen anledning att revidera den förmodan som delvis gjordes i Malmgren 2014, att den svenska, danska och norska (och kanske holländska) *underligt/klokt nog*-konstruktionen uppstått under påverkan av motsvarande engelska konstruktion med *enough*.

Avslutningsvis ska vi gå drygt ett halvsekel framåt i tiden och se efter hur det såg ut i norska tidningar under 1900-talets första decennium, när den första riksmålsnormalen fastställdes och man kan tala om ett självständigt norskt skriftspråk (vid sidan av nynorska). Det ser faktiskt ut som om Swan (1991) har rätt i sin förmodan, att *nok* ännu inte är särskilt produktivt som satsadverbbildande

suffix. Några få situationsvärderande satsadverb är väl företrädda, särskilt *merkelig(t) nog*, *underlig(t) nok* och *karakteristisk nok* (nästan 100 förekomster i tidningsmaterialet mellan 1900 och 1909), men de personvärderande tycks vara ganska sällsynta: exempelvis förekommer *klokt nok* inte en enda gång 1900–1909 och *naivt nok* 5 gånger. Det är därför kanske inte så konstigt att Western (1921) inte nämner *underligt/klokt nog*-konstruktionen.

Det är tydligt att de båda nationalbibliotekens jättekorporusar öppnar helt nya perspektiv för den språkhistoriska forskningen, i synnerhet (tack vare tidningsmaterialen) kanske forskningen kring de senaste 200–250 årens språkhistoria.

LITTERATUR, TRYCKTA ARBETEN

Malmgren, Sven-Göran. 2000. Projektet Det svenska ordförrådets utveckling 1800–2000. Utgångspunkter. Göteborg: Göteborgs universitet, Institutionen för svenska språket. (Rapporter från ORDAT. 1.)

Malmgren, Sven-Göran. 2002. *Faktiskt, förstås och många andra*. Om förändringar i det svenska formordssystemet under 1800- och 1900-talet. I *Språk och stil* 12 (ny följd), 97–146.

Malmgren, Sven-Göran. 2014. *Konstigt nog, klokt nog ...* Om en ganska ny typ av satsadverb i svenskan och andra språk. *Fint språk / Good Language. Festschrift till Lars-Gunnar Andersson*. Göteborg: Göteborgs universitet, Institutionen för svenska språket, 109–116. (Meijerbergs arkiv för svensk ordforskning. 41.)

ODS = *Ordbog over det danske sprog*. København 1918–56.

OED = *Oxford English Dictionary*. Oxford 1882–1928.

Ramat, Paolo & Davide Ricca. 1998. Sentence Adverbs in the languages of Europe. I Johann van der Auwera (red.), *Adverbial Constructions in the Languages of Europe*. Berlin & New York, 187–275.

SAG = Andersson, Erik, Staffan Hellberg & Ulf Teleman. 1999. *Svenska Akademiens grammatik*. Stockholm.

SAOB = *Ordbok över svenska språket utgiven av Svenska Akademien*. Lund 1893. (I maskinläsbar och sökbar form: OSA.)

Swan, Toril. 1991. Adverbial Shifts: Evidence from Norwegian and English. I Dieter Kastovsky (red.), *Historical English Syntax*. Berlin & New York, 409–438.

Western, August. 1921. *Norsk riksmålsgrammatikk*. Christiania.

WNT = *Woordenboek der Nederlandsche Taal*. 's-Gravenhage/Leiden 1882–1998.

ELEKTRONISKA MATERIAL

Kungliga bibliotekets korpus över svenska dagstidningar. Stockholm, <http://tidningar.kb.se>

Litteraturbanken. Göteborgs universitet: Institutionen för svenska språket, <http://litteraturbanken.se>

Nasjonalbibliotekets korpus över norsk litteratur. Oslo, www.bokhylla.no

Språkbanken. Göteborgs universitet: Institutionen för svenska språket, <http://spraakbanken.gu.se>

KONTAKT

Sven-Göran Malmgren
Prof. emeritus, Göteborgs universitet
sg.malmgren@gmail.com

STORE TEKSTKORPUS TIL BRUK I SPRÅKRÅDETS SPRÅKOBSERVASJON, SPRÅK- NORMERING OG SPRÅKRÅDGIVNING

ÅSE WETÅS

SAMMENDRAG

Artikkelen tar opp endringene tilgangen til store tekstkorpus har gitt i Språkrådets arbeid med språkoobservasjon og språknormering. Korpusene har gjort det mulig å studere språkbruk og formvalg i store tekstmengder. Faktisk språkbruk (usus) i normdannende tekster skal tillegges vekt i normeringen av begge de to norske skriftspråkene. Tilgang til store tekstkorpus bidrar til å heve kvaliteten og styrke legitimiteten på et arbeidsfelt som har stor oppmerksomhet i offentligheten, og der meningene ofte er både mange og sterke. Artikkelen ser særlig på rollen Leksikografisk bokmålskorpus har fått i den løpende normeringen av bokmålet.

[1] INNLEDNING

Denne artikkelen¹ vil belyse endringene som har skjedd i Språkrådets arbeid med normering og språkoobservasjon som et resultat av at vi etter årtusenskiftet har fått muligheten til å bruke store tekstkorpus i arbeidet vårt. Språkrådet har ansvar for å normere begge de to norske skriftspråkene, men artikkelen legger særlig vekt på normering av bokmålet, og på den viktige rollen ressurser som Leksikografisk bokmålskorpus (LBK) spiller i normeringsarbeidet. I tillegg vil artikkelen også komme nærmere inn på hvordan tilgang til store leksikalske databaser og ressurser bidrar til god kvalitet i arbeidet med andre av Språkrådets kjerneoppgaver. Første del av artikkelen vil være tilbakeskuende og gjøre rede for hvordan både organiseringen av det statlige normeringsarbeidet og den normeringsfaglige praksisen har endret seg fra årtusenskiftet til i dag. Artikkelen vil gjennom eksempler på nyere normvedtak vise hvordan LBK får sin anvendelse i Språkrådets løpende arbeid med å følge opp den norske språkpolitikken på normeringsfeltet.

[1] Artikkelen baserer seg på et foredrag til festsymposium for professor Ruth Vatvedt Fjeld i juni 2018. Den er en takk til Ruth både for det store arbeidet hun har lagt ned i å utvikle Leksikografisk bokmålskorpus, og for hennes mangeårige innsats for norsk leksikografi og leksikologi.

[2] BAKGRUNN

Språkrådet ble etablert i 2005 som statens fagorgan for språk, og rådet er en etat direkte underlagt Kulturdepartementet. Rådet har ingen selvstendige forskningsoppgaver, men forvalter den offentlige norske språkpolitikken gjennom oppgaver som enten er vedtektsfesta, har strategisk prioritet innenfor rammene av den gjeldende språkpolitikken, eller er formulerte i de årlige tildelingsbrevene fra overordnet myndighet. Språkrådet har direkte ansvar for arbeid med norsk, norsk tegnspråk og de nasjonale minoritetsspråkene kvensk, romanes og romani. Rådets oppgaver omfatter både arbeid med språk i offentlig sektor, språk i skole og høyere utdanning, språk i arbeids- og næringslivet og språk på kultur- og mediefeltet. I tillegg har Språkrådet strategisk ansvar for arbeid med norskspråklig fagterminologi og praktisk ansvar for løpende språkbobservasjon og normering av de to skriftspråkene bokmål og nynorsk. Rådet driver et omfattende kurs-, veilednings- og rådgivningsarbeid overfor både offentlige og private aktører på alle våre arbeidsfelt. I tillegg tar vi imot henvendelser fra enkeltpersoner som søker svar på språkspørsmål.

Språkrådets forgjenger var Norsk språkråd (1972-2004). Organisasjonen ble styrt av et faglig råd med 21 representanter for hvert av de to skriftspråkene bokmål og nynorsk, til sammen 42 medlemmer. Rådsmedlemmene representerte til sammen 30 ulike instanser, som omfattet både sentrale språkpolitiske aktører i Norge og relevante fag- og forskningsmiljøer. Norsk språkråd hadde i tillegg ei fagnemnd med fire medlemmer fra hver målform, og et sekretariat av fast ansatte medarbeidere. Norsk språkråd var delt inn i en bokmåls- og en nynorskseksjon, og organisasjonens hovedoppgaver var språknormering og språkrådgivning. Etter hvert fikk rådet også tilsynsoppgaver på mållovsfeltet og strategisk ansvar for utvikling av språkteknologi på norsk. I tillegg hadde rådet i sitt mandat å fremme toleranse og respekt for språk, og å bidra med kunnskap om norsk i offentligheten. I Norsk språkråds oppgaveportefølje lå dessuten ansvar for kvalitetssikring og godkjenning av lærebøker til bruk i skolen. Denne godkjenningsordningen ble avviklet i år 2000.

Språkrådet (2005-) har ansvar for å iverksette den norske språkpolitikken slik den er vedtatt av Stortinget, og på fullmakt fra overordnet departement. Dagens organisasjon har et styre med fire medlemmer oppnevnt av kulturministeren og én representant for de ansatte. Språkrådets styre har overordna økonomi- og resultatansvar for organisasjonen, og styret gjør også løpende normeringsvedtak i tråd med organisasjonens normeringsfullmakter. Språkrådet har i tillegg knyttet til seg fire fagråd der eksperter fra ulike samfunnsfelt er hentet inn for å gi råd og ny kunnskap på organisasjonens arbeidsområder. Ett av fagrådene, fagrådet for normering og språkbobservasjon,

er vedtektsfestet. En hovedoppgave for dette fagrådet er å arbeide med løpende språkobservasjon og språknormering. Jeg vil litt seinere i artikkelen vise fram et par konkrete eksempler på arbeidet til fagrådet for normering og språkobservasjon, for å illustrere hvordan metoder og fagressurser i dette arbeidet har endret seg i den digitale tidsalderen. Men først skal vi ta et nærmere blikk på statusen til og innholdet i dagens bokmålsnorm.

[3] 2005-NORMEN FOR NORSK BOKMÅL

Den gjeldende skriftspråksnormen for norsk bokmål ble vedtatt av Kulturdepartementet i 2004, med iverksettelse fra 2005. Den store, prinsipielle endringen som kom med 2005-reformen var at det gamle tonivåsystemet i bokmålet ble fjernet. Dermed forsvant systemet med hovedformer og sideformer (også kalt «klammeformer»), som hadde kommet inn med 1938-normalen. Systemet med sideformene hadde som sitt primære formål å lage et «venterom» for former på veg ut av normen eller på veg inn i normen, og det fungerte gjennom en periode i norsk normeringshistorie der det langsiktige målet var en gradvis tilnærming mellom de to målformene, med utgangspunkt i folkemålet. For nynorsk ble tonivåsystemet i normen opprettholdt fram til 2012, da dagens nynorskrettskriving ble innført. Opprinnelig var planen på tidlig 2000-tall at begge de to norske målformene skulle gå over til en norm med bare ett nivå i parallelle prosesser, men etter at flertallet i Norsk språkråds fagnemnd stemte mot etnivåsystem for nynorsk, samtidig som det kom tydelige og sterke innvendinger fra høringsinstansene da et forslag til ny nynorsknorm med bare ett normeringsnivå i 2002-2003 var på høring, ble denne endringen utsatt for nynorskens del. Den gamle tonivånormen levde derfor videre, men behovet for en ny og strammere nynorsknorm var tydelig. Språkrådet fikk etter en tid nytt mandat til å sette i verk arbeid med en større rettskrivingsendring for nynorsk. Dette arbeidet ledet fram til det som i dag har fått benevnelsen «2012-normalen for nynorsk».

Det er viktig å presisere at målet med å fjerne tonivåsystemet i bokmålsrettskrivingen i 2005 ikke var å fjerne all valgfrihet innenfor normen. Det var heller ikke et mål i seg selv å reversere resultatene av tilnærminglinja i språknormeringen gjennom 1900-tallet, jf. også Lars Vikør (2007: 202). Vikør skriver at mange tolket opphevingen av tilnærmingssparagrafen i den daværende *lov om Norsk språkråd* som at tilnærmingen mellom målformene skulle reverseres, men han påpeker at det faktisk ikke lå i Stortingets vedtak fra 2002. Selv om en del lite brukte former ble tatt ut av normen, er det fremdeles en god del variantformer i bokmålet, men fra 2005 har disse formvariantene vært formelt sidestilte, slik at de i prinsippet skal ha samme status i bruk. I praksis ser vi likevel at systemene for implementering av hele innholdet i de offisielle språknormene ikke

er gode nok, og mange store private normaktører (forlags- og mediehus) viser i liten grad fram den faktiske variasjonsbredden i bokmålet. Mange brukere kjenner derfor dessverre ikke til valgmulighetene de har innenfor normen, og en del valgfrie former er derfor mindre synlige i offentligheten enn det som er ønskelig.

