

LEXICONORDICA

LEXICONORDICA

32 · 2025

NORDISKE ORDBØKER
– OPPDATERING, UTVIKLING OG
TILGJENGELEGGJERING

NORDISK FORENING FOR LEKSIKOGRAFI

LexicoNordica 32 · 2025

Nordiske ordbøker – oppdatering, utvikling og tilgjengeleggjering

Hovudredaktørar

Henrik Hovmark

Terje Svardal

Redaksjonskomité

Helga Hilmisdóttir

Louise Holmer

Caroline Sandström

Bente Selback

Liisa Theilgaard

© 2025 LexicoNordica og forfattarane

Omslag og sats: Laurids Kristian Fahl

Trykt hos: Tarm Bogtryk a-s, Danmark

Takk til Voksenåsen kulturcentrum for å ta imot

LexicoNordica-symposium 32



VOKSENÅSEN

LexicoNordica blir trykt med økonomisk stønad frå

Nordplus Nordiske Sprog



Nordplus

ISSN 0805-2735

ISSN 1891-2206 (online)

Innhald

Terje Svardal & Henrik Hovmark

Nordiske ordbøker – oppdatering, utvikling og
tilgjengeleggjering 9

Tematiske bidrag

Magnus Ahltopf & Laila Höglund

Framtidssäkrade samhällsord i Lexin 17

Trausti Dagsson & Ellert Þór Jóhannsson

Revitalisering af Íslensk orðsifjabók, islandsk etymologisk
ordbog 37

Hanne Lauvstad

Litteraturen i den diakrone dokumentasjonsordboken
– en beskrivelse med utgangspunkt i *Det Norske Akademis*
ordbok (NAOB) 59

Pär Nilsson

Vetenskaplig akribi under tidspress – och andra utmaningar
att hantera vid revideringen av en stor historisk ordbok 81

Sanni Nimb & Agnes Aggergaard Mikkelsen

Den Danske Begrebsordbog – opdatering, udvikling og
tilgængeliggørelse 107

Bente Selback, Klara Sjø, Knut E. Karlsen & Sturla Berg-Olsen
Revisjonen av *Bokmålsordboka* og *Nynorskordboka*:
oppdatering og tilpasning av ordbøkene til en ny tid..... 131

Ásta Svavarsdóttir & Helga Hilmisdóttir
Gamle historiske sprogarkiver og deres værdi i nutiden..... 155

Lars Trap-Jensen
Fremtidens ordbøger: udfordringer og tilpasninger..... 177

Emma Josefin Ölander Aadland
Kjønnsrepresentasjon og -fremstilling i *Bokmålsordbokas*
brukseksempel 197

Ikkje-tematiske bidrag

Andreas Brimberg & Linnea Hanell
Har de kommit för att stanna? Om nyordslistornas ord i
SO och SAOL 223

Poul Hansen
Den universelle First-Letter Law og dens anvendelse til
statistisk undersøgelse af skandinaviske ordbøger..... 243

Bokmeldingar

Holger Juul
Anmeldelse af *Retskrivningsordbogen*, 5. udgave 265

Eva Thelin
En rikhaltig ordbok om dialekterna i Thy..... 281

Meldingar

Einar Freyr Sigurðsson

Nyt fra bestyrelsen for Nordisk Forening for Leksikografi..... 297

Redaktionsanvisninger 301

Nordiske ordbøker – oppdatering, utvikling og tilgjengeleggjering

Terje Svardal & Henrik Hovmark

Vi har gleden av å presentere det 32. nummeret av *LexicoNordica*, tidsskriftet som blir gitt ut av Nordisk forening for leksikografi (NFL). Som vanleg inneheld publikasjonen ein tematisk del, som i år omfattar ni bidrag. Desse bygger på føredrag haldne på *LexicoNordica*-symposiet 2025. Dette blei arrangert i Oslo på Voksenåsen, Noregs nasjonalgåve til Sverige. I tillegg til dei tematiske bidraga inneheld dette nummeret to ikkje-tematiske bidrag, bokmeldingar og avslutningsvis meldingar frå styret i Nordisk forening for leksikografi og dessutan redaksjonsreglane for tidsskriftet.

Årets tema er oppdatering, utvikling og tilgjengeleggjering av nordiske ordbøker. Dei er viktige for å sikre nordisk språkforståing og dokumentasjonen av felles nordisk kulturarv. For at ordbøkene skal kunne fylle denne funksjonen, er det ein føredsetnad at dei er kvalitetssikra og innholdsmessig oppdaterte. I tillegg må dei vere tilgjengelege på måtar som møter brukarane sine behov og brukarane si verkelegheit. Nordiske ordboksressursar har i vår tid ei rekke utfordringar i samanheng med dette: Samtidsordbøker møter stor konkurranse frå ikkje-kvalitetssikra webordbøker og frå applikasjonar som bygger på og leverer maskingenerert tekst og innhald. Samtstundes forventar publikum at ordbøkene ikkje berre gir korrekt informasjon til for eksempel skulevesenet, offentleg forvaltning og andre sektorar og brukargrupper, men at dei òg blir løpande oppdaterte med omsyn til representasjon og framstilling av kulturar og minoritetar i samfunnet. Historiske ordbøker med omfattande opplysningar om kulturarv har særleg utfordringar med tanke på omstillinga til moderne teknologiar og tilgjengeleggjering for eit breiare publikum.

Bidraga i årest *LexicoNordica*-utgåve belyser temaet på ulike måtar. Felles for Pär Nilsons artikkel om revisjonen av *Svenska Akademiens ordbok*, SAOB, og Sturla Berg-Olsen, Knut E. Karlsen, Klara Sjo og Bente Selbacks bidrag om *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) er at dei belyser problem som dukkar opp i ein revisjon og korleis ein må tilpasse ambisjonsnivået til ressursane. Hanne Lauvstad viser ved hjelp av arbeidet med litterære sitat i *Det Norske Akademis ordbok*, NAOB, verdien og nytten av kritisk gjennomgang av sitatutval med tanke på periode og kjønn, mens Emma Aadlands kjønnsperspektiv på sitata nytta i BOB belyser kor viktig det er å ha tanke om at ulike brukargrupper kjenner seg representerte i ei ordbok. Også Sanni Nimb og Agnes Aggergard Mikkelsen viser drøftingar som ligg til grunn for endringar i definisjonar, lemmautval og bruksmarkeringar i *Den Danske Begrebsordbog*. Det er her tale om ein thesaurus, ei onomasiologisk ordbok, der ein prøvde å tilpasse begrepshierarkiet til notida i den opphavlege utgåva frå 2014. Men allereie etter få år viste det seg at ei lang rekke omgrep måtte omdefinierast for å vere i betre samsvar med den aktuelle samfunnsutviklinga.

Den tekniske utviklinga er både utfordring og moglegheit for eit ordbokprosjekt. Lars Trap-Jensen belyser bruken av kunstig intelligens, KI, og peiker på positive og negative konsekvensar for leksikografien. Det blir dessutan argumentert for at det er naudsynt å omorganisere informasjonen i ordbøkenes databasar fordi brukarane ikkje lengre hentar informasjonen i den trykte bokas løpande framstilling, noko som inneber at den enkelte artikkelen ikkje kan utnytte forskjellige former for nedarving, men i høgare grad skal presentere brukaren for ein full informasjon.

Eit viktig aspekt med tanke på relevans for framandspråkdbukkarar blir drøfta av Magnus Ahlthorp og Laila Höglund, som tar for seg såkalla *samhällsord* i Lexin, korleis finne lakuner i lemma-lista og korleis ein kan bøte på ein slik mangel.

Historiske ordbøker er rike kjelder til kulturarven og Trausti

Dagsson og Ellert Þór Jóhannsson viser korleis ny teknologi kan gi ei eldre utgiving betre og meir informasjon til publikum. Dei historiske ordbøkene bygger på historisk tekstmateriale, og sikringa og formidlinga av dette kjeldematerialet er av stor verdi. Ásta Svavarsdóttir og Helga Hilmsdóttirs artikkel gir innblikk i arbeidet med historiske ressursar og fordelar og ulemper knytte til digitalisering av desse. Eit hovudbodskap er at så vel eldre kjeldemateriale (setelsamlingar m.v.) og nyare korpus rommar verdifulle data for både leksikografar og språkforskarar, og derfor supplerer kvarandre.

Under ikkje-tematiske bidrag undersøker Poul Hansen Den universelle First-Letter Law og korleis den kan brukast til statistisk undersøking av skandinaviske ordbøker. I *Har de kommit för att stanna? Om nyordslistornas ord i SO och SAOL* presenterer Andreas Brimberg og Linnea Hanell i kva grad neologismar på nyordslista som blir publisert av Språkrådet blir innlemma i samtidsordbøker utgjevne av Svenska Akademien.

Årets *LexicoNordica* byr på meldingar av to danske utgjevingar. Holger Juul ser på 5. utgåva av den offisielle danske rettskrivingsordboka, *Retskrivningsordbogen*, og diskuterer bl.a. ordbokas formidling og framstilling av bøyingsformer, kommareglar og endra staveformer. Forholdet mellom den trykte og den digitale utgåva blir også tematisert. Eva Thelin tar for seg *En rikhaltig ordbok om dialekterna i Thy*, ei ordbok over målføret i Thy i det nordvestlege Jylland. Denne byggjer på ei rik, eldre setelsamling.