Bokmålsnormen fra 2005 bygger på normeringsvedtak gjort av Norsk språkråd i perioden 2002-2004. Som nevnt forsvant en del lite brukte former ut av bokmålet på dette tidspunktet, samtidig som det ble tatt inn en del nye brukte former som enten hadde vært sideformer, eller som hadde ligget utenfor normen. I 2002 hadde § 1 i *lov om Norsk språkråd* blitt endret av Stortinget, slik at loven ikke lenger la til grunn at det langsiktige målet for språknormeringen var å fremme utviklingstendenser som førte bokmål og nynorsk nærmere hverandre. Fra tidlig 2000-tall blir de to norske målformene dermed normerte på selvstendig grunnlag og med utgangspunkt i egen tradisjon. Dagens retningslinjer for normering av bokmål og nynorsk presiserer samtidig at det i normeringen av de to målformene skal unngås unødige forskjeller i ortografiske prinsipper (f.eks. i stavemåten av lånord).²

Hvordan skaffet språknormererne på 1900-tallet seg innsikt i hvordan brukspråket så ut og hvilke former som var mye og lite brukte, og hvilke kilder brukte de? I arbeidet med 2005-normen for bokmål ble det i stor grad fremdeles brukt tradisjonelle metoder for å hente inn empiri om faktisk språkbruk. Et eksempel er de to større studiene Norsk språkråd på 1990-tallet bestilte av normuttrykket i norsk skjønnlitteratur (bokmål og nynorsk). I disse studiene ble 59 romantekster fra ulike tiår på 1900-tallet gjort maskinlesbare – noe som for så vidt representerte et visst framskritt i det empiriske grunnlagsarbeidet, sammenliknet med tidligere større rettskrivingsendringer. Maskinlesingen skulle gjøre det mulig å kartlegge formvariasjon i en større mengde tekster under ett. Lars Vikør (1995) skriver at de to parallellstudiene ble initierte som en konsekvens av motstand mot 1981-reformen for bokmål. Motstanden hadde utgangspunkt i at det ikke var gjort større empiriske undersøkelser av den faktiske utviklingen av bokmålet gjennom 1900-tallet. Det var Språkrådet som stod for utvalget av de 59 romantekstene, og for bokmål kom Egil Pettersens studie «Språknormering og forfatterne» i 1993. I denne studien undersøkte Pettersen ortografi og morfembruk i skjønnlitterære tekster fra ti utvalgte bokmålsforfattere fra hvert av årene 1937, 1957 og 1977. Lars Vikør (1995): «Rettskriving hos nynorskforfatarar» er en tilsvarende undersøkelse og analyse av 29 nynorskforfatteres tekster fra de samme tre tiårene.

Egil Pettersen skriver i forordet til sin studie at «Ideelt sett burde alle typer

[2] Jf. Språkrådets retningslinjer for normering, <https://www.sprakradet.no/globalassets/sprakavare/norsk/retningslinjer-for-normering-bokmalsversjon.pdf>

litteratur vært med i en undersøkelse som skulle være fullt ut representativ for språkbruken. Vi hadde gjerne sett at aviser, ukeblad og tidsskrifter også hadde fått plass i prosjektet. Men det ville ha sprengt alle rammer og ligget utenfor våre muligheter» (Pettersen 1993: 1). I sluttordet til studien peker Pettersen (1993: 163) på at materialet han har analysert, er spinkelt og vanskelig å trekke slutninger fra. Dette har selvsagt sammenheng med at et utvalg på til sammen 30 tekster er altfor lite til at man kan finne sikre og gode funn på alle undersøkelsesområder, samtidig som det blir svært synlig at materialet mangler nødvendig sjangerspredning. Pettersens og Vikørs studier er solide og grundige, men metoden med i hovedsak manuell kartlegging av ortografi og morfembruk er ekstremt arbeidskrevende. Det er samtidig viktig å understreke at de to studiene ikke hadde som intensjon å kartlegge særlig fremragende bruk av språket, men å undersøke hvilket gjennomslag rettskrivingsreformene i 1917, 1938 og 1959 hadde hatt i mye leste tekster. Med andre ord var forutsetningen for studiene ikke å bygge den videre språknormeringen direkte på de utvalgte forfatternes formvalg, men snarere å dokumentere om og hvordan språkbruken i de skjønnlitterære verkene hadde blitt påvirket av innholdet i de respektive rettskrivingsreformene.

Samme modell og undersøkelsesmetoder som hos Pettersen (1993) og Vikør (1995) lå også til grunn for to av i alt fire studier jeg selv leverte Språkrådet i 2001 (Wetås 2001a, b). Disse studiene kartla formvalg i ulike typer publisert og upublisert nynorsktekst, og de skulle utgjøre underlagsmateriale til det da pågående arbeidet med ny nynorsknorm. Selv om det på dette tidspunktet faktisk var utviklet noen relativt små digitale tekstkorpus for norsk, foregikk kartleggingsarbeidet stadig hovedsakelig i form av mer eller mindre manuell ekserpering fra lokalaviser, NOU-er og utvalgte sakprosaverk.

Her kan det skytes inn at det selvsagt også ville vært prinsipielt mulig å bygge språknormeringen på skriftspråksuttrykket til et utvalg fremragende forfattere, eller på skjønnet til noen utvalgte spesialister, men det ville representere en framgangsmåte og en normeringspolitikk som står i motsetning til det som har ligget til grunn for den norske språknormeringstradisjonen. Gjennom hele 1900-tallet var tilnærminga mellom bokmål og nynorsk på folkemålsgrunn det langsiktige formålet med språknormeringen. Dermed ligger det i sakens natur at mye brukte former i det norske folkemålet (talt norsk) skulle tillegges særlig vekt i normeringssammenheng, og at tonivåsystemet hadde en viktig funksjon i å få realisert dette målet.

Selv om tilnærminglinja mellom de to målformene i praksis alt var forlatt, førte Stortingets enstemmige vedtak fra 2002 om at bokmål og nynorsk skulle

normeres på egen grunn, og uten sideblikk til den andre målformen, til en endring i vektleggingen av normeringsprinsippene som har vært brukt for de norske skriftspråkene. En annen viktig endring i betingelsene for den skriftspråklige normeringen av norsk på 2000-tallet, er den radikale økningen i tilgang på store mengder data om faktisk språkbruk i ulike typer brukstekster. Den sistnevnte utviklingen kom raskt til å endre kravene til empirisk fundament for normeringsvedtak. I tillegg kunne åpne dataløsninger ha den gode tilleggs-effekten at det sikret allmenn tilgang til å kikke språknormererne i korta, og dermed gav åpne, tilgjengelige korpus språkbrukerne mulighet til å vurdere i hvilken grad retningslinjene for normering ble etterlevd av normererne. Når dette er sagt, er usus (språkbruk) bare ett av til sammen sju prinsipper for normering av bokmål (jf. nedenfor). Retningslinjene legger dermed ikke opp til at det skal være et entil-en-forhold mellom bruksfrekvens i normdannende tekster og faktisk representasjon i skriftspråksnormen.

[4] SPRÅKRÅDETS NORMERINGSFULLMAKTER OG NORMERINGSPRINSIPPER

Etter at Språkrådet var etablert, gav Kulturdepartementet i 2012 organisasjonen nye normeringsfullmakter. Fra dette tidspunktet har det offentlige normeringsoppdraget vært nedfelt i Språkrådets vedtekter. Paragraf 3 *Normering av bokmål og nynorsk* slår i første ledd fast at Språkrådet kan gi råd om staving og bøyning av tidligere unormerte ord og gjøre bindende normeringsvedtak når språkutviklinga tilsier det. I tredje ledd slås det fast at Språkrådet også kan vedta ny stavemåte og bøyning av tidligere normerte ord, dersom faktisk språkbruk eller systemhensyn tilsier det. Det ligger likevel klare avgrensninger i Språkrådets normeringsfullmakter. De gjelder nemlig bare for normering av enkeltord eller grupper av ord under systemnivå. Alle *gjennomgripende* endringer eller *systemendringer* i de to norske skriftspråksnormene er det Kulturdepartementet som skal godkjenne, etter forslag fra Språkrådets styre.

Språkrådets fagråd for normering og språkbobservasjon leverte i 2014 utkast til interne retningslinjer for normering av både bokmål og nynorsk, som styret i Språkrådet vedtok med virkning fra 1. mars 2015. Normeringen av bokmål og nynorsk følger etter disse retningslinjene samme hovedprinsipper, men med en litt ulik vektlegging for de to målformene. I normeringen av bokmålet skal følgende prinsipper vektlegges:

- (i) **Selvstendighetsprinsippet:** Målformen skal normeres på egen tradisjon og eget grunnlag, men ikke slik at det oppstår nye og unødvendige forskjeller mellom de to målformene når det gjelder ortografi, valg av norvagserte

former osv. Det er f.eks. ingen grunn til at ord som *kviss* og *nikab* skal skrives ulikt på de to målformene.

- (ii) **Stabilitetsprinsippet:** Normen skal ligge fast over tid, og det kan bare gjøres mindre justeringer. Dette skal gi språkbrukerne tilgang til en langsiktig og stabil norm som skal være enkel å gjøre seg kjent med og lett å kunne bruke over tid.
- (iii) **Bruksprinsippet (ususprinsippet):** For bokmål er det vesentlige grunnlaget nettopp skriftspråket slik det kommer til uttrykk i tekster skrevet på bokmål, og da særlig tekster som regnes som normdannende. Vektleggingen av «normdannende tekster» kan problematiseres, for en rigid bruk kan gjøre normeringsprosessen sirkulær. Språkrådets fagråd for normering og språkoobservasjon startet i 2019 derfor et arbeid med å presisere normeringsprinsippene på dette punktet.
- (iv) **Stramhetsprinsippet:** Det skal ikke være et selvstendig mål med normeringen å stramme inn eller utvide valgfriheten i bokmålsnormen. På lang sikt sier retningslinjene at det likevel bør vurderes om normen kan strammes inn *noe*. Her gjelder det likevel at Språkrådet bare har en avgrensa normeringsfullmakt og ikke på eget initiativ kan gjøre mer gjennomgripende endringer.
- (v) **Enkelhetsprinsippet:** Det bør være færrest mulig bøyingsmønstre, og normeringen skal heller gjelde klasser eller grupper av ord, snarere enn enkeltord. Det er ikke vanskelig å se at dette prinsippet kan støte mot bruksprinsippet, og her legger retningslinjene faktisk inn en avgrensing i favør av bruksfrekvens. Hvis høyfrekvente ord har et avvikende bøyingsmønster, bør dette mønsteret stå ved lag.
- (vi) **Tradisjonsprinsippet:** Både bokmål og nynorsk har mer enn hundre års skrifttradisjon bak seg, og det gjør at mye leste eldre tekster også bør være med i det normdannende tekstgrunnlaget for skriftspråket. Dette prinsippet knytter seg tett til stabilitetsprinsippet.

For nynorsk gjelder i tillegg **talemålsprinsippet**. Det slår fast at ved innstramning av normen skal formene som har best grunnlag i talemålet stå igjen, mens det ved eventuelle utvidinger av normen helst skal velges former som har et solid talemålsgrunnlag.

For at normeringsprinsippene skal kunne praktiseres på en god måte, er det

nødvendig for Språkrådet å ha tilgang til store mengder relevant empiri som kan gi informasjon om hvordan språket i bruk faktisk ser ut, og hvordan det har endret seg over tid. Av de nevnte normeringsprinsippene for norsk bokmål forutsetter i alle fall både bruksprinsippet, enkelhetsprinsippet og tradisjonsprinsippet tilgang til kvalitativt og kvantitativt solid empiri.

Min egen bakgrunn fra leksikografisk arbeid og som medlem i nemnda som i 2011 leverte forslag til ny nynorskrettskriving, har vist meg veldig klart at man i normeringssammenheng vanskelig kan stole på introspeksjon og egen intuisjon når det gjelder hvilke formvarianter som er i faktisk bruk, og hvilke varianter som dominerer i bredt sammensatte tekstmengder. Det er bare store tekstsamlinger som vil kunne gi sikker innsikt på dette punktet, og da vil Leksikografisk bokmålskorpus (LBK) og andre store digitale tekstsamlinger være helt nødvendige arbeidsverktøy i den løpende normeringen av de to norske skriftspråkene.

[5] BRUKEN AV LEKSIKOGRAFISK BOKMÅLSKORPUS (LBK)³ OG ANDRE STORE DIGITALE KORPUS

Språkrådets styre står for de endelige vedtakene i normering av enkeltord i bokmål og nynorsk. Språkrådets fagråd for normering og språkoobservasjon leverer saksunderlag og anbefalinger til ny eller endret normering til styret, og saksunderlagene gjør grundig greie for det empiriske grunnlaget for fagrådets anbefalinger. Fagrådet tar løpende opp normeringsspørsmål som trenger avklaring, men de arbeider i tillegg også tematisk. De tre siste årene har Språkrådets fagråd for normering og språkoobservasjon arbeidet med å kartlegge behovet for nedskjæring i genusvariasjonen innafor begge de to skriftnormene. På oppdrag fra Språkrådet gjorde forsker Philipp Conzett ved Universitetet i Tromsø i 2016 en empirisk studie av genusvariasjon i norsk (Conzett 2017). I dette arbeidet var LBK og Det nynorske tekstkorpuset, begge utviklet for leksikografiske formål ved Universitetet i Oslo, viktige kilder til informasjon om bruken av de ulike genusvariantene. I 2018 satte fagrådet dessuten i gang et prosjekt for å undersøke nærmere statusen til norvagiseringsvedtak eldre enn 25 år. Også i dette arbeidet vil det være helt avgjørende å ha tilgang til store moderne tekstsamlinger som kan vise i hvilken grad norvagiserte skrivemåter er tatt i bruk.

For å vise den konkrete bruken av LBK og andre sentrale og nyere datakilder i normeringsarbeidet vårt, vil jeg vise fram underlagsmateriale fra to normeringssaker som ble behandlet av Språkrådets styre i 2016 og 2017, og der styret

[3] Leksikografisk bokmålskorpus (LBK) er et balansert leksikografisk korpus på ca. 100 mill. ord, og det inneholder tekst fra perioden 1985-2013. Sammensetningen av korpuset skal være representativ for både ulike typer tekster og mengden av disse teksttypene en gjennomsnittlig leser møter. I presentasjonen av korpuset (<https://www.hf.uio.no/iln/om/organisasjon/tekstlab/prosjekter/lbk/>) går det fram at 95 % av tekstene som inngår, er redaksjonelt bearbeida og publiserte av forlagshus og mediehus.

gjorde vedtak om endret rettskriving. Det må presiseres at eksemplene omfatter bare to del saker i det som er større sakskomplekser, og de utgjør slik sett en liten del av det samla normeringsarbeidet til rådet.

[5.1] Normeringen av substantivene *mormon* og *mormoner*

Både i bokmål og nynorsk var formen *mormon* eneform fram til 2017, men vi ønsket å vurdere om normen burde justeres. Langformen *mormoner* var rett nok inne i bokmålsrettskrivingen fra 1984 til 1986, men ble tatt ut igjen, som en konsekvens av at Jesu Kristi kirke av de siste dagers hellige *selv* bare brukte kortformen. Da fagrådet for normering og språkoobservasjon i 2017 behandlet saken, mente de at det ikke kunne legges avgjørende vekt på hva trossamfunnet selv foretrekker. Ønsker fra enkeltaktører ligger utenfor settet av normeringsprinsipper som gjelder for norsk, og kan ikke uten videre tas hensyn til.

Søk i korpus gir interessante funn om bruken av kortformen *mormon* (KF) og langformene *mormoner* og *mormonar* (LF), jf. tabell 1 (nedenfor).

		KF	LF	n ⁴
nb.no (bøker)	Nasjonalbibliotekets korpus av digitaliserte bøker og aviser, 1960–	47 %	53 %	1 357
nb.no (aviser o.a.)	Norske aviser 1998–2015 (bokmålsdelen)	28 %	72 %	1 917
Norsk aviskorpus	Balansert korpus av bokmålstekster, 1985–2013	25 %	75 %	267
Leksikografisk bokmålskorpus	Nynorsktekster fra ulike sjangre fra 1800-tallet til i dag	67 %	33 %	48
Nynorskkorpuset	Norske nettsider 2009–2010	76 %	24 %	17
NoWaC		30 %	70 %	139

TABELL 1: Oversikt over korpus treff på søk «mormon», «mormonar» og «mormoner»

Her kommer en tydelig forskjell til uttrykk mellom LBK og nynorskkorpuset på den ene sida og de andre korpusene på den andre sida. Begge de to leksikografiske korpusene har en sterk overvekt av kortformen *mormon*. En forklaring på at disse to korpusene skiller seg ut, kan være at en større andel av teksttilfanget der er normert etter forlags- og mediehusnormer enn tilfanget i de andre korpusene (og da særlig NoWaC). Med den bakgrunnen kan det gi ekstra styrke til de totale tallene at både LBK og Nynorskkorpuset likevel omfatter en relativt stor mengde forekomster av en formvariant som ligger utenfor den offisielle skriftspråksnormen. Denne entydige empiriske dokumentasjonen var en del av bakteppet for det seinere vedtaket om å ta opp skrivemåten *mormoner* ved siden

[4] n = totalt antall korpus treff. Tabellen er hentet fra saksframlegget til styresak 32/17, behandlet av Språkrådets styre 2.6.17.

av *mormon* som en del av den offisielle normen.

[5.2] Normeringen av ordene *dølahest*, *dølafe* og *dølaku*

I bokmålsnormen var *dølehest*, *dølefe* og *døleku* (med *e*-fuge) eneformer fram til 2016. Språkrådets fagråd for normering og språkoobservasjon så nærmere på denne ordgruppa, og søk i korpus gav også her interessante funn.

		<i>dølahest</i>	<i>dølehest</i>
nb.no (bøker)	Nasjonalbibliotekets korpus av digi-	55 %	45 %
nb.no (aviser o.a.)	taliserter bøker og aviser, 1960–	60 %	40 %
Retriever	Norske aviser og andre media	60 %	40 %
Norsk aviskorpus	Norske aviser 1998–2015	48 %	52 %
Leksikografisk	Balansert korpus av bokmålstekster	56 %	44 %
bokmålskorpus	fra 1985 til i dag		
NoWaC	Norske nettsider 2009–2010	42 %	58 %
Google	Google.no	83 %	17 %

TABELL 2: Oversikt over korpustreff på søk «*dølahest*» og «*dølehest*»⁵

Også for disse ordene, i tabell 2 representert ved *dølahest* og *dølehest*, viste det seg at formvarianten som ikke var en del av den offisielle skriftspråksnormen, var svært mye brukt. Overraskende nok hadde den ikke-normerte varianten med *a*-fuge faktisk flere treff både i LBK og i Nasjonalbibliotekets digitale bokhylle enn den normerte formen med *e*-fuge.

I normeringssaker er det også andre forhold enn bruksfrekvens som skal vurderes, ikke minst hensynet til stabilitet og system (jf. oversikten over normeringsprinsipper). Det endelige vedtaket i saken ble derfor ikke å fjerne de tradisjonelle formvariantene med *e*-fuge (som faktisk også viste seg å ha en god del bruk bak seg) til fordel for de mer brukte med *a*-fuge. Løsningen ble i stedet å ta de sistnevnte formene (*dølahest*, *dølafe*, *dølaku*) inn i normen, og la dem stå som likestilte med *dølehest*, *dølefe* og *døleku*.

[6] BRUK AV LBK TIL SPRÅKOBSERVASJON OG NORMERING

Sammenlignet med andre store tekstkorpus er en umiddelbar fordel med Leksikografisk bokmålskorpus at det er representativt og vektet, og at det omfatter tekst fra en hel rekke ulike sjangre. Jeg vil i denne sammenhengen særlig løfte fram sjangeren TV-teksting. Denne teksttypen representerer en vesentlig kilde til kunnskap om skriftlig norsk for mange som i liten grad leser andre typer trykt eller digital tekst. I språknormering er derfor dette en viktig

[5] Tabellen er hentet fra saksframlegg til styresak 40/16, behandlet av Språkrådets styre 23.8.16.

teksttype. LBK dekker tidsperioden fra 1985 til 2013, og vektingen i korpuset gir i det hele tatt et godt bilde av hvilke normframstillinger språkbrukerne møter i det daglige. Tekstkorpuset viser selvsagt ikke hvilke formvarianter språkbrukere flest foretrekker når de skal produsere egne tekster, men det gir god og kvalitetssikret dokumentasjon på hvilke operative normer forlag, mediehus og andre store tekstprodusenter bruker i sine tekster, og dermed hvilke formvarianter språkbrukerne faktisk eksponeres for. LBK dokumenterer at dette i stor grad viser seg å være former som ligger i det som ofte omtales som den «konservative» eller «moderate» enden av bokmålsnormen, slik også Kjersti W. Kola har dokumentert i masteroppgaven sin fra 2014. Det er selvsagt grunn til å tro at disse normframstillingene i tur preger språkbrukernes oppfatninger av hva som er rammene for det korrekte, normerte språket. Svaret er likevel at vi faktisk vet veldig lite om hvilke former språkbrukerne foretrekker eller oppfatter som del av normen, annet enn at det er åpenbart at språkbrukere velger ulike normuttrykk i ulike brukssituasjoner. For eksempel viser en befolkningsundersøkelse Opinion utførte for Språkrådet i 2016 at svært mange voksne språkbrukere rapporterer å bruke dialekt når de skriver i sosiale medier. Denne tendensen er særlig uttrykt i responsen fra språkbrukere i Nord-Norge og Midt-Norge, den er tydeligere for kvinner enn for menn, og den gjelder yngre språkbrukere i større grad enn eldre. Det skal skytes inn at undersøkelsen er basert på egenvurdering av skriftspråklig praksis, så hva den enkelte respondenten legger i det å skrive på dialekt, kan være ulikt. Vi ønsker oss mer kunnskap på dette området, men studier av unormert og upublisert tekst produsert av «vanlige» språkbrukere (i motsetning til profesjonelle språkbrukere) er vanskelig å gjennomføre. En hovedgrunn er at rettighetsproblematikk og personvernlovgivning hindrer innhøsting og bruk av tekst fra sosiale medier. En forskningsbasert undersøkelse av språkbrukernes praktiske normkunnskaper og (sjangerbestemt) normvariasjon står likevel høyt på Språkrådets ønskeliste.

For Språkrådet er LBK av stor nytteverdi i det løpende arbeidet med språkbobservasjon og normering, og det er også verdifullt i arbeidet med språkrådgivning. I tillegg brukes LBK i arbeidet med vedlikehold og utvikling av *Bokmålsordboka* og *Nynorskordboka*, som Språkrådet fra 2016 eier sammen med Universitetet i Bergen (UiB). I et prosjekt med oppstart i 2018 og tentativ slutføring i 2023 skal de to tvillingordbøkene gjennomgå en fullstendig revisjon. Prosjektet er finansiert av Kulturdepartementet, med bidrag fra UiB og Språkrådet. Viktige delprosjekter i dette arbeidet vil være kritisk gjennomgang og revisjon av alle orddefinisjoner og brukseksempler. Et korpus spesielt utviklet til bruk i leksikografisk

forskning er av stor verdi i dette arbeidet, og det utgjør en sentral ressurs i revisjonsprosjektet. Vi vet at språket på noen områder endrer seg raskt, og til arbeid med Språkrådets rettskrivningsordbøker skulle vi gjerne sett at de viktige leksikografiske korpusressursene for norsk løpende ble bygget ut med nytt teksttilfang.

[7] OPPSUMMERENDE MERKNADER

Språkrådet er et forvaltningsorgan uten forskningsakkreditering. For å kunne utføre arbeidsoppgavene våre, må vi ha inngående kjennskap til språk og språkendring. I dette arbeidet er vi avhengige av tett kontakt med relevante forskningsmiljøer innenfor fagfelt som leksikografi og leksikologi, onomastikk og terminologi. Det er mye vi ønsker mer kunnskap om. Er det for eksempel endring i den skriftspråklige bruken av *han* og *ham* i bokmål? Opptrer objektsformen *dykk* stadig oftere på subjekt plass i tekster på nynorsk, og hvordan kan det i så fall forklares? Har utbredte trekk fra norsk dialekt syntaks som «*Hvem bil?*» gjennomslag også i skrift? Hva er utsiktene for at et nytt kjønnsnøytralt personlig pronomen *hen* vil etablere seg i skriftspråket? Dette er bare et lite utvalg eksempler på spørsmål der det er viktig for Språkrådet å ha tilgang til relevant, oppdatert empiri. Uten tilgang til store og kvalitetssikra korpusressurser vil det være tilnærmet umulig å få oversikten vi trenger over hvordan språket i bruk faktisk tar seg ut, både når det gjelder ortografi, morfologi og syntaks.