Redaksjonen for årets nummer består av dei to hovudredaktørane Henrik Hovmark og Terje Svardal og landredaktørane Liisa Deth Theilgaard (Danmark), Helga Hilmsdóttir (Island), Louise Holmer (Sverige), Caroline Sandström (Finland) og Bente Selback (Noreg).

Redaksjonen har bestemt at tema for neste symposium er:

2026: Brugerundersøgelser og -involvering i nordiske
ordbøger

Redaksjonen tar imot forslag til føredrag og føredragshaldarar på desse symposia og også idéar til tema for framtidige symposium. Vi minner også om heimesida til og nyheitsbrev frå NFL som kjelde til informasjon.

Symposiet blei arrangert i samarbeid med *Språkrådet* i Noreg og *Svenska Akademiens Ordbok*. Vi takkar for godt samarbeid!

Vi takkar landredaktørane for innsatsen i løpet av året. Ein særskild takk går til Kjetil Gundersen, som har gjort ein stor og verdfull innsats som landredaktør for Noreg. Innsatsen var stor både med arbeid med artiklar og å finne føredragshaldarar. Takk til styret, især kasserar Pär Nilsson for ei god hand med det praktiske arbeidet med søknader og gjennomføring av symposiet. Ein takk går også til Laurids Kristian Fahl som enno ein gong har tatt seg av settinga av tidsskriftet og gitt det ei profesjonell utforming. Til slutt takkar vi Nordplus Nordens sprog for økonomisk støtte og Voksenåsen kultursentrum for husrom og mat.

Terje Svardal
leksikograf
Språksamlingane
Universitetet i Bergen
Postboks 7808
NO-5020 Bergen
terje.svardal@uib.no

Henrik Hovmark
lektor, ph.d.
Ømålsordbogen
Institut for Nordiske Studier og
Sprogvidenskab
Københavns Universitet
Emil Holms Kanal 2
DK-2300 København S
hovmark@hum.ku.dk

TEMATISKE BIDRAG

Framtidssäkrade samhällsord i Lexin

Magnus Ahlthorp & Laila Höglund

Much has happened since the Lexin dictionary series for immigrants learning Swedish was first launched in Sweden in the early 1980s. Technical innovations, increased immigration from non-European countries, together with an ongoing digitisation have led to increased demand for advanced Swedish literacy skills and more rapid integration. An increased number of students going through university education today, along with a growing importance of digital literacy across the labour market in previously mainly manual industrial roles, now means an increased need for more specialised terminology for immigrants learning Swedish.

As language usage changes to reflect societal change, regular revision is important to keep contemporary dictionaries up-to-date. Most affected are the so-called *samhällsord* ‘societal terms/words’ in Lexin, words covering specific issues, celebrations of cultural importance, and names of institutions in Swedish society. These are words with additional factual information, which have been included in the Lexin dictionaries as they are “of importance to students wishing to familiarise themselves with modern Swedish society”, as stated in the preface of the Lexin Swedish-English dictionary edition from 1997. “As society evolves, however,” it concludes, “these explanations will gradually become outmoded”. This article focuses on ways of identifying and updating this particular group of lemmas through categorisation and partial automatisisation.

1. Inledning

Lexikonserien Lexin tillkom inom ett svenskt forsknings- och utvecklingsprojekt som initierades 1977 med syfte att utreda och skapa förutsättningar för produktion av språklexikon för invandrare och flyktingar som studerar svenska. Svenskkunskaper är viktiga för invandrades möjligheter att integreras i samhället och språk-

lexikon är ett betydelsefullt hjälpmedel när man ska lära sig ett nytt språk. (Se bl.a. SOU 1983:58, Beijer & Fre Woldu 1997:4ff. För utförligare bakgrund och lägesbeskrivning, se Gellerstam 1999:3.) Språklexikon var även ett viktigt hjälpmedel i genomförandet av riksdagsbeslutet om att främja aktiv tvåspråkighet.

Lexin förvaltas sedan 2007 av Språkrådet, Institutet för språk och folkminnen. I dag består Lexin av tryckta och digitala inlärarlexikon på ett tjugotal olika minoritetsspråk, en bildordbok samt tematiska videofilmer. Den primära målgruppen för Lexinlexikonen är nyanlända invandrare och flyktingar som lär sig svenska och lexikonen i serien är unika i att de i hög grad är utarbetade utifrån denna målgrupps behov och förutsättningar. Urvalet av uppslagsord är anpassat till den vardag de möter i Sverige, med arbete eller studier, familj, skola och bostad och utgörs av högrekventa och vardagliga ord ur allmänspråket, men även specifika ord som invandrare behöver känna till för att kunna orientera sig i det svenska samhället, ord som är vanligt förekommande i politiska eller samhälleliga sammanhang, exempelvis *skatt*, *fackförbund*, *riksdagsval* och *dagis*. Till dessa s.k. *samhällsord* finns förutom språklig information även sakupplysningar. (Se bl.a. Gellerstam 1999:5, Hult 2017:35.) Figur 1 visar artikeln för samhällsordet *medborgarskap* med en förklarande sakupplysning i webbgärnsnittet lexin.se.

1.1. Bakgrund till projektet

Sedan Lexinseriens tillkomst i början av 1980-talet har världen förändrats och med den ordboksanvändarnas behov. Dagens digitala informationssamhälle ställer krav på avancerad litteracitet. Genom ny teknik, fri rörlighet och utbyte av varor och tjänster samt alltmer globala diskussioner och trender i sociala medier förändras språkbruket i snabb takt. En större medvetenhet kring social rättvisa i dagens Sverige påverkar också språket, och ord-



Figur 1: Samhällsorden i Lexin är försedda med extra sakupplysningar eller förklaringar.

boksartiklar där exempelvis föräldrade eller missvisande namn på grupper förekommer kan behöva revideras och uppdateras utan dröjsmål. Ord som söker beskriva samhället, dess beståndsdelar och kulturyttringar påverkas i högre grad av dessa språkförändringar än vardagliga substantiv som rör familj, hem eller husdjur.

Det är därtill av särskild vikt att kategorin av lemman som kallas *samhällsord* är aktuell och relevant för att underlätta för andraspråksinlärare att orientera sig i det svenska samhället. Andraspråksinlärare har ett specifikt lexikografiskt behov av att få förklaringar till en rad kulturspecifika förhållanden som inte självklart förstås av en utomstående och inkluderar ett urval av ord som rör relevanta lagar, samhällsinstitutioner och kulturfenomen (Tarp 1999:107, 2008:143). Samhällsorden är särskilt relevanta att lära sig tidigt för att lära känna och bli en del av det nya samhället (Loenheim & Hult 2020:14). Samhällsinstitutioner, kulturföretelser och lagar i urvalet bör därför vara aktuella och relevanta för Lexinanvändarna.

Åtskilligt har gjorts för att bibehålla Lexinlexikonens relevans, aktualitet och användaranpassning, senast genom den uppdatering och genomgripande metalexikografiska förbättring av den svenska orddatabasen som genomfördes vid Göteborgs universitet 2008–2010 (Malmgren 2012). Även samhällsorden har reviderats genom åren, och urvalet har kompletterats framför allt inom det ekonomiska området (Beijer & Fre Woldu 1997:9), men någon genomgripande översyn och revidering har inte gjorts på senare tid.

Att kontinuerligt revidera och uppdatera samhällsorden i Lexin är nödvändigt för att innehållet ska vara aktuellt och relevant (Se bl.a. Loenheim 2012:302, Loenheim & Hult 2020:50, Höglund & Malmgård 2022:32). Då det finns begränsade medel för Lexin ämnar redaktionen på Språkrådet revidera och uppdatera åtminstone samhällsorden, även om förhoppningen finns att kunna revidera och uppdatera hela ordbasen med jämna intervall.

Med tiden har emellertid den ursprungliga uppmärkning av innehållet i orddatabasen med referens till källtexter som gjordes (Gellerstam 1978:26–27) gått förlorad och det är därför inte helt enkelt att avgöra vilka ord som är samhällsord. Eftersom samhällsorden inte är uppmärkta i Lexins orddatabas eller angivna i de xml-format som används i dag behöver de först och främst identifieras, varefter en revidering och uppdatering med aktuella nya ord kan genomföras. Som ett första steg i arbetet har vi genomfört den här studien.

1.2. Syfte

Syftet med denna artikel är att redogöra för resultaten av vår studie, i vilken vi undersökte möjliga metoder för att identifiera och uppdatera samhällsorden. Vi undersökte även möjligheten att manuellt kategorisera samhällsorden (tematiskt) och märka upp dem för att möjliggöra automatiserad redigering av en delmängd i taget och därmed smidigare och effektivare uppdateringar i framtiden.