Alle Språkrådets arbeidsfelt er avhengige av god tilgang til forskning og forskningsinfrastruktur utviklet og forsket fram av relevante forskningsmiljøer. Vi trenger dessuten en tett og god dialog med forskningsmiljøene for å utføre forvaltningsoppgavene våre på beste vis. Språkrådet ser derfor med bekymring på at den systematiske forskningen i det norske språkets leksikon lever en pressa og utrygg tilværelse ute i universitets- og høgskolesektoren. For forskningsfeltene leksikografi og leksikologi vurderer vi at situasjonen nå er prekær. Det samme gjelder for de leksikalske «nabodisiplinene» terminologi og stedsnavnforskning. Vi ser dessuten at det ikke bare er forskningsfeltene i seg selv som er under press, men at det også er svært krevende å skaffe nødvendige ressurser til å opprettholde og videreutvikle relevant forskningsinfrastruktur på de nevnte fagområdene. Disiplinene deler skjebne med en rekke andre såkalte nasjonale empirier på kulturfeltet, og det er nærliggende å se utviklingen i sammenheng med finansieringsmodellene i sektoren. Språkrådet engasjerer seg i disse spørsmålene for å sikre at det norske samfunnet og alle andre som ønsker det, skal ha tilgang til relevant infrastruktur og forskningsressurser for norsk språk. For at vi skal kunne dokumentere hvilken relevans dette har for oss som fagorgan, forsøker vi også selv å løfte fram og synliggjøre den konkrete

nytteverdien Språkrådet har av både fagkompetansen og de enkelte leksikalske infrastrukturressursene fra forskningssektoren. I dette bildet er Leksikografisk bokmålskorpus én av flere tunge og viktige ressurser, og korpuset bidrar både til å sikre god datakvalitet og gjøre Språkrådets arbeid med språkobservasjon og språkrådgivning mer effektivt.

LITTERATUR

- Conzett, Philipp. 2017. *Genusvariasjon i norsk skriftspråk*. Ei undersøking av genusbruken ved substantiv som er normerte som maskulinum og nøytrum. (Rapport utarbeidet for Språkrådet.) <https://septentrio.uit.no/index.php/SapReps/article/view/4077>.
- Kola, Kjersti W. 2014. *Bokmålsbruk – hvorledes/hvordan/åssen og hvorfor? Om bruken av morfologiske og ortografiske varianter i bokmålsnormalen*. Masteroppgave i nordisk, særlig norsk, språkvitenskap, Universitetet i Oslo. <https://www.duo.uio.no/bitstream/handle/10852/40996/Kola-Master.pdf?sequence=1&isAllowed=y>.
- Pettersen, Egil. 1993. *Språknormering og forfatterne. Ortografi og morfembruk hos ti bokmålsforfattere for hvert av årene 1937, 1957 og 1977*. Eigenproduksjon nr. 49/93. Bergen: Nordisk institutt, Universitetet i Bergen.
- Språkrådets retningslinjer for normering av bokmål og nynorsk*. <https://www.sprakradet.no/globalassets/spraka-vare/norsk/retningslinjer-for-normering-bokmalsversjon.pdf>
- Vikør, Lars. 1995. *Rettskriving hos nynorskforfattarar*. Norsk språkråds skrifter nr. 3. Oslo: Norsk språkråd.
- Vikør, Lars. 2007. *Språkplanlegging*. Prinsipp og praksis. Oslo: Novus forlag.
- Wetås, Åse. 2001a. Undersøking av formbruken i ulike typer upublisert materiale på nynorsk. (Rapport utarbeidet for Norsk språkråd.) <https://www.sprakradet.no/localfiles/nyno02awe3.pdf>
- Wetås, Åse. 2001b. Undersøking av formbruken i ulike typer publisert materiale på nynorsk. (Rapport utarbeidet for Norsk språkråd.) <https://www.sprakradet.no/localfiles/nyno02awe2.pdf>

KONTAKT

Åse Wetås

direktør, Språkrådet

ase.wetas@sprakradet.no

LEKSIKOGRAFISK BOKMÅLSKORPUS (LBK) – BAKGRUNN OG BRUK

RUTH VATVEDT FJELD, ANDERS NØKLESTAD OG KRISTIN HAGEN

SAMMENDRAG

Denne artikkelen er en introduksjon til Leksikografisk bokmålskorpus (LBK). Vi starter med en historisk oversikt over ordboksarbeid som er utført for norsk språk, og forklarer bakgrunnen for at LBK ble bygd opp på den måten det ble. Deretter gir vi en oversikt over innholdet i korpuset, før vi til slutt viser hvordan man kan søke i korpuset ved hjelp av korpussøkeverktøyet Glossa.

[1] INNLEDNING

Leksikografisk bokmålskorpus (LBK) er et representativt, vektet korpus laget for leksikalsk utforskning av moderne bokmål. LBK inneholder tekster fra perioden 1985 til 2013, i alt omlag 100 millioner ord. Korpuset er bygd opp ved Universitetet i Oslo og lagt inn i søkegrensesnittet Glossa, som er administrert av Tekstlaboratoriet ved Institutt for lingvistiske og nordiske studier på Humanistisk fakultet. Korpuset er fritt tilgjengelig for språkforskning, det kreves kun at man logger seg inn og godkjenner et sett av bruksregler. Prosjektet ble ledet av professor Ruth Vatvedt Fjeld, ved daværende avdeling for bokmålsleksikografi ved Institutt for lingvistiske og nordiske studier, Universitetet i Oslo.

Korpusets design og oppbygning er spesielt tilrettelagt for utforskning av leksikalsk og morfologisk variasjon i moderne norsk bokmål, og er også et godt materiale for å dokumentere endringer i bruken av bokmålsordforrådet over en periode på tretti år. Korpuset er bygd opp etter mønster fra blant annet *Den Danske ordbogs elektroniske korpus* og *British National Corpus*.

Denne artikkelen er tredelt: I kapittel 2 gis det en historisk oversikt over leksikografi i Norge og Norden. Her gjør vi også rede for behovet for et elektronisk, leksikografisk innrettet tekstkorpus av en viss størrelse.

I kapittel 3 beskriver vi en del hindere og problemer som oppsto i arbeidet med å bygge et balansert korpus, og hvilke løsninger som ble valgt.

Kapittel 4 gir en praktisk og instruktiv innføring i bruken av korpusets mange søkemuligheter og hvordan man kan velge ut subkorpus for spesialundersøkelser og få resultatene vist og eventuelt lastet ned i ulike formater.

[2] BAKGRUNN OG METODEVALG

[2.1] *Materiale for tradisjonelt ordboksarbeid*

De fleste ordbøker for norsk språk er skrevet uten en systematisk oppbygd ordboksbase som materiale (jf. Nordisk leksikografisk ordbok, 1997:199), dvs. man har ikke undersøkt tekster etter en gjennomtenkt plan for å finne hvilke ord som skal dokumenteres i ordboka, men har samlet ord og uttrykk etter relativt tilfeldige metoder. Det betyr i praksis at ordbøkene ikke er redigert ut fra et tekstmateriale som er metodisk innsamlet for å dekke et visst ordforråd. Man trenger for eksempel et annerledes tekstmateriale for å lage spesialordbøker enn om man vil lage allmennordbøker som skal dekke hele det ordforrådet en språkkultur normalt bruker.

Tradisjonelt har norske leksikografer tatt som utgangspunkt de oppslagsordene de fant i eldre ordbøker, og lagt til mer eller mindre tilfeldig registrerte nye ord eller uttrykk som de mente manglet. Metoden var å notere seg ord man kom på, leste eller hørte, som man så skrev ned på sedler sammen med diverse andre opplysninger. Sedlene ble så systematisert alfabetisk i arkiver som utgjorde grunnlaget for redigering av den planlagte ordboka. Et slikt materiale ble selv sagt svært preget av den eller de som hadde samlet sedlene.

Et annet problem var at de eksisterende ordene ikke ble undersøkt for bruks- eller betydningsendring, noe som skjer jamt og trutt med ordforrådet i alle levende språk.

Et tidlig unntak fra denne tradisjonen i Norge er *Norsk riksmålsordbok* (Knudsen og Sommerfelt (red.) 1937-1957), som hadde som mål å dokumentere det norske litterære skriftspråket. Utgangspunktet var den norske delen av Brynhildsens *Norsk-engelsk ordbok*, men redaksjonen ble fort klar over at den ble for knapp, og de brukte også andre tilgjengelige ordbøker i tillegg til systematisk å ekserpere mange utvalgte norske forfatterskap som tilleggsmateriale. Etter hvert hadde redaksjonen en omfattende beleggssamling fra mønstergyldige tekster skrevet på norsk riksmål, blant annet ble Ibsens ordforråd godt dokumentert. Resultatet ble en normativ og vel dokumenterende ordbok over det norske litteraturspråket.

De fleste andre norske ordbøker er kommet i stand på den tradisjonelle måten, bygd på mer eller mindre tilfeldig innsamlet grunnlagsmateriale i tillegg til tidligere ordbøkers ordforråd. Det kan bli gode ordbøker av det, men svakheten ved en slik metode er at ordbøkene lett kan dokumentere språklige avvik eller nyheter i stedet for det vanlige og mest brukte ordforrådet, siden det er det spesielle eller avvikende som det er lettest å merke seg. (Ifølge den danske professor i leksikografi Henning Bergenholtz (Bergenholtz, 1996:6) kan resultatet i verste

fall bli en samling av perversiteter.) Grunnen til det er at de dagligdagse ordene og uttrykkene i det språket en er flytende i, lett overses, og dermed kan de komme til å mangle i ordbøkene. Særlig kan metoden være farlig om ordboksforfatteren har et personlig mål eller en sterk ideologi med arbeidet sitt. En av de mest kritiserte ordbøkene i så måte er Arakins russisk-norske ordbok (1963), som omtales som notorisk upålitelig (Nesset & Trosterud 205:273). Arakins mål var å lansere ei ordbok med størst mulig lemmautvalg, og redaktøren komponerte derfor selv mange udokumenterte, men tenkbare sammensetninger. Av ideologiske grunner kom nok også Knud Knudsen til å konstruere mange udokumenterte ord i norsk i den ordboka han selv regnet som sitt storverk, *Unorsk og norsk, eller fremmedordenes avløsning* (Knudsen 1881), der målet var å gi alternativer til fremmedord i den norske språket. Som kjent fikk svært få av hans forslag innpass i norsk språk.

Beleggsamlinger der en og samme ekserptor kan ha bidratt med flere hundre registreringer, vil uvegerlig også medføre en skjevhet i ordutvalg og bruken av ordene. Svært ofte er det ordboksredaktørens eget ordutvalg som fører til skjevhet eller feil beskrivelser og merking. Det er også lett å finne eksempler på bruksmarkeringer vi er uenige om i mange ordbøker.

Alle ordbøker må gjøre et utvalg av ord som skal beskrives, et såkalt lemmautvalg. Det fins internasjonalt noen få uttømmende ordbøker som er laget manuelt over berømte tekstsamlinger eller viktige forfatterskap i såkalte tesauruser eller konkordanser. I tesauruser lemmatiserer man med stor nøyaktighet hvert eneste ord som forekommer i de utvalgte tekstene og utarbeider definisjon eller forklaring av bruken av dem, og ordene blir vanligvis ordnet etter begrep, altså etter semantisk betydning, ikke etter alfabetet. Konkordanser, derimot, er ordnet alfabetisk, og hvert ord har henvisning til hvor det er brukt i den teksten man har gransket. Både tesauruser og konkordanser var før dataalderen en meget tidkrevende og kostbar dokumentasjon av ordforråd, og ble bare gjort med svært viktige tekster, som regel historiske eller religiøse skrifter. For norsk fins bare *Norsk bibelkonkordans* fra 1907 som er laget på den måten, med en oversikt over 70 000 av de ordene som var brukt i den første norske Bibelen fra 1904.

[2.2] Nye muligheter for leksikografien i dataalderen

I siste halvdel av 1990-tallet forsøkte mange leksikografer å finne nye og bedre metoder for ordbokskrivning, og med framveksten av elektronisk databehandling av tekst ble det plutselig mulig å behandle store mengder data med letthet. Da datamaskiner ble tatt i bruk i ordforskningen, ble det en faglig revolusjon for leksikografene, og det ga faget en helt ny status. Tidligst ute med bruk av edb (elektronisk databehandling) i norsk leksikografi var sannsynligvis professor

Harald Noreng, som i samarbeid med daværende NAVFs edb-senter for humanistisk forskning i Bergen i perioden 1975-1986 utarbeidet en fullstendig oversikt over ordforrådet i Ibsens verker (Noreng 1993). Det var et banebrytende arbeid i norsk leksikografi.

I dataalderen ble det mulig å samle tekster i søkbare databaser, kalt elektroniske tekstkorpus. Utbredelsen av moderne korpusleksikografi begynte med John Sinclairs forelesninger og studier utover 1980-tallet, og hans lærebok *Corpus Concordance Collocation* fra 1991 fikk stor betydning verden over. Internasjonalt begynte mange ordboksredaksjoner å lage egne, søkbare tekstsamlinger som grunnlag for sine ordbøker. Med moderne elektroniske tekstkorpus kunne man lage både konkordanser og tesauruser ut fra et nesten uendelig stort tekstmateriale, noe som åpnet opp for en helt ny æra i ordboksarbeidet. Sinclair hadde lenge hevdet at man måtte ha store tekstsamlinger for å kunne si noe viktig eller sikkert om språk i det hele tatt, og når det gjaldt leksikografi, formulerte han det slik: «It is, therefore, necessary to have access to a large corpus because the normal use of language is highly specific, and good representative examples are hard to find.» (1991:101). Han kritiserte ellers de tradisjonelle ordbøkene for bare å være egnet for resepsjon, ikke for produksjon. Produksjonsordbøker krever mer dyptpløyende semantisk beskrivelse av de leksikalske enhetene, og mer detaljert grammatisk informasjon om bruken av dem. Det var særlig konstruerte eksempler som Sinclair var motstander av, han mente at bare ekte brukseksempler kunne anvendes som dokumentasjon på ord i bruk (Krishnamuthy 2008:237); med hans egne ord: «usage cannot be thought up - it can only occur» (Sinclair 1987:3). Språklig evidens kan man bare finne enten ved introspeksjon eller observasjon, og på leksikalsk nivå er bare observasjon brukbart, hevdet han. Sinclair mente at leksikografi først og fremst var en deldisiplin av språkteknologien. Han demonstrerte hvordan man kunne lage ordbøker ved hjelp av språkteknologisk metode i Cobuild-prosjektet *Looking up!* (Sinclair 1987). Rapporten fra dette prosjektet havnet i posthylla som en gave til en stipendiat ved det daværende Leksikografisk institutt, som seinere ble prosjektleder for LBK. Den boka virket som en indirekte oppfordring: norsk leksikografi har stort behov for et skikkelig leksikografisk korpus, og rapporten hadde stor betydning for at arbeidet med LBK ble igangsatt.

Men det er ikke bare å sette i gang og samle inn tekster og tro man lager et nyttig leksikografisk korpus. Hovedproblemet med å bruke korpus som materiale for ordboksarbeid er at man må være svært bevisst på hvilke tekster man anvender, hvor store tekstmengder man trenger og hvordan tekstmassen bør være sammensatt for å få fram et så dekkende bilde som mulig av et språks ordforråd og hvordan det brukes i autentisk språkbruk. Utvelgelsesprosessene her

er avgjørende for kvaliteten på ordbøker som lages med korpus som materiale. Samtidig med at det ble utviklet metoder for å bygge opp store tekstkorpus, vokste innsikten i statistiske metoder brukt i språkforskningen. Særlig innen leksikografien har det blitt utviklet gode og brukervennlige programmer, som Sketch Engine (Kilgarriff et al. 2004) og den nordiske DeepDict analysis (Bick 2010). Ved hjelp av et sett algoritmer kan man nå analysere store tekstmengder automatisk og regne ut hvilke fraser og sammensetninger som er vanlige, og hvilke som er uvanlige eller må regnes som feil. Særlig nyttig er slike hjelpemidler ved kartlegging av et språks fraseologi.

[2.3] Korpusbasert leksikografi i Norden

Den danske ordbog (DDO) ble laget med et systematisk sammensatt tekstkorpus på 40 millioner ord som materiale. Ordboka kom ut i 1995 som den første korpusbaserte ordboka i Norden. Alle ord som forekom i korpuset, ble vurdert som kandidater for ordboka. Etter en grundig utvelgelsesprosess i redaksjonen ble de aktuelle ordene lemmatisert, definert og belagt med levende og autentiske eksemplere fra korpuset. Å redigere en ordbok ut fra slikt materiale var da noe helt nytt i nordisk leksikografi, og det lå en annen tenkning bak selve redigeringsarbeidet enn slik man hadde gjort ved de tradisjonelle ordbøkene. Leksikografi ble med disse metodene ikke sett på som en gudbenådet samleroppgave, som man tidligere tenkte om fagfeltet. Det var særlig den første internasjonale læreboka i leksikografi som dokumenterte dette fagsynet: *The art and craft of lexicography* (Landau 1984).

De nye datamaskinelle metodene bød på en mer vitenskapelig og nøyaktig utforsket presentasjon av et språks ordforråd og hvordan det fungerte i reell bruk. For ordboksredigering er korpus selvsagt bare ment som et tryggere grunnlag og hjelpemiddel, både for lemmaseleksjon, definisjon og bruksmarkeringer. En redaksjonell vurdering av enhver opplysning i en ordbok vil naturligvis aldri bli overflødig. Men grunnlaget vurderingen tas på, er mye bedre enn før.

Metodene er beskrevet i seinere nordiske lærebøker i leksikografi, som Bo Svenséns *Handbok i leksikografi* (2005), Sven Görans *Malmgrens Svensk lexikologi – Ord, ordbildning, ordböcker och orddatabaser* (1994) og Fjeld og Vikørs *Ord og ordbøker* (2008). Språkdata ved Göteborgs universitet var tidlig ute med systematiske ordforrådsstudier, med Sture Allén og den store forskergruppen rundt ham, som utførte korpuslingvistiske undersøkelser helt tilbake på 1960-tallet. Det fins mange store og gode korpus over svensk språk, men det har ennå ikke blitt utarbeidet noe balansert og vektet korpus over svensk skriftspråk som grunnlag for ordbøker.

I Norge var det på slutten av 1990-tallet fortsatt stor mangel på systematisk innsamlet materiale for å lage ordbok, og da det ble lyst ut midler til forskningsinfrastruktur på HF-fakultetet ved Universitetet i Oslo, søkte vi og fikk i 1999 aksept for å bygge et bokmålskorpus på 40 millioner løpeord. Ruth Vatvedt Fjeld, Avdeling for leksikografi ved Institutt for lingvistiske og nordiske studier, sto som prosjektleder med Anders Nøklestad ved Tekstlaboratoriet på samme institutt som teknisk fagansvarlig.

Det ble samtidig søkt om midler til å bygge opp et korpus for nynorsk under ledelse av Lars S. Vikør. Nynorskprosjektet valgte å ta inn alle tilgjengelige tekster uten å relatere til vektning eller en viss spredning i materialet, sannsynligvis fordi det var mye mindre tekst å velge blant for nynorsk. Dermed utviklet disse to korpusene seg svært forskjellig, og det var lite reelt samarbeid underveis om design og oppbygning av dem. Det nynorske korpuset skulle først og fremst være grunnlag for redigering av det nasjonale prosjektet *Norsk Ordbok*, der nynorsk skriftspråk og norske dialekter skulle dokumenteres.

[3] OPPBYGGING AV DET LEKSIKOGRAFISKE BOKMÅLSKORPUSET

De eldste tekstene som ble samlet inn for LBK, var fra 1985, siden det var det første året det ble vanlig å publisere tekster elektronisk. Målet var å lage et åpent («open-ended») korpus, der man fylte på nye tekster i vektet mengde for hvert år, inntil man hadde et korpus som var stort nok for det man ville dokumentere. Målet på 40 millioner ord ble nådd i 2008. I løpet av perioden 1999–2008 var elektronisk publisering blitt mye mer vanlig og lettere tilgjengelig. Derfor søkte vi om flere midler for å bygge et korpus på 100 millioner ord, slik at størrelsen skulle bli sammenliknbar med andre internasjonale korpus i tiden, bl.a. British National Corpus.

Med et korpus på 100 millioner løpeord kunne vi dokumentere bedre både enkeltord og fraseologien i moderne norsk bokmål. Forskingen på fraseologi hadde skutt fart på totusentallet. Et språks inventar består ikke bare av enkeltord, mange leksikalske enheter består av flere ord som gjerne opptrer sammen i flerordsenheter eller fraser, kalt MWEs («multiword expressions»). Statistiske analyser av store tekstmengder kan gi bedre oversikt over hvilke flerordsenheter som eksisterer i et språk, eller hvilke som er gangbare i moderne språkføring, og hvilke som er foreldet. Flerordsenheter, både med og uten metaforisk betydning, endrer seg mye raskere i et språk enn enkeltord gjør, og en god ordbok bør kunne vise det. De statistiske analysene som trengs for kartlegging, krever imidlertid et svært omfattende materiale om resultatene skal bli gode nok. Etter en statistisk analyse av MWEs i LBK2008 (Bick 2010) fant vi ut at et materiale på 40 millioner løpeord var for lite for å sortere tilfeldige sammenstillinger av ord fra

flerordsenheter med egen leksikalsk betydning. Ved 100 millioner løpeord ble resultatene mer treffsikre og nyttige i kartlegging av faste fraser og flerordsenheter i forskjellige teksttyper. Fagspråk og allmennspråk har som regel svært ulik fraseologi, dermed er det også viktig å kunne sortere tekstene i subkorpus etter teksttype og sjanger for å kunne beskrive fagrelevant fraseologi bedre.

Vi fikk forlenget prosjektet med dette målet for øye, med støtte fra Universitetet i Oslos infrastrukturprogram. Tekstinnsamlingen ble avsluttet i 2013 med vel 100 millioner løpeord, og LBK inneholder dermed tekster fra en periode på 28 år.

At et korpus er balansert, vil si at korpuset består av en beregnet mengde forskjellige typer tekster etter et definert mål. Vi kom fram til at det skulle være 20 % periodika, 45 % sakprosa, 25 % skjønnlitteratur, 5 % tv-tekst, og 5 % upublisert, og dermed unormert, tekst. Balansen ble bestemt etter en avveining av statistiske data fra Norsk Mediebarometer i 2003 om folks lesevaner¹, samt hva det faktisk var mulig å få tak i av elektroniske tekster uten større kostnader enn prosjektet kunne tåle. Mediebarometeret viste at folks lesevaner det året var 35 % internettlesing, 15 % bøker, 40 % aviser, 5 % tidsskrifter og 5 % tegneserier. Det var lite spesifiserte tall, for eksempel visste man ikke hva slags tekster folk leste på internett, og vi jenket fordelingen noe etter hvordan man hadde designet andre korpus, særlig var British National Corpus et viktig forbilde, samt det korpuset som lå til grunn for utvelgelse av grunnbegrepene i EuroWordNet, i og med at LBK også var ment å være grunnlag for utvikling av et norsk ordnett (Vossen, P. et al 1998, Fjeld, R.V et al 2004). I EuroWordNet la man stor vekt på balanse med hensyn til teksttype og tekstforfatterens kjønn, alder og sosiale tilhørighet. Dette var også data som ble registrert for de fleste tekstprodusentene i LBK. Under innsamlingen prøvde vi bevisst å holde disse variablene så balansert som mulig, med de begrensninger materialtilgangen ga. Det ble en avveining av hva slags materiale som var lett tilgjengelig, og hva vi spesielt trengte å finne for å holde balansen. I den prosessen gjorde de vitenskapelige assistentene en stor innsats for å kartlegge de sosiale variablene. Et unntak gjorde vi med avistekstene, som ofte var svært korte og var skrevet av personer som det var vanskelig å skaffe bakgrunnsdata på. Da ble bare avisnavn og publiseringsdato registrert.

[3.1] *Hvorfor lage et balansert leksikografisk korpus for norsk bokmål?*

Et godt balansert korpus over moderne norsk er et nyttig hjelpemiddel for å kartlegge hvilke ord som er brukt av så mange og i så forskjellige sammenhenger at de bør være oppslagsord i en allmennordbok. Et balansert korpus gir kunnskap om det vanligste ordinventaret i forskjellige typer tekst og er et viktig redskap

[1] <https://www.ssb.no/a/publikasjoner/pdf>

for å kunne foreta en godt fundert lemmaseleksjon. Et godt lemmautvalg inkluderer selvsagt alle rimelig frekvente ord, men også nye ord, såkalte neologismer. At ordbøkene skal ha med nye ord, er lett å forstå, bare man ikke tar med alt nytt som kanskje brukes et fåtall ganger av noen få; det må foreligge en viss konvensjonalisering før et ord lemmatiseres.

Videre er det viktig at de ordene som lemmatiseres, er brukt med en viss frekvens og spredning innen det aktuelle språksamfunnet. Dialektuttrykk tas vanligvis ikke med, eller de merkes som typiske for et område eller et spesielt kommunikasjonsfelt. En god ordbok har derfor et sett av såkalte diasystematiske markeringer som gir opplysninger om bruksforhold eller andre pragmatiske restriksjoner. Men å være treffsikker med disse markeringene forutsetter god innsikt i språkbruken generelt, det er ikke nok med egen intuisjon. Et balansert korpus skal nettopp kunne gi det nødvendige bakgrunns materialet for å gjøre slike nødvendige avveier ved oppføring av bruksmarkeringer.