2. Samhällsorden

Förutom rent frekvensbaserade skriftspråkliga ord kom Lexins orddatabas att bestå av dels ett ”praktiskt” ordförråd hämtat ur ordindex till ett antal läromedel för invandrarundervisning, diverse vardagliga och talspråkliga ord, ”svåra” ord enligt förståelseundersökningar samt ”samhällsord, dvs. ord som behövs för att man som medborgare skall kunna orientera sig i det svenska samhället (*moms, daghem, kommunalval* etc.)”. (Se Gellerstam 1999:6.)

Samhällsorden i Lexin rör huvudsakligen offentlig sektor (Gellerstam 1999:5) samt kulturspecifika företeelser som *allemansrätt*. Urvalet bygger dels på Ordlista för tolkar (Beijer & Fre Woldu 1997:9), vilken innehåller fackord och samhällstermer inom rättsväsende, allmän försäkring, sjukvård och arbetsmarknad, och dels ingår ”ett tusental ord speciellt utvalda för detta ändamål” (Beijer & Fre Woldu 1997:9). Gellerstam (1978:27) ger exempel på källtexter ur vilka ord inom samhällsinformation till Lexins orddatabas bl.a. har hämtats, titlar som Språkklyftan, Ord vi behöver, Samhälle 1, 2 & 3 samt Socialkunskap (1978:26). I förteckningen över källor till Lexinprojektet finns även ”Tolkordlista” med (1978:26) och bidrog enligt denna med 8 424 uppslagsord, vilket motsvarar 18 % av uppslagsorden i det ursprungliga (obearbetade) källmaterialet.

Det ursprungliga urvalet av samhällsord utarbetades på uppdrag av Invandrarverket och kompletterades till den andra upplagan av Institutionen för svenska språket vid Göteborgs universitet. Samhällsorden försågs med sakupplysningar, förklaringar, och reviderades ”under åren beroende på de fortlöpande ändringar i den verklighet som orden beskriver” (Beijer & Fre Woldu 1997:9). Medvetenheten fanns att man skulle behöva göra systematiska genomgångar för att fånga upp nya samhällsföreteelser, men även om detta sedermera gjordes ”till stor del i samband med översättningsarbetet” (Beijer & Fre Woldu 1997:10) är det oklart när samhällsorden systematiskt reviderats och uppdaterats på senare tid.

Gellerstam (1999:6) beskriver samhällsorden som ”ord för företeelser i det svenska samhället som inte har motsvarigheter i andra kulturer” och tillägger att en exakt sammanvägning av dessa olika typer av ord är svårgörig. Dahl (1999:49) menar att samhällsorden snarast uppfattas som ”ord med utförligare förklaring och sakupplysning”, dvs. deras anknytning till det svenska samhället är inte entydig. Samhällsord återfinns enligt Dahl dels i ord som är ”namnliknande”, dels som en kompletterande upplysning till ord som också ges en vanlig ordförklaring, exempelvis *barnbidrag*. Dahl argumenterar för en snävare definition av vad som räknas som samhällsord i Lexin och menar att typiskt svenska institutioner som *Korpen* hör dit, liksom maträtter som *köttbullar* och *strömming*, men att *addition* inte borde räknas, eftersom samhällsorden bör ha en förankring i den (specifikt) svenska verkligheten. Han påpekar att inte heller kemiska ämnen eller jordarter kan uppfattas som specifikt svenska. Typiska exempel på termer med stark nutidsanknytning till invandrares verklighet är, enligt Dahl, förkortningar från arbetslivet/arbetsförmedlingen.

3. Metod

Vi har identifierat tre olika grundläggande tillvägagångssätt för att ta reda på vilka samhällsorden i Lexin är och vilka ord som inte ingår i Lexin men som kan vara aktuella att klassa som samhällsord och inkludera i Lexin:

- utgå från en lista med kända samhällsord
- utgå från en definition
- identifiera ord som förekommer oftare i viss typ av text än i ”text i allmänhet” (Detta tillvägagångssätt använder vi oss inte av i den här studien.)

Dessa tillvägagångssätt producerar vad vi kan kalla för råmaterial för studier, vilket sedan måste gås igenom manuellt.

Vi har valt att göra tre olika delstudier, där de två första är manuella, medan den tredje använder språkteknologiska metoder. Delstudie ett utgår från en lista med kända samhällsord och använder kategorisering för att identifiera ord som förekommer i samma kontext och påvisa luckor som nya ord kan fylla. Delstudie två undersöker tänkbara samhällsordskandidater utifrån en definition, och den tredje använder en lista med kända samhällsord för att automatiskt ta fram kandidater och med hjälp av definitionen avgöra vilka av dessa kandidater som är rimliga.

3.1. Delstudie 1: Kategorisering

Den här delstudien syftar till att undersöka hur samhällsorden fördelar sig över olika kategorier av ord (ämnesområden) och huruvida en kategorisering kan synliggöra eventuella luckor i samhällsordsbeståndet. Listan som användes var alla ord (lemman) i xml-filen för *Lexins svenska lexikon (Svenska ord)*, version 3 (hädanefter LSL 3) som har försetts med fältet (xml-taggen) ”Explanation”. Detta gjorde det möjligt att använda en automatiserad sökning och identifiering av samhällsorden. Ord som har tagits bort i LSL 4 togs bort från listan, och alla ord i LSL 4 som har fältet (xml-taggen) ”Explanation” lades till.

Eftersom urvalet till samhällsorden i Lexin gjordes till stor del utifrån Ordlista för tolkar använde vi oss av den befintliga indelningen och kategorierna i dessa: *Arbetsmarknad och pension*, *Rättsväsende*, *Socialförsäkring* samt kategorin *Diverse*. Den kliniskt orienterade kategorin *Medicin och psykiatri* i Ordlista för tolkar anpassade vi dock till det mer allmänspråkliga innehållet i Lexin genom att benämna den *Hälsa och sjukvård*.

Vi jämförde ordbasen för Ordlista för tolkar (Tolkordbasen) med vår lista över samhällsord (samhällsordlistan) för att undersöka hur många ord som är direkt hämtade ur Tolkordbasen och i hur stor utsträckning vi kan använda kategorierna ur den. Samtliga ord i samhällsordlistan som även återfanns i Tolkordbasen tilldelades samma kategoriindelning.

3.2. Delstudie 2: Definitionsbaserad undersökning

Denna delstudie syftar till att undersöka hur en definition kan användas för att hitta nya samhällsord. Först enades vi om en definition av samhällsord för denna studie: ”Ord som definierar eller beskriver det svenska samhället, samhällsinstitutioner och företeelser i det, eller som är viktiga i samhällsdebatten, kulturyttringar och traditioner.”

Vi valde ut 200 ord slumpmässigt av de 4 609 ord som tillkommit i LSL 4 i förhållande till LSL 3. Dessa annoterades av två personer var för sig genom att markera om ordet uppfyllde definitionen av samhällsord eller inte. För att mäta samstämmigheten i annoteringen beräknades Cohens kappa (Landis & Koch 1977).

3.3. Delstudie 3: Utökning av lista

Denna delstudie syftar till att med en befintlig lista av samhällsord som grund se om det är möjligt att automatiskt identifiera nya kandidater till samhällsord. Listan som användes var alla ord (lemman) i LSL 3 som har fältet (xml-taggen) ”Explanation”. Ord som har tagits bort i LSL 4 togs bort från listan, och alla ord i LSL 4 som har fältet (xml-taggen) ”Explanation” lades till. Eftersom Lexin använder presens som verbuppslagsform (Malmgren 1999:4), vilket motiveras av att man lättare kan generera övriga verbformer utifrån presens än utifrån infinitiv, anpassades orden så att de överensstämde med den traditionella lemmaformen (infinitiv för

verb). Dessa 465 ord benämndes ”seed-ord” (Riloff & Shepherd 1997, Ahlthorp et al. 2016).

För att tillverka ett ordrum (vektorrum) baserat på distributionell semantik (se till exempel Forsberg & Sköldberg 2022) användes korpusamlingen SVT nyheter (2004–2023), 240 407 472 tokens, och korpusen 8 Sidor, 4 998 634 tokens, totalt ungefär 245 miljoner tokens (hädanefter kallat ”textmaterialet”) (Språkbanken Text 2024a, 2024b). Ordrum baserade på distributionell semantik placerar paradigmatiskt lika ord nära varandra beroende på vilka ord de samförekommer med.

Ordrummet vi tillverkade var av typen Word2Vec (Le & Mikolov 2014) med 100 dimensioner, med en lägsta frekvens på 20 förekomster i textmaterialet och en fönsterstorlek på 5 tokens, det vill säga att algoritmen tar hänsyn till 5 tokens före och 5 tokens efter varje undersökt token i en mening.

Orden samlades in i sin lemmatiserade form, närmare bestämt attributet ”lemma” i Korps XML-filer. Om lemma inte fanns tillgängligt för ordet användes ytformen. Vissa ord har kolon följt av ett tal i lemmaattributet, detta rensades bort innan vidare bearbetning. Från korpusen 8 Sidor rensades meningar som bara bestod av ”8 SIDOR” och ”8 SIDOR / TT” bort.

Ordvektorrummet reducerades från de ursprungliga 100 dimensionerna först till 40 dimensioner med en linjär metod för dimensionalitetsreducering, principalkomponentanalys, och sedan till två dimensioner med t-SNE, en icke-linjär dimensionalitetsreduceringsmetod som förutom att reducera dimensionerna även i viss mån grupperar vektorerna (van der Maaten & Hinton 2008).