Ved hjelp av et korpus kjørt mot lemmalisten i en eksisterende ordbok kan man også få en viss oversikt over hvilke ord som ikke eller sjelden brukes lenger, slik at man ekskluderer eller merker dem som anakronismer, avhengig av ordbokstype. I skoleordbøker og innlæringsordbøker er det ofte hensiktsmessig å utelate foreldede ord, da denne typen brukere sjelden leser eller anvender all informasjon i en komplisert ordboksartikkel, og dermed kan komme i skade for å skrive anakronismer eller rett og slett utdøde ord i sine moderne tekster, noe som selvsagt ikke er til hjelp for dem. I større dokumentasjonsordbøker og historiske ordbøker bør ord som ikke lenger er vanlige i samtiden, være med, slik at brukerne kan lese og forstå dem. Tidligere ble slike ord gjerne slettet av plasshensyn. I elektroniske dokumentasjonsordbøker bør de bli stående med bruksmarkering som «sjelden» eller «eldre».

Levende språk er i stadig endring, og særlig endrer ordinventaret seg i takt med den generelle samfunnsendringen. Kartlegging av neologismer og anakronismer er derfor en forutsetning for godt ordboksarbeid (Fjeld og Nygaard 2012), og i det arbeidet er gode og moderne korpus av uvurderlig verdi.

[3.2] *Innsamling av tekstene i Leksikografisk bokmålskorpus*

Tekstene som ble samlet inn til korpuset, ble sortert og klassifisert i følgende teksttyper:

<i>Aviser og ukeblader</i>	AV00
Riksaviser	AV01
Regionaviser	AV02
Lokalaviser	AV03
Ukeblad	AV04
TV	TV00
Teksting av nyhets- og dokumentarprogram	TV02
Teksting av TV-serier og filmer	TV03
<i>Skjønnlitteratur</i>	SK00
Epikk	SK01
Drama	SK02
Lyrikk	SK03
Korttekster	SK04
Religiøse tekster	SK05
<i>Sakprosa</i>	SA00
Lærebøker	SA01
Fagbøker	SA02
Biografier	SA03
Artikkelsamlinger (antologi, festskrift o.l)	SA04
Avhandling	SA05
Juridisk dokument	SA06
Rapport	SA07
Oppslagsverk, leksikonartikler	SA09
Reiseskildring	SA10
Uspesifisert	SA99
Fagtidsskrifter	SA11
<i>Unormert materiale, småtrykk, annet</i>	UN00
Reklame, brosjyre, bruksanvisninger m.m.	UN01
Korrespondanse & forretningsdokumenter	UN02
Diskusjonsgrupper	UN04
Blogger	UN05

Korpuset har registrert tre sosiale variabler hos tekstforfatterne: kjønn, alder og oppvekststed. Idealet var å ha jamn fordeling på alle sosiale variabler, slik at

det blant annet ville være mulig å undersøke hvordan forskjellige brukere anvender den åpne bokmålsnormen, utføre statistiske analyser og få sammenliknbare resultater.

Resultater fra korpus bør alltid vurderes etter to vesentlige dimensjoner, frekvens og spredning. Svært mange undersøkelser i store tekstmengder, som hos Google eller liknende, gir gode anslag for frekvenser, men det er ofte vanskelig å finne ut noe etterprøvbart om spredning. I beste fall kan man finne spredning over tid, da publiseringsår ofte registreres, men spredning på fagfelt, sosiale variabler hos forfattere som kjønn og alder eller andre avgjørende variabler er det ikke mulig å sortere ut. Vil man ha sammenliknbare kvantitative resultater etter sosiale variabler fra LBK, kan man i søkegrensesnittet lett skille ut subkorpus med lik størrelse. Et balansert og godt merket korpus kan dessverre heller ikke bli helt perfekt eller treffsikkert, men det kan i hvert fall gi bedre grunnlag for videre undersøkelser enn et søk på for eksempel Google kan.

Det var mange hindringer som skulle overvinnes for å få tilgang til nok elektroniske tekster, særlig ga lov om åndsverk store problemer i de første årene prosjektet pågikk. De ble løst ved at vi avtalte med forlagene å slette ca. 20 % av sidene i hvert verk, tilfeldig spredt utover i teksten, slik at den ble umulig å rekonstruere for piratutgaver, som var det man fryktet mest ved oppstarten av prosjektet. Det lettet søknadsprosessen overfor tekstprodusentene i stor grad. Redselen for misbruk og rekonstruksjon avtok imidlertid etter hvert som digitalisering og nettpublisering ble mer vanlig utover 2000-tallet, og fra ca. 2010 var det få som ba om slik tekstreduksjon. I tillegg ble det undertegnet forpliktende avtaler med alle tekstgivere om at tekstene ikke skal brukes til kommersielle formål, kun til språkforskning, og at tekstene bare offentliggjøres gjennom et avtalt grensesnitt som er tilgjengelig for registrerte brukere via Tekstlaboratoriet ved Universitetet i Oslo.

I begynnelsen var det også vanskelig å få fatt i nok elektronisk tilgjengelige tekster uten vederlag. Men gjennom gode kontakter, først og fremst forlagsredaktør Øystein Eek i Kunnskapsforlaget, fikk vi avtaler med forlagene Gyldendal, Kunnskapsforlaget og Aschehoug. Forlagsredaktør Inger-Ma Gabrielsen har også bidratt jevnlig ved å gi UiO kopi av det meste som ble utgitt på Cappelen forlag. Seinere fikk vi også avtale med Verbum forlag, samt en lang rekke mindre tekstprodusenter.

Tekstene har ellers blitt samlet inn på svært forskjellige måter, bl.a. sendte prosjektlederen i 2001 ut et bøneskrift til alle universitetsansatte i Norge, med overskriften «Bli ordgiver du også». Det var svært mange som ga positiv respons på det og bidro med sin egen faglige produksjon, og slik ble mye av fagprosaden i korpuset mettet, og det ga god spredning på fagfelt. Vi fikk i den forbindelse

også både publiserte og upubliserte vitenskapelige tekster. Sakprosatekstene er fordelt på humaniora, samfunnsfag, realfag, juss, helsefag samt sport og fritid. Dermed er det nå mulig å kartlegge faguttrykk som er gått inn i allmennspråket, gjerne med utvidet eller endret betydning, og som da bør defineres ut fra både faglig og allmennspråklig bruk.

Populærvitenskapelige tekster fant vi for det meste på Internett og i aviser. Også venner og bekjente av mange som arbeidet på korpuset, har bidratt med tekster. Vi har ellers fått tekster fra en lang rekke forfattere som har gitt oss sine råmanuskripter slik de var før de ble vasket av forlagene, slik at de kunne sammenliknes med de endelige, publiserte tekstene. Slike dublettekster er interessante for leksikografisk forskning og for normeringsarbeid generelt, og materialet har blitt analysert i interessante studentoppgaver som blant annet ble publisert på prosjektets lanseringsseminar i 2008 (Ims 2008). Dette materialet kan lett sorteres ut som et subkorpus som det er viktig å hegne spesielt om, da råtekstene i prinsippet er upubliserte og uferdige.

I tillegg til forlagstekster og privat materiale har vi høstet inn offentlige tekster på Internett etter hvert som det ble en vanlig publiseringskilde, så mot slutten av arbeidsperioden ble det viktigere å velge de rette tekstene, mens det i starten var et hovedproblem å få tak i tekster i det hele tatt. Hele tiden måtte vi ha den rette balansen for øye, og sørge for at balansen skulle være der for hvert enkelt årstall.

Det mest krevende var å få inn teksten talemål fra tv og radio, noe som ble løst ved å legge inn NRKs teksting for hørselshemmede over en viss periode. På den måten fikk vi i det minste registrert det nesten-muntlige ordforrådet, som er viktig i et leksikografisk underlagsmateriale. Tekstekontoret ved NRK fortjener en ekstra takk for fleksibilitet ved å gi oss tilgang til dette.

[3.3] *Resultatet*

Korpuset består av 27 082 dokumenter som til sammen utgjør 99 959 468 token, dvs. ordformer og tegnsettingstegn. Alle metadata er registrert manuelt. Målet med korpuset var som nevnt en balansert sammensetning av teksttyper. Videre ønsket vi å ha antall løpeord balansert for hvert enkelt år, samt god spredning på forfatternes sosiale variabler. Det har ikke vært lett å holde den balansen, og dersom man ønsker å gjøre direkte statistiske analyser ut fra visse variabler, kan det bli nødvendig å velge ut likevektige subkorpus. For eksempel var det vanskelig å finne like mange tekster skrevet av kvinner som av menn. Siden det totale korpuset er veldig nær 100 millioner token, er det enkelt å finne prosentandel for de enkelte variablene. For eksempel er det knapt 53 prosent tekst produsert av menn og tett oppunder 30 prosent av kvinner. De resterende tekstene har vi

ikke opplysninger om forfatterens kjønn på. Det gjelder først og fremst avistekster.

Om vi ellers ser på balansen i det ferdige korpuset, får vi også et noe annet bilde enn det som var målet. I forhold til de oppsatte målene har korpuset knapt 5 % mer sakprosa enn planlagt og 10 % mer skjønnlitteratur, og 15 % mindre fra periodika. Samtidig med arbeidet med LBK ble det utviklet et omfattende avis-korpus ved Universitetet i Bergen, Norsk Aviskorpus. Siden LBKs annoterings-system var krevende for nettopp korte avistekster, valgte vi derfor å øke mengden skjønnlitteratur, og redusere avistekster. Man kan lett kontrollere ordforrådet i avistekster fra LBK mot Norsk Aviskorpus, og vi anser ikke denne balanseforskyvningen som betydningfull for verdien av LBK. Norsk aviskorpus er dessuten et særdeles godt hjelpemiddel i den redaksjonelle vurderingen som alltid foretas før et ord ordboksføres.

LBK er et ikke-kommersielt produkt som kan brukes av alle som logger seg inn og godkjenner lisensavtalen. Republisering er ikke tillatt, ellers kan korpuset gjerne brukes i språkforskning som gir salgbare produkter. Avtalen vi har med forlagene og andre tekstgivere, er at Universitetet i Oslo har rett til å bruke tekstene til forskningsformål mot at tekstgiverne selv også får tilgang til å bruke korpuset. Videre kan Universitetet i Oslo gi andre forskere utvidet tilgang etter spesiell avtale.

En viktig tanke bak LBK var nemlig at det ikke bare skulle fungere som materiale for ordboksredigering, men også kunne brukes i språkforskning generelt. For eksempel skulle korpuset kunne dokumentere bruken av moderne norsk morfologi, og spesielt vise hvilke brukere som foretrakk de forskjellige formene. Det er særlig interessant i språk som norsk, der det har foregått en politisk styrt normering over mange år, samtidig som normen er svært vid og gir mange valgmuligheter. Korpuset kan også gi svar på spørsmål om i hvilken grad eksisterende normering har virket, eller hvilke ikke-tillatte former som er så mye i bruk at de burde tas inn i normen. Dette er bl.a. demonstrert tydelig i masteroppgaven til Kjersti Wictorsen Kola (Kola 2014).

Korpuset har foreløpig ikke blitt utnyttet til å lage en korpusbasert ordbok, slik DDO ble laget over et eget korpus, men det har vært benyttet ved revisjon av både Bokmålsordboka og flere andre bokmålsordbøker. I Det Norske Akademis ordbok (NAOB, 2017) ble korpuset brukt flittig, både som leksikografisk verktøy for å undersøke morfologi, kollokasjoner og annet, og for å hente ut sitater. Pr. 2019 er det 30277 ordboksartikler i NAOB som har sitat fra LBK. Ellers er det skrevet flere fagartikler, masteroppgaver og PhD-avhandlinger med grunnlagsmateriale fra LBK.

[4] SØKEPROGRAMMET GLOSSA

Leksikografisk bokmålskorpus er gjort tilgjengelig i søkesystemet Glossa (Nøklestad et al. 2017), som er utviklet av Tekstlaboratoriet ved Universitetet i Oslo. Glossa er web-basert, noe som innebærer at man ikke trenger å installere programvare for å bruke det; alt man trenger er en nettleser og en datamaskin som er koblet til internett. Systemet støtter innlogging med Feide², noe som betyr at ansatte og studenter på norske universiteter og høyskoler (og på mange andre offentlige institusjoner) vil kunne logge inn med det brukernavnet og passordet de bruker på institusjonen sin. Det samme gjelder personer ved utenlandske akademiske institusjoner som støtter innlogging med eduGAIN³, og personer med CLARIN-konto⁴. Om man ikke har tilgang til noen av disse innloggingsmetodene, kan man også få tilgang ved å kontakte Tekstlaboratoriet⁵.

[4.1] Søkegrensesnitt

Glossa gjør det mulig å søke etter bestemte ord, eller deler av ord, fraser, lemma (grunnformer eller oppslagsformer), ordklasser og grammatiske trekk. Man kan søke i et utvalg av tekstene, for eksempel bare i aviser eller i tekster av en bestemt forfatter. Resultatene blir presentert som en konkordans, dvs. en liste med tekstutdrag som viser søkeordet med litt kontekst foran og etter. Konkordansen kan lastes ned i Excel-format eller som tab- eller kommaseparert tekst, noe som gjør det enkelt å jobbe videre med resultatene i annen programvare, som Excel eller statistikkprogrammer. Man kan også få resultatene i form av frekvenslister.

En ulempe med mange tidligere korpussøkegrensesnitt er at de krever teknisk kunnskap, både om såkalte regulære uttrykk og om kodene som representerer metadatakategorier og grammatisk informasjon i korpuset. Glossa er derimot utformet med tanke på at det skal være enkelt å bruke for personer uten slike spesielle tekniske kunnskaper.

Hovedsøkesiden til LBK er vist i Figur 1, der fargede rammer er satt inn for å utheve viktige deler av grensesnittet. Til venstre (grønn ramme) er alle metadatakategoriene det går an å søke i. Disse kategoriene representerer ulike typer informasjon om tekstene i korpuset: Tekst-ID, Tittel, Publikasjon, Kategori, Underkategori, Utgiver, År, Sted, Oversatt, Emne, Emne (detaljer), Navn på forfatter/oversetter, Kjønn og Fødselsår. Glossa sørger for at man ikke kan velge inkompatible verdier fra de ulike kategoriene; har man først valgt Dag Solstad under Navn på forfatter/oversetter, så vil Tittel-lista blir redusert til bare å inneholde boktitler av Solstad. Over metadatakategoriene kan man se hvor mange

[2] <https://www.feide.no/>[3] <https://edugain.org/>[4] <https://www.clarin.eu/>[5] tekstlab-post@iln.uio.no

tekster man har valgt, og hvor mange token (ordformer og tegnsettingstegn) de inneholder til sammen. Knappen *Show texts* (også grønn ramme) gir en oversikt over alle tekstene eller det utvalget tekster som har blitt valgt (se Figur 2). Øverst i Figur 1 finner vi to knapper (rød ramme). Med *Hide filters* kan man skjule meta-datakategoriene til venstre, mens *Reset form* gir en blank søkeside.

The screenshot shows the main search interface for the Lexicographical Norwegian Corpus (LBK). At the top left, it says 'Glossa'. On the right, there are logos for 'CLARINO' and 'tekstlab. Logged'. Below the logos, there are two buttons: 'Hide filters' and 'Reset form', both highlighted with a red border. The main heading is 'Leksikografisk bokmålskorpus'. Below this, there are three search options: 'Simple', 'Extended', and 'CQP query', with 'Simple' highlighted in yellow. To the right of these options is a green 'Search' button. Below the search options is a search input field. Underneath the input field is a button labeled 'Or...' and a green button labeled 'Show texts', which is highlighted with a green border. At the bottom, there is a list of links: 'Les om korpuset', 'Rapporter om feil i korpuset', 'Bruk det gamle søkegrensesnittet', and 'Hvordan referere til korpuset'. On the left side, there is a sidebar with a list of filters, including 'All 27081 texts (99959468 tokens) selected', 'Tekst-ID', 'Tittel', 'Publikasjon', 'Kategori', 'Underkat.', 'Utgiver', 'År', 'Sted', 'Oversatt', 'Emne', 'Emne (detaljer)', 'Navn på forfatter/oversetter', 'Kjønn', and 'Fødselsår'. The entire sidebar is enclosed in a green border.

FIGUR 1: Hovedsøkesiden til LBK.

Glossa tilbyr tre forskjellige søkegrensesnitt (gul ramme i Figur 1), der valgene fra venstre mot høyre gir økende muligheter for å formulere avanserte søk, men også en viss økning i vanskelighetsgrad. Grunnleggende søk etter ordformer eller fraser kan foretas ved hjelp av en enkel søkeboks som minner om det man finner i Google eller andre web-søkemotorer, og som de fleste brukere derfor vil være kjent med. Det er dette grensesnittet som er vist i Figur 1.

Corpus texts

x



2 of 27081 texts (1137 of 99959468 tokens) selected

Tekst-ID	Tittel	Publikasjon	Kategori	Underkat.	Utgiver	År	Sted
SA06AAxx12	Arbeidsavtale		SA	SA06	Troms Fylkeskommune	0000	Tromsø
SA06AAxx32	Universitetet i Tromsø har inngått følgende arbeidsavtale med arbeidstaker - midlertidig ansettelse		SA	SA06	Universitetet i Tromsø	0000	Tromsø

FIGUR 2: Oversikt over tekstene som er valgt.

Vil man søke etter lemma, deler av ord eller grammatiske trekk, kan man bytte til et grensesnitt som presenterer slike valgmuligheter i form av avkrysningsbokser, knapper og nedtrekksmenyer. Det gjør man ved å klikke på *Extended* over søkeboksen, som vist Figur 3. Man får da opp en søkeboks med en rekke avkrysningsknapper under, og disse gjør det mulig å spesifisere at søkeordet skal være lemma, begynnelsen eller slutten på et ord, eller i begynnelsen eller slutten av en setning. I dette grensesnittet er det også mulig å spesifisere at man vil ha et antall tilfeldig utvalgte resultater i stedet for alle resultater. Dette kan være nyttig hvis man vil ha ut et begrenset antall eksempler, men likevel vise variasjonen som fins i kildetekstene (siden man ellers bare ville få eksempler fra de første tekstene i korpuset).

Vil man oppgi grammatiske kriterier, kan man gjøre det ved å klikke på en av knappene til venstre for søkeordet, som vist i Figur 4. Knappen rett til venstre for ordet åpner en meny der man raskt kan velge ordklasse. Vil man i tillegg oppgi bøyingstrekk, kan man i stedet klikke på knappen lengst til venstre og få opp et vindu med ordklasser og tilhørende morfosyntaktiske trekk, som avbildet i Figur 5. Merk at det er fullt mulig å spesifisere både ordform eller lemma og grammatiske kriterier for ett og samme søkeord.

Glossa CLARINO  tekstlab. 

All 27081 texts
(99959468 tokens)
selected

Hide filters Reset form

Leksikografisk bokmålskorpus

Tekst-ID
Tittel
Publikasjon
Kategori
Underkat.
Utgiver
År
Sted
Oversatt
Emne
Emne (detaljer)
Navn på
forfatter/oversetter
Kjønn
Fødselsår

Simple **Extended** CQP query Search

+

Lemma Start End
 Sentence initial Sentence final

Or... Show texts random results (with seed:)

- Les om korpuset
- Rapporter om feil i korpuset
- Bruk det gamle søkegrensesnittet
- Hvordan referere til korpuset

FIGUR 3: Utvidet søk.

Simple **Extended** CQP query Search

+

Lemma Start End
 Sentence initial Sentence final

Noun plural x

FIGUR 4: Knappene for ordklassesøk og søk etter andre morfologiske trekk.

Nederst i vinduet i Figur 5 kan man også utelukke bestemte ordformer eller lemma fra søket. Denne muligheten er nyttig hvis man for eksempel vil søke etter verb, men utelukke de verbformene som har «være» eller «ha» som lemma.

The screenshot shows a search interface with two main sections. The top section, titled "Parts-of-speech", contains buttons for: adjective, adverb, determiner, infinitive marker, interjection, conjunction, preposition, pronoun, subjunction, noun, unknown, and verb (which is highlighted in blue). The bottom section, titled "Morphosyntactic features for verb", contains buttons for: Voice: passive; Mood: imperative; Tense: present tense, past tense; Type: infinitive, past participle, no inflection. Below these sections is an "Exclude lemma" dropdown menu, an "OK" button, and two red buttons labeled "lemma:være x" and "lemma:ha x". At the bottom are "Clear", "Search", and "Close" buttons. A note at the bottom left says "Click to select; shift-click to exclude".

FIGUR 5: Vindu med ordklasser og tilhørende bøyningstrekk, samt mulighet for å spesifisere eller ekskludere lemma eller ordform.

Hvis man vil søke etter en frase, kan man legge til flere søkeord ved å klikke på den blå pluss-knappen lengst til høyre for søkeuttrykket. Vil man fjerne et av ordene igjen, kan man klikke på minus-knappen til høyre for det aktuelle søkeordet. Man kan også oppgi minimum og/eller maksimum antall uspesifiserte ord som skal kunne forekomme mellom to oppgitte søkeord; dette gjør det mulig for eksempel å søke etter ordet «en» etterfulgt av et substantiv, men med minst ett og maks to uspesifiserte ord (for eksempel adjektiver med eventuelle adverb foran) mellom. Disse funksjonene er uthevet i Figur 6.

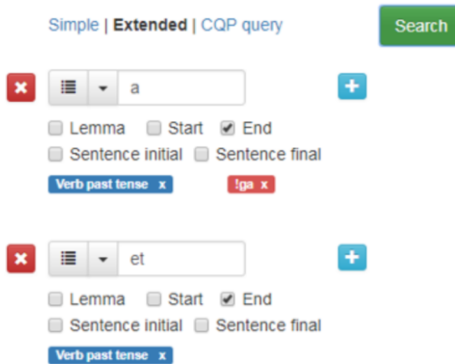
The screenshot shows a search interface with a search bar containing the text "en". To the right of the search bar are buttons for "-" and "+". Below the search bar are checkboxes for "Lemma", "Start", "End", "Sentence initial", "Sentence final", and "Noun x". A red box highlights the "Extended" tab and the search bar. Another red box highlights the "-" and "+" buttons and the "1 min" and "2 max" input fields. A third red box highlights the "Noun x" button.

FIGUR 6: Søk på flere ord.

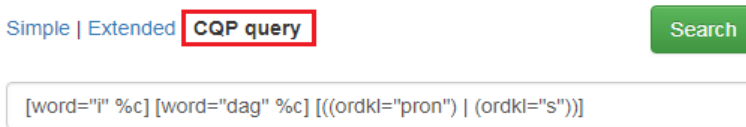
Det er også mulig å søke etter flere alternative varianter av et ord eller en frase. Hvis man klikker på knappen merket *Or...* under søkeboksen, får man opp en ny rad der man kan oppgi et alternativt søkeuttrykk. Figur 7 viser et eksempel på dette, der man har valgt å søke etter verb i preteritum som slutter på *-a* eller *-et*. Ordformen *ga* er ekskludert fra søket ved hjelp av mekanismen som ble vist i Figur 5.

Til slutt kan man velge å bruke et grensesnitt som gir tilgang til alle former for avanserte søk som er støttet av den underliggende søkemotoren (*The IMS Open Corpus Workbench*, Evert og Hardie 2011), men som til gjengjeld krever kunn-

skap om regulære uttrykk og grammatiske koder. Figur 8 viser hvordan søkeboksen ser ut med et regulært søkeuttrykk. I stedet for å utforme et slikt søkeuttrykk fra bunnen av kan man spesifisere søket så langt det er mulig under *Extended* og deretter bytte til *CQP query*-grensesnittet (som da vil vise uttrykket for søket man har spesifisert så langt) og justere det videre der.



FIGUR 7: Samtidig søk etter alternativer.



FIGUR 8: Avansert søk med regulære uttrykk. Søkeuttrykket inneholder ordformen *i* direkte etterfulgt av ordformen *dag*, igjen direkte etterfulgt av et ord med ordklasse «pron» (pronomen) eller «s» (substantiv). «%c» spesifiserer at forskjellen mellom store og små bokstaver skal ignoreres.

[4.2] Søkeresultater

Søkeresultatene for LBK blir presentert i form av konkordanser og frekvenslister, med konkordanser som standardvisning (se Figur 9). Hvert søkeresultat blir vist som en rad i tabellen, med kolonner for tekst-ID, kontekst før søkeordet, selve søkeordet, og kontekst etter søkeordet. Kontekstlengden kan justeres ved å skrive inn et nytt tall i boksen over tabellen og trykke *Enter*. Ved hjelp av de uthevede knappene over tabellen kan resultatene sorteres eller lastes ned i Excel-format eller tab- eller kommaseparert format. Hvis man holder musa over søkeordet eller et ord i konteksten, får man opp informasjon om

lemma og grammatiske tagger, som vist i Figur 10. Ved å klikke på tekst-IDen i første kolonne får man listet opp metadataene som er registrert for teksten som søkeresultatet er hentet fra (Figur 11).