För varje seed-ord hittades motsvarande ordvektor, en ”seed-vektor”. Kandidater för nya samhällsord valdes sedan ut genom att för varje vektor bland seed-vektorerna hitta de närmaste 10 vektorerna i det reducerade ordrummet.

Kandidaterna sorterades enligt avståndet till den närmaste seed-vektorn, och de 100 kandidaterna med kortast avstånd val-

des ut för utvärdering. Utvärderingen skedde genom att välja ut 100 slumpmässigt valda ord från de ord i textmaterialet som hade minst 20 förekomster och blanda dessa med de 100 kandidaterna. Inga av de slumpmässigt valda orden råkade bli någon av kandidaterna. De totalt 200 orden presenterades i en slumpmässig ordning för två annoterare (samma ordning för båda annoterarna), och annoterarna fick samma uppgift som ovan, att markera om ordet uppfyllde definitionen av samhällsord eller inte. För att mäta samstämmigheten i annoteringen beräknades Cohens kapp (Landis & Koch 1977).

Listan med de 200 orden gicks igenom av båda annoterarna och varje ord som någon hade markerat diskuterades och en ensad uppfattning antecknades utan att annoterarna visste vilka ord som var kandidater och vilka som var slumpmässigt utvalda. Efter ensningen delades orden upp i kandidater och slumpmässigt utvalda. Ett χ^2 -test (Campbell, David & Walters 2007:132) genomfördes för att se om kandidaterna hade signifikant fler samhällsord än de slumpmässigt utvalda.

4. Resultat

4.1. Delstudie 1: Kategorisering

Av totalt 465 ord i samhällsordlistan fanns 234 ord (50 %) även i Tolkordbasen och indelades därmed i samma kategorier. Vid en sortering av listan fick vi en överblick över mängden ord per kategori för de ord som även ingick i Tolkordbasen: Arbetsmarknad och pension 68 ord, Diverse 14 ord, Hälsa och sjukvård 28 ord, Rättsväsende 70 ord, Socialförsäkring 54 ord.

231 ord fanns inte med i Tolkordbasen. Dessa indelades i kategorier utifrån semantiskt innehåll. Vi tillämpade så långt det

var möjligt samma kategorisering som i Tolkordbasen, med viss försiktighet och i förekommande fall semantisk granskning. Vid denna indelning utökades kategorin Diverse i Lexin från 14 ord till 91, varför en indelning av givna ord även i alternativa (under-) kategorier, exempelvis Sveriges landskap, kan vara till nytta.

Delstudie 1 visade att kategorier kan vara användbara för att synliggöra luckor i en samling av utvalda ord, vilka kan utgöras av såväl luckor (saknade ämnesområden) bland de kategorier som ingår i listan och luckor (saknade ord) inom en viss kategori av ord. Exempelvis återfanns såväl *valborgsmässoafton* (se figur 2) som *fruntimmersveckan* i listan över samhällsord, men inte *mid-sommar*.



The screenshot shows the Lexin website interface. At the top, there is a dark teal header with the 'LEXIN' logo on the left, a language dropdown menu set to 'svenska' in the center, and a hamburger menu icon on the right. Below the header is a search bar containing the text 'valborgsmässoafton' with a magnifying glass icon to its right. The search results are displayed in a light blue box. The main entry is 'valborgsmässo | afton' with the phonetic transcription '[²və:lbårjsmes:oaf:tån]'. Below this, it is identified as a 'substantiv' and lists related terms in parentheses: '(valborgsmässo | afton · valborgsmässoaftonen · valborgsmässoaftnar · valborgsmässoaftnarna)'. To the right of the word is a speaker icon. Below the word, it says 'den 30 april (som anses symbolisera vårens ankomst)'. Underneath, there is an 'Exempel:' section with a bullet point: 'de samlades runt bålet på valborgsmässoafton och sjöng vårsånger'.

Figur 2: Artikeln för ordet valborgsmässoafton på lexin.se, vilken har försetts med extra information inom parentes, men ingen specifik förklaring.

Kategorisering visade sig även användbar för att synliggöra regelbundenheter och inkonsekvenser vad gäller redaktionell hantering. I användargränssnittet har vissa av orden getts såväl definition som förklaring och andra har försetts med enbart en förklaring.

Lemmat *Högsta domstolen* har exempelvis såväl en definition som förklaring (där även ordet ”Förklaring:” har angetts), vilket visas i figur 3. Jämför vidare orden *inkomstpension*, som har försetts med förklaring, och *ålderspension*, där ingen förklaring har angetts.



Figur 3: Ordet *Högsta domstolen* på lexin.se har försetts med både definition och förklaring.

4.2. Delstudie 2: Annotering av 200 nya ord ur LSL 4

Av de 200 undersökta orden i delstudie 2 var de två annoterarna överens om att 9 av orden uppfyllde den överenskomna definitionen av samhällsord. För ytterligare 17 ord var annoterarna oense om ordet uppfyllde definitionen. Samstämmigheten enligt Cohens kapp beräknades till 0,47 (måttlig enligt Landis & Koch 1977), se tabell 1.

	Ja	Nej
Ja	9	5
Nej	12	173

Tabell 1: Annoteringen av de 200 nya orden ur LSL 4 gav en samstämmighet på 0,47 (Cohens kapp, måttlig enligt Landis & Koch 1977).

4.3. Delstudie 3: Utökning av lista

I delstudie 3 förekom 129 438 ord minst 20 gånger i textmaterialet (SVT nyheter + 8 Sidor) och utgjorde därmed underlaget för ordrummet. Bland dessa (hög)frekventa ord fanns 392 av seed-orden. Av de totalt 465 seed-ord som identifierades i befintliga Lexin kunde 73 seed-ord alltså inte användas i framtagandet av kandidater, eftersom de inte fanns med i textmaterialet minst 20 gånger (och därmed beräknades inga vektorer för dem). Resultatet av reduceringen till två dimensioner kan ses i figur 4, där varje ordvektor representeras av en punkt i figuren. Seed-vektorerna är märkta med mörkgul färg och kandidaterna med turkos färg i figur 4. En detaljbild med ord relaterade till fackföreningar finns i figur 5.

Av de 200 orden som valts ut för annotering var de två annoterarna överens om att 35 av orden uppfyllde definitionen av samhällsord. För ytterligare 38 ord var annoterarna oense om ordet uppfyllde definitionen. Samstämmigheten enligt Cohens kapp beräknades till 0,53 (måttlig enligt Landis & Koch 1977), se tabell 2.

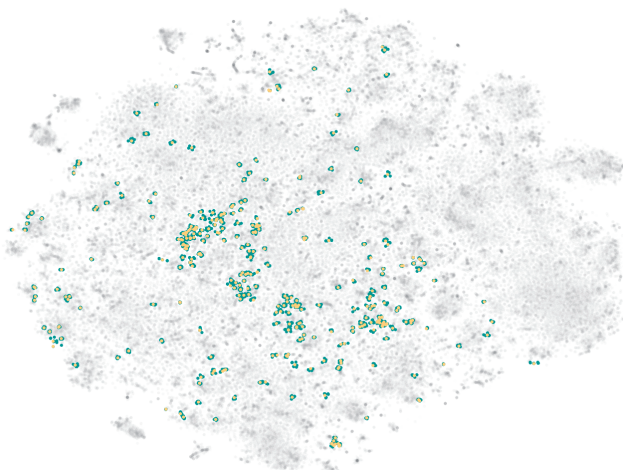
	Ja	Nej
Ja	35	8
Nej	30	127

Tabell 2: Annoteringen av 100 kandidater och 100 slumpmässiga ord (totalt 200 ord) gav en samstämmighet på 0,53 (Cohens kapp, måttlig enligt Landis & Koch 1977).

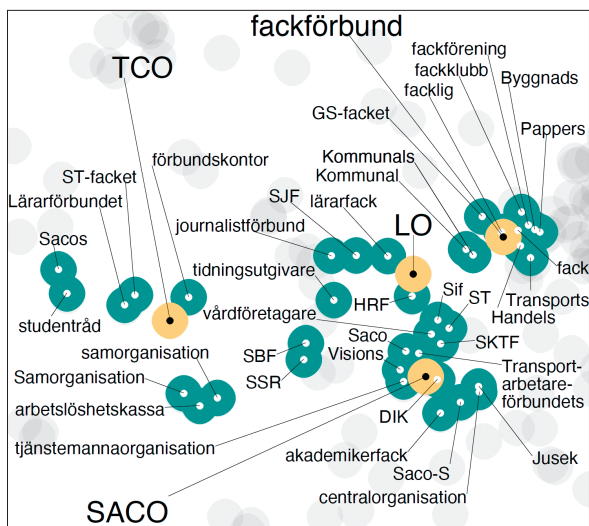
Ensningdiskussionen gav även upphov till att några ord som ingen hade markerat blev markerade vid ensningen.

	Kandidater	Slumpmässiga ord
Samhällsord	57	14
Inte samhällsord	43	86

Tabell 3: Resultat efter ensning.



Figur 4: Alla ord som förekommer minst 20 gånger i textmaterialet, med seed-vektorna märkta med mörkgul färg och kandidaterna märkta med turkos färg.



Figur 5: Detalj som visar ord som relaterar till fackföreningar, med seed-vektorna märkta med mörkgul färg (centrum visas med en svart prick) och kandidaterna märkta med turkos färg (centrum visas med en vit prick). Ord som inte är kandidater visas med en något mörkare grå i förhållande till figur 4.

Resultatet av den ensade annoteringen visas i tabell 3, där 57 kandidater av 100 ansågs vara samhällsord. En fullständig lista på dessa 57 kandidater finns i tabell 4.

Chi2-testet gav resultatet 38,52, vilket gör att vi kan förkasta nollhypotesen att skillnaden mellan kandidaternas 57 % bedömda samhällsord och de slumpmässiga ordens 14 % bedömda samhällsord bara beror på slump med en sannolikhet $p < 0.001$. Skillnaden mellan listan av kandidater och listan av slumpmässiga ord är alltså signifikant.

socialbidrag	delbetänkande
Bris	april-firande
vårdboende	länstyrelse
Region	Östersjölandskapen
hockeyallsvenska	elevkår
kallelse	miljöparti
Elitserien	V
prövningstillstånd	regionval
Sjuhäradsbygden	förvaltningsrätt
f-kassa	steroid
polisstyrelse	pastorat
Gävleborg	påskdag
nyårsdag	riksbank
a-kassa	hetvägg
arbetsmiljöverk	juldag
midsommardag	allsvenska
vänsterparti	anhöriginvandring
byggrätt	kammarrätt
universitetsadjunkt	Sörmland
Västernorrland	ob-tillägg
betänkande	Nordsjön

registerutdrag	Naturvårdsverk
brottmål	brottsmål
skatteväxling	arbetsbrist
europaparlamentsval	advokatfirma
allhelgonahelg	familjeåterförening
påskafton	studiestöd
valkuvert	sörmland
sjuklön	

Tabell 4: De 57 automatiskt framtagna kandidater till samhällsord där annoterarnas konsensus var att ordet skulle räknas som samhällsord.

5. Diskussion

Delstudie 1 visade att en indelning i kategorier kan vara användbar för mer frekventa och tidsmässigt hanterbara framtida deluppdateringar genom att man kan revidera och uppdatera en eller några kategorier i taget. Vi fann att det vore lämpligt att märka upp på det här sättet och när det är möjligt använda samma kategorier som i Ordlista för tolkar.

För att möjliggöra detta bör man vara konsekvent med att lägga till kategorier i xml vid revidering och uppdatering, samt därvid tillämpa ett konstant regelverk. Möjlig märkning kan vara att lägga till ett attribut på taggen "Explanation" alternativt på taggen för lemmat samt döpa attributet (exempelvis "Category"). Vissa ord kan tillhöra flera olika kategorier. Det är därför en god idé att redan från början bestämma huruvida man vill tillåta indelning i flera kategorier, åtskilda med kommatecken exempelvis. Man bör givetvis vara konsekvent och tillämpa samma principer vid kategorisering och märkning även över tid.

Både i delstudie 2 och 3 var samstämmigheten mellan annoterarna måttlig. Samstämmigheten kan antagligen förbättras genom att förfina definitionen av vad som utgör samhällsord (se avsnitt

3.2) och genom löpande diskussioner mellan annoterarna om gränsdragningar. Var gränsen går mellan vad som är samhällsord och vad som är fackord eller termer inom något av områdena politik, ekonomi eller juridik är inte alltid helt lätt att avgöra.

Delstudie 3 visade att det var signifikant mer effektivt att leta samhällsord i en automatiskt extraherad lista baserad på närhet i ett ordrum än att gå igenom slumpmässigt valda ord. Det är därför ett sätt att få tag i så många samhällsord som möjligt, men däremot inget säkert sätt att få en uttömmande lista. Där kan en lista indelad i kategorier utgöra ett bra komplement genom att visa på luckor inom en viss ordkategori eller domän och därmed möjliggöra analys och fokuserad (manuell eller automatiserad) uppdatering av orden i en eller några kategorier åt gången. Begränsad arbetsinsats krävs för att uppdatera en avgränsad del av listan och gör det därmed möjligt att mer regelbundet och effektivt uppdatera samhällsorden i Lexin i framtiden.

Litteratur

Ordböcker och korpusar

LSL 3 = *Lexins svenska lexikon*, version 3 [Nätversionen av] *Svenska ord: med uttal och förklaringar* (1995). 3 uppl. Stockholm: Myndigheten för skolutveckling/Liber.

LSL 4 = *Lexins svenska lexikon*, version 4 [Nätversionen av] *Svenska ord* (2011). 4 uppl. Språkrådet, Institutet för språk och folkminnen.

Lexin svensk-engelskt lexikon (2004). Stockholm: Myndigheten för skolutveckling.

Språkbanken Text (2024a). SVT nyheter [Data set]. Språkbanken Text. <<https://doi.org/10.23695/vvd5-j8482>>.

Språkbanken Text (2024b). 8 Sidor (Uppdaterad: 2024-03-07) [Data set]. Språkbanken Text. <<https://doi.org/10.23695/f6ds-fo45>>.

Annan litteratur

- Ahltorp, Magnus, Maria Skeppstedt, Shiho Kitajima, Aron Henriksson, Rafal Rzepka & Kenji Araki (2016): Expansion of medical vocabularies using distributional semantics on Japanese patient blogs. I: *Journal of Biomedical Semantics* 7 (58), 1–18. <<https://doi.org/10.1186/s13326-016-0093-x>>.
- Beijer, Mai & Kiros Fre Woldu (1997): *Detta är LEXIN. Lexikon för invandrare. Lägesbeskrivning våren 1997*. Stockholm: Skolverket.
- Campbell, Michael J., David Machin & Stephen J. Walters (2007): *Medical Statistics: a Textbook for the Health Sciences*. 4. ed. Chichester: Wiley.
- Dahl, Adolf (1999): Lexinlexikonen och översättaren. I: *LexicoNordica* 6, 47–53.
- Forsberg, Markus & Emma Sköldberg (2022): Ordvektorer i lexicografiskt arbete. I: Elena Volodina, Dana Dannélls, Aleksandrs Berdicevskis, Markus Forsberg & Shafqat Virk (eds.): *LIVE and LEARN – Festschrift in honor of Lars Borin*. Gothenburg: Department of Swedish, Multilingualism & Language Technology. 46–50.
- Gellerstam, Martin (1999): LEXIN – Lexikon för invandrare. I: *LexicoNordica* 6, 3–18.
- Gellerstam, Martin (1978): Att välja sina ord: om ordurval till enspråkiga ordböcker. *Rapporter från Språkdata* 9. Göteborgs universitet, Institutionen för språklig databehandling.
- Höglund, Laila & Sofia Malmgård (2022): *Framtidens Lexin – forskningsöversikt*. Rapporter från Språkrådet 13. Institutet för språk och folkminnen.

- Landis, J. Richard & Gary G. Koch (1977): The Measurement of Observer Agreement for Categorical Data. I: *Biometrics* 33(1), 159–174. <<https://doi.org/10.2307/2529310>>.
- Loenheim, Lisa & Ann-Kristin Hult (2020): *Framtidens Lexin – forskningsöversikt*. Rapporter från Språkrådet 12. Institutet för språk och folkminnen.
- Malmgren, Sven-Göran (2012): Från Svenska ord 3 (Lexin) till Svenska ord 4. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi* 11. *Rapport från Konferensen om lexikografi i Norden, Lund 24–27 maj 2011*. 454–465.
- Malmgren, Sven-Göran (1999): De svenska LEXIN-ordböckerna ur ett (ställföreträdande) användarperspektiv. I: *LexicoNordica* 6, 79–90.
- van der Maaten, Laurens & Geoffrey Hinton (2008): Visualizing data using t-SNE. I: *Journal of Machine Learning Research* 9(86), 2579–2605.
- Le, Quoc & Tomas Mikolov (2014): Distributed representations of sentences and documents. I: *Proceedings of the 31st International Conference on Machine Learning (ICML ‘14)* 32. 1188–1196.
- Riloff, Ellen & Jessica Shepherd (1997): A Corpus-Based Approach for Building Semantic Lexicons. I: *Proceedings of the Second Conference on Empirical Methods in Natural Language Processing (EMNLP-2)*. 117–124.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2 uppl. Stockholm: Norstedts Akademiska Förlag.
- Tarp, Sven (2008): *Lexicography in the Borderland between Knowledge and Non-Knowledge. General Lexicographical Theory with Particular Focus on Learner’s Lexicography*. Lexicographica Series Maior 34. Tübingen: Max Niemeyer Verlag.

Tarp, Sven (1999): Lørnerordbøger for indvandrere og andet godtfolk. I: *LexicoNordica* 6, 107–132.