Concordance		Statistics		Found 76569 matches (1532 pages)	
Sort by position ▼	Download	Context: 15 words		<	>
AV01AF930007.33	første kunstneren som foreviger slike ekspedisjoner . Det ble malt akvareller fra Scottekspedisjonen . Men	kanskje	er han den første med oljebilder ? Resultatet skal skal stilles ut i Oslo til		
AV01AF930025.39	, denne gangen . Straffen er fullbyrdet . De slo ham aldri . Det var	kanskje	hans ulykke . Hadde de slått , ville han ha blitt trodd av dem som		
AV01AF930033.43	etter dyr som er forsvunnet . - Vi kan lete dagevis etter ett dyr .	Kanskje	sitter det fast og pines , sier reingjeterne . Før snøscooteren kom bodde reindriftsfamilien ved		
AV01AF930048.28	fant mening i det . Selv om det muligens ikke var fornuftig . Det er	kanskje	ikke fornuftig å dra på bildekk i 1994 heller . Men man kan jo bli		
AV01AF930048.31	, ennå finnes : Det kan finne mening i det ubegripelig meningsløse , ja ,	kanskje	nettopp der . Samme hva mediene måtte mene på våre andres , mindre uvørnes ,		

FIGUR 9: Konkordansvisning av søkeresultater.

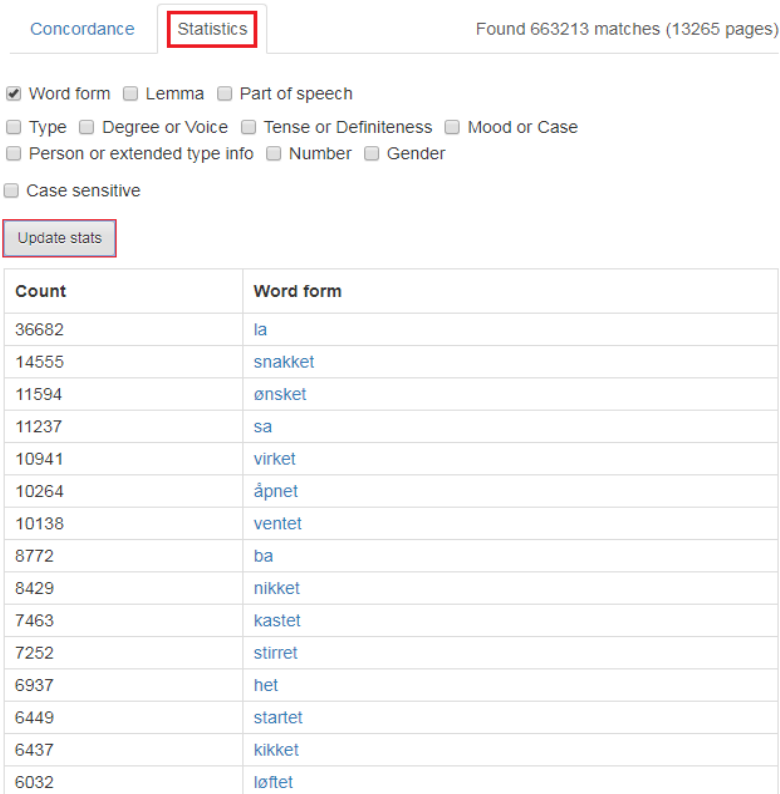
"annen" det fl dem
: litt mykere med andre ord

FIGUR 10: Visning av lemma og grammatisk informasjon for et ord i søkeresultatet.

G		CLARINO		tekslab	
Tekst-ID	AV01Da9706				
Tittel	Dagbladet. En europeisk pådriver				
Publikasjon	Dagbladet				
Antall ord	1565				
Kategori	AV				
Underkat.	AV01				
Utgever	Dagbladet				
År	1997				
Sted	Oslo				
Korpusdato	2000-09				
Oversatt	original				
Kjønn	M				
Navn på forfatter/oversetter	Borchgrevink, Aage Storm				
Fødselsår	1969				
Emne (detaljer)	SAM05				
Emne	SAM%: Samfunnsfag				
Oversatt	AV01U89706.57	ptre uavhengig av OSSEs politiske organer . Uavhengighet er det som karakteriserer OSSEs	kanskje	mest vellykkete institusjo minoriteter . Akkurat som	
Emne	AV01Da9804.60	, der man våger å skrelle av gammelt , umoderne , tungt ideologisk tankesett	Kanskje	man våger å gi slipp på rr	

FIGUR 11: Visning av metadata for et søkeresultat.

Under den andre fanen med søkeresultater, *Statistics*, kan man få vist frekvenser over søkeresultatet; se Figur 12. Man kan velge å se frekvenser for ordformer, lemma og/eller grammatiske trekk. I figuren har vi valgt å vise ordformer. Hvis man klikker på en ordform, får man se en konkordans med bare de søkeresultatene som inneholder den bestemte ordformen.



FIGUR 12: Frekvenser for ordformene i et søkeresultat.

[5] OPPSUMMERING

LBK er et av flere norske tekstkorpus som er utviklet på 2000-tallet. Det som skiller LBK fra andre korpus, er at det er relativt godt balansert og har svært mange metadata. Dermed kan LBK utnyttes til flere formål enn å kartlegge leksikalske enheter. De enkelte ordboksoppslag (lemmaer) kan utstyres med rikere informasjon enn det som er vanlig i norske ordbøker, f.eks. morfologisk variasjon, leksikalske forskjeller hos forskjellige grupper språkbrukere, kjønns spesifikt ordvalg, leksiko-syntaktiske endringer hos yngre brukere, for å nevne noe. Korpuset

har et vell av metadata, både relatert til språkbrukerne og til de forskjellige teksttypene og –sjangrene. Ved grundig analyse av LBK kan man dessuten framskaffe sikre data i den spesielle norske språkpolitiske og språkideologiske debatten. Søkesystemet Glossa gjør at det er enkelt å anvende for statistisk analyse og er gjort greit tilgjengelig for språkforskere.

Korpuset har alt i alt kostet rundt 10 årsverk, alt finansiert av Universitetet i Oslo, og vil være en viktig ressurs for moderne språkforskning og leksikografisk dokumentasjon av moderne norsk bokmål for perioden 1985-2013.

TAKK

Det har vært mange medarbeidere på prosjektet for tekstsamling, tekstrensing og merking. Lars Nygaard, Anne Engø, Preben Wik, Rune Lain Knudsen og Arash Saidi har etter tur vært tilsatt på timebasis som dataingeniører, og har bidratt mye til den teknologiske utviklingen av korpuset, hele tiden under ledelse av senioringeniør Anders Nøklestad. Det har vært mange vitenskapelige assistenter som har arbeidet med innsamling av tekster, tekstrensing og innlegging i korpuset. Ålov Runde var tilsatt som vit.ass. i oppbyggingsfasen og gjorde et stort arbeid med praktisk tilrettelegging av selve databasen. Videre i perioden har mange assistenter bidratt, særlig bør nevnes vit.ass.-ene Carina Nilstun, Julie Torjusen og Kjersti Wictorsen Kola, som alle har vært studenter på bachelorkursen i leksikografi og har bidratt vesentlig i arbeidet med LBK, både med gode ideer, systematisering og organisering av korpusinnholdet, og som har fungert som hjørnesteiner i et særdeles godt og kreativt arbeidsmiljø over mange år.

REFERANSER

Arakin, Vladimir D. (red.). 1963. *Norsk-russisk ordbok*. Moskva: Statsforlaget for ordbøker på fremmede språk og nasjonalitetsspråk.

Bergenholtz, Henning. 1996. Korpusbaseret leksikografi. I *LexicoNordica* 3, 5-18.

Bick, Eckhard. 2010. DeepDict - et korpusbaseret relationelt leksikon. I Ruth Vatvedt Fjeld & Henrik Lorentzen (red.), *Lexico Nordica 17 - 2010, Leksikografi og språkteknologi i Norden*, 17-34 . *LexicoNordica*. ISSN 0805-2735.

British National Corpus. <http://www.natcorp.ox.ac.uk/corpus/creating.xml>

Den danske ordbog. <https://ordnet.dk/ddo>

Den Danske ordbogs elektroniske korpus. <https://ordnet.dk/ddo/fakta-om-ddo/metode-og-kilder/en-korpusbaseret-ordbog>

Det Norske Akademis ordbok (NAOB). <https://www.naob.no/>

Evert, Stefan & Andrew Hardie. 2011. Twenty-first century Corpus Workbench: Updating a query architecture for the new millennium. I *Proceedings of the Corpus Linguistics 2011 conference*, University of Birmingham, UK.

Fjeld, Ruth E Vatvedt, Wik, Preben & Lars Nygaard. 2004. Managing Complex and Multilingual Lexical Data with the Simple Editor. I *Euralex international congress; 2004-07-06*.

Fjeld, Ruth E. Vatvedt & Lars S. Vikør. 2008. *Ord og ordbøker*. Fagbokforlaget.

Fjeld, Ruth E. Vatvedt & Lars Nygaard. 2012. Lexical neography in modern Norwegian. I *Exploring newspaper language : using the web to create and investigate a large corpus of modern Norwegian*. John Benjamins Publishing Company, 221-240. ISBN 978-90-272-0354-0.

Ims, Ingunn. 2008. *Kan dublett-tekstene fortelle noe om et forlags normeringsfilter?* Presentasjon på lanseringsseminar, Institutt for lingvistiske og nordiske studier ved Universitetet i Oslo.

Kilgarriff, Adam. et al. 2004. The Sketch Engine. I G. Williams og S. Veissier (red.), *Proceedings of the Eleventh EURALEX International congress, Euralex 2004*, Université de Bretagne-Sud, 105-116.

Knudsen, Knud. 1881. *Unorsk og norsk eller fremmedordenes avløsning*. Cammermeyer forlag.

Knudsen, Rune Lain & Ruth Vatvedt Fjeld. 2013. LBK2013: A balanced; annotated national corpus for Norwegian Bokmål. I *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013; May 22-24; Oslo; Norway*. NEALT Proceedings Series 19.

Knudsen, Trygve & Alf Sommerfelt (red.). 1937-1957. *Norsk Riksmålsordbok*. Utgitt av Riksmålsvernet.

Kola, Kjersti Wictorsen. 2014. *Bokmålsbruk – hvorledes/hvordan/åssen og hvorfor?: Om bruken av morfologiske og ortografiske varianter i bokmålsnormalen*. Masteroppgave, Universitetet i Oslo.

Krishnamurthy, Ramesh. 1997. Corpus-driven lexicography. I *International Journal of Lexicography* vol. 21, nr. 3, 231-242.

Landau, Sidney I. 1984. *The Art and Craft of Lexicography*. Schribner.

- Malmgren, Sven Göran. 1994. *Svensk lexikologi - Ord, ordbildning, ordböcker och orddatabaser*. Studentlitteratur AB.
- Neset, Tore & Trond Trosterud. 2005. Ny norsk-russisk ordbok: Ei leksikografisk storhending. I *LexicoNordica* 12, 273-284.
- Nordisk leksikografisk ordbok*. 1997. Universitetsforlaget.
- Noreng, Harald. 1993. *Konkordans over Henrik Ibsens dramaer og dikt*. Ibsensamlingen. Oslo.
- Norsk aviskorpus*. <http://uni.no/nb/uni-computing/clu/norsk-aviskorpus/>
- Nøklestad, Anders, Kristin Hagen, Janne Bondi Johannessen, Michal Kosek, & Joel Priestley. 2017. A modernised version of the Glossa corpus search system. I Jörg Tiedemann (red.), *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, 251-254.
- Sinclair, John. 1987. *Looking Up: An Account of the COBUILD Project in Lexical Computing and the Development of the Collins COBUILD English Language Dictionary*. Collins ELT.
- Sinclair, John. 1991. *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.
- Svensén, Bo. 2005. *Handbok i lexikografi. Principer och metoder i ordboksarbeidet*. Es-selte 1987 og 2005.
- Vossen, Piek, Laura Bloksma, Paul Boersma, Felisa Verdejo, Julio Gonzalo, Horacio Rodriguez, German Rigau, Nicoletta Calzolari, Carol Peters, Eugenio Picchi, Simonetta Montemagni & Wim Peters. 1998. *EuroWordNet Tools and Resources Report*. EuroWordNet (LE-4003) Deliverable D021D025, University of Amsterdam.

KONTAKT

Ruth Vatvedt Fjeld
Institutt for lingvistiske og nordiske studier, Universitetet i Oslo
r.e.v.fjeld@iln.uio.no

Anders Nøklestad
Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo
anders.noklestad@iln.uio.no

Kristin Hagen
Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universitetet i Oslo
kristin.hagen@iln.uio.no