Magnus Ahltop
språkteknolog
Nakajima Koen Research Institute
Hildebergsvägen 26
SE-117 62 Stockholm
magnus@nakajimakoen.org

Laila Höglund
projektledare för Lexin
Språkrådet, Institutet för språk och
folkminnen
Alsnögatan 11
SE-116 41 Stockholm
laila.hoglund@isof.se

Revitalisering af *Íslensk orðsifjabók*, islandsk etymologisk ordbog

Trausti Dagsson & Ellert Þór Jóhannsson

This article discusses *Íslensk orðsifjabók*, an etymological dictionary of modern Icelandic, its structure and material. This work came out in print in 1989 but has recently been revitalized and made available on a new online platform. The platform has several innovative features, including visual graphs showing relations between words and languages. The online version improves access to the dictionary, though some features need improvement and copyright restricts further development.

1. Indledning

Íslensk orðsifjabók (IO) er den eneste etymologiske ordbog over det moderne islandske sprog. Ordbogen gør rede for det islandske ordforråds historie og har ca. 25.000 opslagsord. IO blev første gang udgivet på tryk i 1980'erne (Magnússon 1989), men senere er materialet også blevet publiceret online. IO er et forfatterværk under ophavsret. Det betyder at værket forbliver som det er, og at det ikke kan revideres eller opdateres indholdsmæssigt.

I denne artikel giver vi et overblik over IO, de typer oplysninger der præsenteres i værket, og det som vi kalder revitaliseringen af IO. Vi diskuterer en ny webplatform, IOW, hvor hele ordbogsmaterialet er blevet gjort tilgængeligt og søgbart samt opgraderet med diverse søgemuligheder og andre forbedringer, idet ordbogens egentlige indhold forbliver det samme som før.

Artiklen er struktureret på følgende vis. Efter indledningen gives der i andet afsnit en kort baggrund om ordbogen og dens forfatter efterfulgt af en diskussion af de typer informationer man finder i etymologiske ordbøger. Vi giver et overblik over islandsk

og dets rødder i norrønt og urgermansk, og ser nærmere på et typisk opslag i den trykte udgave.

I tredje afsnit præsenterer vi IO som en digital udgave og dens indhold, herunder strukturen og organiseringen af informationer. Vi redegør for hvordan IOW er blevet til, og hvordan revitaliseringsprocessen er foregået. Vi adresserer også nogle udfordringer, både praktiske og lidt mere fundamentale. Vi diskuterer hvad det betyder for IO at der er tale om et statisk værk som begrænser mulighederne for videreudvikling, og vi ser på fremtidsperspektiverne for hvordan man fortsat kan forbedre platformens indhold. Til sidst er der afsluttende bemærkninger.

2. Baggrund

IO er udarbejdet af Ásgeir Blöndal Magnússon (1909-1987) som var lingvist og leksikograf ved *Orðabók Háskólans*, det leksikografiske institut ved Islands Universitet, og leder af instituttet i mange år. Han arbejdede sideløbende på sit etymologiske opslagsværk allerede fra 1960'erne, men først efter at han gik på pension i 1980, kunne han fokusere mere på arbejdet på denne ordbog. Desværre døde Ásgeir Blöndal før *Íslensk orðsifjabók* udkom endeligt på tryk i 1989, men inden da nåede han at skrive indledningen og korrekturlæse størstedelen af det færdige manuskript (jf. Magnússon 1989:xxii).

2.1. Etymologiske oplysninger

Etymologiske ordbøger udgør en særlig type ordbøger der kan beskrives som diakroniske, da de redegør for ordenes historiske udvikling (jf. Svensén 1987:18). Der er en lang tradition for etymologiske ordbøger for europæiske sprog. Ordbøgerne kan afvige betydeligt fra hinanden med hensyn til både faglig rækkevidde

og graden af leksikografisk detaljeringsniveau (jf. Malkiel 1976). Det historiske plejer at udgøre den vigtigste del af etymologiske oplysninger, fx information om ordenes oprindelse, den tidligst kendte form af et ord og hvilken type relationer ordet har til andre beslægtede ord (jf. Svensén 2004:398-399).

Ord kan grupperes efter hvordan de er blevet en del af sproget. De fleste ord hører til en af de følgende tre kategorier (jf. diskussion i Kvaran 2005:357):

- Arveord som gennemgår regelmæssige lydændringer og morfologiske udviklinger over tid.
- Låneord som et sprog overtager fra et andet sprog, typisk gennem kulturel eller sproglig kontakt.
- Nydannelser, ofte afledte og sammensatte ord som tit er dannet af eksisterende ordstof.

I tilfælde af et arveord kan man spore dets rødder tilbage i tiden, ofte til tilsvarende ord fra ældre sprogstadier og muligvis rekonstruktioner af hypotetiske urformer som ikke er belagt i skriftlige kilder. Tit kan brugeren danne sig et overblik over den historiske udvikling, dvs. hvordan ordet har ændret sig i form og betydning over tid, herunder fonetiske, morfologiske eller semantiske ændringer, og for at illustrere udviklingen bliver der oplistet kognater, dvs. beslægtede ord i samme eller andre sprog der deler en fælles rod eller oprindelse og viser sproglige forbindelser.

En del ord er kommet ind i et sprog ad andre veje end fra ældre sprogstadier som arveord. Mange af dem ville høre til kategorien låneord i bred forstand. Låneord bevarer ofte træk fra deres oprindelsessprog og er ikke et resultat af naturlig intern sproglig udvikling. Fremmedord er en type låneord. I disse tilfælde er der tale om lånehistorie, dvs. information om hvorvidt ordet er lånt fra et andet sprog og muligvis også hvornår det skete, ofte med angivelse af relevante kilder.

En anden kategori af ord består af nydannelser som ofte er sammensætninger og afledte ord: Det kan være ord som er tæt beslægtet med både arveord og låneord, men som ikke har direkte tilsvarende former på ældre sprogstadier eller i beslægtede sprog.

Ved alle kategorier af ord kan det være relevant at diskutere semantisk udvikling, dvs. ændringer i ordets betydning, herunder fx udvidelse eller indsnævring af betydningen. Det samme gælder oplysninger om den tidligst belagte form af ordet i historiske kilder. I nogle tilfælde kan det være gavnligt at diskutere variationer inden for det samme sprog hvis de eksisterer, fx dialektale former eller regionale anvendelser der har afvejet fra standardformen.

I mange tilfælde er der tvivl eller uklarhed vedrørende ordenes herkomst, fx kan låneord være meget gamle eller kommet ind i tidligere stadier af sproget, og de bliver på den måde også en slags arveord. Låneoversættelser eller kalker er også en slags blanding af låneord og nydannelser. Beslægtede sprog kan have identiske processer ved orddannelse uafhængigt af hinanden, så det er ikke altid muligt at se om et ord er arvet i to beslægtede sprog fra et fælles tidligere sprogstadium, eller om det er dannet uafhængigt i to sprog på baggrund af lignende morfologiske processer. Der findes andre kategorier af ord, såsom neologismer eller forskellige analogiske formationer, men de bygger tit også på eksisterende ordstof.

Som man kan se af dette korte sammendrag, er der mange typer informationer og mange forskellige processer i spil når man skal gøre rede for et ords oprindelse og slægtskab med andre ord. Det er vigtigt at en etymologisk ordbog kan fremsætte disse typer oplysninger på en klar og tydelig måde, og at brugeren kan danne sig et hurtigt overblik over de store træk i et ords historie.

2.2 Islandsk og dets rødder

For at kunne diskutere islandske etymologier er det nødvendigt at gøre rede for hvordan islandsk hænger sammen med andre beslæg-

tede sprog. Islandsk er et nordisk sprog, dvs. i familie med andre sprog i de nordiske lande (jf. Haugen 1976). Sammen med de øvrige nordiske sprog repræsenterer islandsk de nordgermanske sprog som er den ene af de tre germanske sproggrene (vestgermansk, østgermansk og nordgermansk) som har mange udviklingstræk til fælles (jf. Harbert 2006). Germansk er en af de ti undergrupper som tilsammen udgør indoeuropæiske sprog: germansk, romansk, græsk, keltisk, balto-slavisk, indo-iransk, albansk, armensk, tokarisk og anatolisk. Man diskuterer stadig relationerne mellem forskellige undergrupper, og hvordan de har udviklet sig fra protoindoeuropæisk (se diskussion i Olander 2022).

2.3 Struktureringen af oplysninger i IO

I figur 1 kan man se etymologiske oplysninger fra den trykte IO som knytter sig til ordet *vatn* 'vand' (Magnússon 1989:1110). Her er der en del forskellige oplysninger, og man kan se hvor veldokumenteret etymologien til dette ord er.

vatn h. 'vökvi, lögur, blávatn; stöðuvatn, fljót, ...'; sbr. fær. og nno. *vatn*, sæ. *vatten*, d. *vand* (fsæ. *vatn*, fd. *watn*, *wadn*), gotn. *wato*; < germ. **watan*-, gamall n-st. sem líkl. er leif af heteróklítískri beygingu (með (*e*)r-endingu í nf. og þf. en (*e*)n-endingu í öðrum aukaföllum et.), sbr. fe. *wæter*, fsax. *water*, fhþ. *wazzer* (ne. *water*, nhþ. *wasser*). Sk. fsl. og rússn. *voda*, hett. *ya-a-tar* og gr. (*h*)*yídōr* (hljsk.), lith. *vanduō* (s.m.) og lat. *unda* 'alda'. Óvíst er um tengsl við nno. *vete*, *vetu*, *vutu* kv. 'vætuborin lægð, blautt engi, ...'; < germ. **wetōn*? Af *vatn* er leidd so. **vatna** 'brynnna, fljóta yfir, ...', sbr. fær., nno. og sæ. *vatna* s. og d. *vande* í svipaðri merk.; *vatn* (ie. *(*a*)*ue-d-* e.t.v. < *(*H*₂)*ue-d-*) líkl. fjarskylt *aur* (1) og *úr* (2), *væsa* (1) og *unn(u)r* (2) (ie. *(*a*)*ue-r-*, *(*a*)*ue-s-*, *(*a*)*ue-t-*). Sjá -*vetni*, *vát(u)r*, *votur*, *vaska*, *vetur*, *otur* (1), sbr. og *vatli*, *vatr*, *vaturpassi*, *vati* (1), *væska* og *væta*.

Figur 1: Opslagsordet *vatn* 'vand' fra den trykte ordbog.

Her får brugeren klart overblik over beslægtede ord i andre germanske og indoeuropæiske sprog. Man kan se at ordet oprindeligt har haft to vekslende stammetyper der endte på *r* eller *n*. Denne veksling har overlevet indtil i det mindste germansk tid, som det fremgår hvis vi ser på resultatet i norrønt/islandsk: *vatn* i modsætning til *water/wasser* i engelsk og tysk. Det fremgår også tydeligt at der er flere beslægtede ord der er attesteret i islandsk. Verbet *vatna* 'vande' er et afledt ord markeret med fed skrift. Andre beslægtede ord som tillige er opslagsord i ordbogen, er gengivet med kursiv, fx ordet *aur* 'vådt mudder' der stammer fra samme rod, adjektivet *vátr/votur* 'våd' og verbet *vaska* 'vaske'. Andre interessante slægtskaber er ordet *vetur* 'vinter' (oprindeligt betydning måske 'våd-sæson') og *otur* 'odder' = 'vand-dyr'. De ord vi ser allernederst i artiklen, som *vatr* og *vaturpassi*, viser sig at være specialiserede låneord der er lånt fra tysk, muligvis via dansk. Her er *r* en del af stammen, hvilket er et vest-germansk træk og afslører ordenes oprindelse som låneord i islandsk.

3. En digital etymologisk ordbog

Som det fremgår af afsnit 2, er det ganske mange oplysninger som præsenteres i en etymologisk ordbog. Det kan være udfordrende at finde frem til dem alle sammen i et trykt værk hvor man kun kan søge alfabetisk efter opslagsord. Informationernes natur og kompleksitet gjorde at man i 2010'erne indså at det ville øge IO's brugbarhed markant hvis ordbogen kunne digitaliseres og gøres tilgængelig og søgbar online.

3.1 Oprindelige digitale filer

IO var et af de første publicerede værker fra *Orðabók Háskólans* der blev produceret digitalt. Selvom Ásgeir Blöndal Magnússon

udførte størstedelen af sin forskning manuelt, dvs. med noter nedskrevet på papir, blev hans udkast behandlet digitalt før publikationen. Materialet blev i første omgang tastet ind på en computer, og efterfølgende blev bogen struktureret og sat op med TeX-software, hvilket var en god måde at håndtere de mange forskellige bogstaver og diakritiske tegn som kan forekomme når man har eksempler fra mange forskellige sprog samt rekonstruerede urformer (jf. indledningen i Magnússon 1989). Det resulterede i en klar og læselig trykt bog der fik en positiv modtagelse af alle der var interesserede i det islandske ordforråds historie.

Efter at den trykte bog udkom i 1989, blev de oprindelige TeX-filer bevaret i det leksikografiske instituts digitale arkiver. Ikke alle computerfiler fra firserne har imidlertid overlevet opbevaring og flytning fra en computer til en anden. Det var heller ikke tanken, fordi man dengang mente at når den trykte bog var udkommet, var der ingen særlig grund til at gemme gamle digitale filer, især ikke hvis de optog meget plads på computerne. Heldigvis blev TeX-filerne fra arbejdet med IO skånet, hvilket gjorde arbejdet meget nemmere da man for alvor begyndte at overveje en online digital version af bogen tyve år senere.

3.2 Digital udgave af IO på en webportal

Udarbejdelsen af en digital udgave af IO blev udført i forbindelse med en ny webportal, *Málið.is*, der blev lanceret i 2016 (jf. Steingrímsson et al. 2018). Formålet med portalen var at give adgang til en række sprogdata-baser som Árni Magnússon-Instituttet for Islandske Studier på det tidspunkt havde ansvar for, herunder *Beygingarlýsing íslensks nútímamáls* (BÍN, 'database over alle ords bøjningsformer'), *Íslensk nútímamálsorðabók* (OMI, 'Ordbog over Moderne Islandsk'), *Íslensk stafsetningarorðabók* (ÍS, 'Islandsk retskrivningsordbog') og *Íðorðabankinn* (ÍOB, 'termbanken'). Det blev besluttet at denne første version af *Málið.is* også skulle inde-

holde adgang til den etymologiske ordbog, og man gik derfor i gang med at klargøre materialet til en online udgave.

Selvom de oprindelige TeX-filer kunne tilgås, indeholdt de en stor mængde tilpasset opmærkning, hvilket betød at det ikke var muligt at parse filerne efter en moderne standard (TeX-parser). For at imødegå denne udfordring blev der skrevet en brugerdefineret parser i Python som læste alle filerne igennem og konverterede dem til HTML. Denne parser udtrak også nøgleord fra opslagene i IO og organiserede dataene i en SQL-database til webadgang på *Málið.is*. Resultatet blev en digital version af IO som kunne publiceres online, og en omfattende database hvor oplysningerne blev struktureret og gemt.

Den første webudgave af IO var dog ufyldstgørende. Den gav kun brugerne mulighed for at søge på opslagsord, man fik ikke overblik over struktur og indhold, og det var ikke muligt at se hvilke opslagsord der fandtes i databasen. Adgangen var begrænset til ét ord ad gangen, og man kunne ikke blade frem og tilbage som i den trykte bog. Derudover manglede alle metaoplysninger, fx kildefortegnelse og forkortelser.

Derudover opdagede man at nogle opslag manglede i webudgaven på grund af fejl i TeX-markeringen, men det var svært at afgøre hvilke ord der var tale om. På *Málið.is* fandtes i alt 24.054 opslag fra IO, men da den trykte bog ikke indeholdt oplysninger om det samlede antal opslagsord, var det svært at bedømme helt præcist hvor mange og hvilke ord der manglede.

Den originale TeX-parser der blev skrevet til den første digitale version på webportalen, blev nu brugt til at søge efter de manglende poster som var blevet sprunget over. For at finde de manglende poster søgte man mere eller mindre tilfældigt efter ord som man vidste var i den trykte udgave. Et af de ord der manglede, var ordet *rostungur* 'hvalros'. I dette tilfælde var ordets opmærkning fejlbehæftet, og det førte til at ordet ikke blev taget med af parseren. Der viste sig at være flere tilfælde hvor lignende fejl havde afstedkommet at ordet blev sprunget over. Denne proces med at identificere

fejl, rette dem i parseren og køre processen igen resulterede i i alt 24.471 opslag fra TeX-filerne, 417 flere end i den oprindeligt publicerede version på webportalen, og forhåbentlig alle opslag der forekommer i den trykte udgave af IO.

3.3 IOW – en ny webplatform for IO

Efter at denne begrænsede version af IO havde eksisteret som en del af webportalen i nogle år, begyndte man i 2022 at arbejde med at udvikle en mere avanceret digital udgave der kunne revitalisere IO ved at præsentere materialet på en ny måde. For at lave den skulle man udtrække en del oplysninger fra de oprindelige TeX-filer for at få en computer til at forstå hvordan ordenes indbyrdes relationer blev beskrevet i den trykte tekst og tolke det på en systematisk måde. Resultatet blev en ny digital webplatform, IOW, der blev publiceret online med adgang til hele IO's materiale.

For at illustrere processen med at udtrække oplysninger fra TeX-opmærkningen skal vi se nærmere på opslaget for ordet *heimur* 'verden', da det er en ret kompliceret artikel som er rig på både krydsreferencer og referencer til ord fra andre sprog (jf. figur 2).

Opslagsord	heimur k. 'jarðríki, veröld; alheimur; veraldargæði; mannfolk; þheimkynni, bústaður'; sbr. fæf. heimur
Andre sprog	'veröld', nno. heim 'heimkynni, sjónhringur, svæði, lofthvel, veröld', sæ. hem , d. hem 'heimkynni, heimili', fe. hām (nc. <i>homē</i>), fhþ. heim 'heimkynni, bústaður', gotn.
Ord på andre sprog	haims 'þorp'; < germ. * <i>haima-</i> , * <i>haimi-</i> ; sk. fsl. sémja
Betydning (andre sprog)	'hjú, þræll', lith. šeimýna , lett. sāime 'hjú'; upphafli. merk. '(fjölskyldu)bústaður, aðsetur'. Sbr. gr. κεῖται
Henvisning (andre opslag)	'ligger', koimāō 'háttá (e-n), svæfi', lat. cūnāe 'vagga, hreiður' (< * <i>koi-nā</i>), af je. * kei- 'liggja'
	Merkingarþróunin, 'heimastöðvar > veröld', á sér ýmsar hlöðstaður, sbr. rússn. mir 'þorpsfélag; friður, veröld', gotn. midjungards , fe. middangeard 'veröld' (eiginl. <i>miðgarður</i>). Heim- , -heimur kemur fyrir sem liður í mannanöfnum og örn. Sjá heim , heima (1), heimill , heimta , hið , hjón , hjú (1), hýbýli , hyr og hyski .

Figur 2: Opslaget *heimur* fra IO, hvor datatyper er blevet analyseret og markeret.

Figur 2 viser artikelstrukturen. Den indeholder først opslagsordet, efterfulgt af grammatiske oplysninger (dvs. *k* = *karlkyn* 'maskulin') og forklaring af ordets betydning inden for enkelte anførselstegn. Derefter følger en liste over lignende ord på andre relaterede sprog, der viser etymologiske forbindelser. Hver del indeholder en forkortelse af det relevante sprogs navn, ordet det refererer til på det andet sprog, og betydningen af dette ord på det pågældende sprog (fx nno [nynorsk]. *heim* 'heimkynni, sjónhringur, svæði, lofthvel, veröld')¹. Nederst i artiklen er der anført andre ord inden for samme opslag, dvs. forleddet *heim-* og efterleddet *-heimur*, som forekommer i person- og stednavne, efterfulgt af henvisninger til andre opslagsord i ordbogen (dvs. *heim*, *heima* (1), *heimill* osv.). Disse oplysninger åbner op for muligheden for at koble opslag sammen på en måde der ikke var mulig i den trykte bog. Hvis der i andre artikler forekommer henvisninger til de samme ord fra andre sprog, er der mulighed for at linke artiklerne sammen via disse ord.

For at løse disse opgaver og udtrække den relevante information blev der brugt regulære udtryk ("regular expressions") i Python. ChatGPT blev også brugt til at hjælpe med formuleringen af udtrykkene. Alle artikler blev behandlet på denne måde, og information om referencer blev gemt i specifikke tabeller i databasen. Hvis der opstod fejl, blev de rettet ved at justere det relevante regulære udtryk og lave en ny kørsel. Resultatet blev en stor tabel med referencer både til beslægtede sprog og til andre opslagsord i IOW.

Ved udviklingen af brugergrænsefladen til den nye platform blev disse referencer brugt for at gøre det muligt for brugeren at tilgå materialet på en innovativ måde med øget funktionalitet. Det blev gjort muligt at søge både efter opslagsord og i artikeltekst. Det betyder at når en bestemt artikel bliver vist, har man mulighed for at klikke videre til andre opslagsord der refereres til. Man har også links til ord fra andre sprog, jf. eksempel i figur 3.

1 Dvs. 'hjemegn, horizont, område, himmelhvælving, verden'.

heimur, Heim-, -heimur

Færsla Venslanet

heimur k. 'jarðríki, veröld; alheimur; veraldargæði; mannfólk; †heimkynni, bústaður'; sbr. [fær.](#) *heimur* 'veröld', [nno.](#) *heim* 'heimkynni, sjónhringur, svæði, lofthvel, veröld', [sæ.](#) *bem*, [d.](#) *hjem* 'heimkynni, heimili', [fe.](#) *bām* (ne. *home*), [fhp.](#) *heim* 'heimkynni, bústaður', [gotn.](#) *batims* 'þorp'; < [germ.](#) **haima*-, **haimi*-; sk. [fsl.](#) *sémtíja* 'hjú, þræll', [lith.](#) *šeimýna*, [lettn.](#) *sàime* 'hjú'; upphaf. merk. '(fjölskyldu)bústaður, aðsetur'. Sbr. [gr.](#) *keṓtai* 'liggur', *koimáō* 'háttá (e-n), svæfi', [lat.](#) *cūnae* 'vagga, hreiður' (< **koi-nā*), af [ie.](#) **kei*- 'liggja'. Merkingarþróunin, 'heimastöðvar > veröld', á sér ýmsar hliðstæður, sbr. [rússn.](#) *mír* 'þorpsfélag; friður, veröld', [gotn.](#) *midjungards*, [fe.](#) *midðangeard* 'veröld' (eiginl. *miðgarður*). **Heim-, -heimur** kemur fyrir sem liður í mannanöfnum og örn. Sjá [heim](#), [heima](#) (1), [heimill](#), [heimta](#), [bið](#), [bjón](#), [hjú](#) (1), [býbýli](#), [hýr](#) og [hyski](#).

eiginl. eiginlegur, eiginlega e-n einhvern
k. karlkyn merk. merking

Sjá

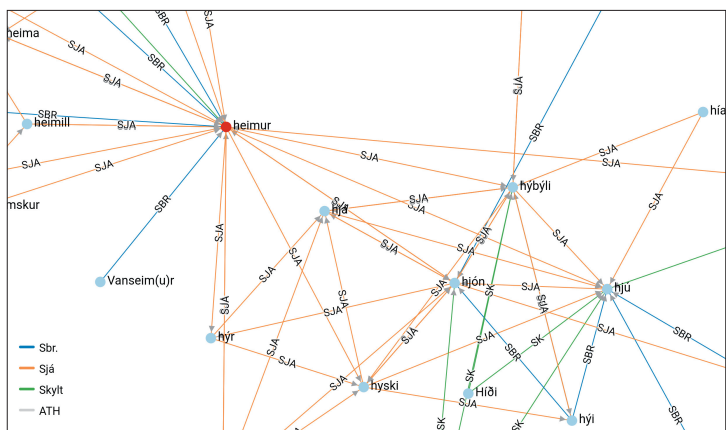
→ heim → heima (1)
→ heimill → heimta
→ hið → hjón
→ hjú (1) → hýbýli
→ hýr → hyski
← Freyland ← heima (1)
← heiman ← heimi
← heimill ← heimskur
← hið ← hjón

Skylt

← Heimir

Figur 3: Opslagsordet *heimur* i IOW.

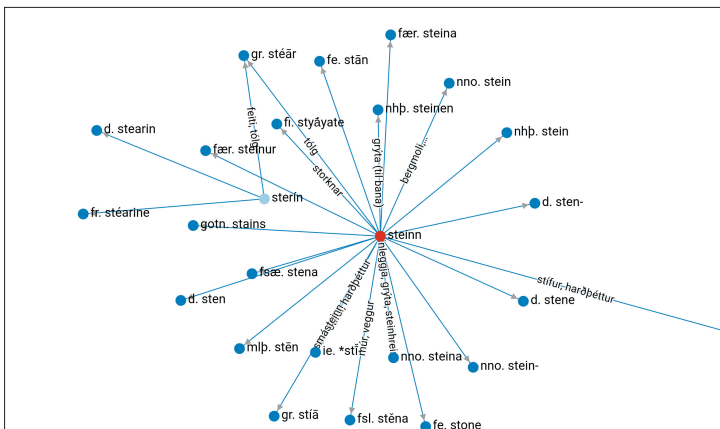
Figur 3 viser et typisk opslag i IOW. Ud over oplysningerne fra den trykte bog er der mulighed for at få en mere visuel oversigt over relationer mellem ord og sprog. Det gør man ved at generere netværksgrafer, dvs. grafiske diagrammer af ordenes indbyrdes koblinger. Det betyder at brugeren har mulighed for hurtigt at danne sig et overblik over hvordan et bestemt ord hænger sammen med andre ord. Brugeren kan se tilknytninger til andre artikler der refereres til i den pågældende artikel, såvel som artikler der refererer til den. Det visuelle diagram viser de andre relevante opslag i i alt tre "generationer" fra den oprindelige artikel. Et eksempel på et sådant diagram vises i figur 4.



Figur 4: En del af netværksgrafen der viser relationer til ordet *heimur* 'verden' (jf. IOW).

I figur 4 kan man se opslagsordet *heimur* og hvordan dette ord hænger sammen med andre ord. Man kan fx se en kobling til ordet *hjón* 'ægtepar' som ligeledes har en henvisning til *heimur*. Endvidere kan man se ordet *hirð* 'hof, hird' som der henvises til fra *hjón*, ligesom der henvises fra *hirð* til *hérað* 'herred'. Der skelnes mellem fire typer relationer, som har hver sin farve: *sjá* 'se', *skýlt* 'beslægtet', *sbr.* 'jf.' og *ATH* 'obs.'. Hele netværket for *heimur* er ganske stort, men diagrammets interaktivitet gør det nemt for brugeren at fokusere på ét ord ad gangen og dets umiddelbare slægtninge.

Man kan tilgå lignende diagrammer der viser forbindelserne mellem ord i andre beslægtede sprog, som det ville være helt umuligt at finde frem til i den trykte bog. I de tilfælde bruger man information om ord fra en bestemt artikel og kobler dem sammen med andre artikler hvor det samme ord diskuteres. Et eksempel vises i figur 5.



Figur 5: Netværksgraf der viser en række ord fra beslægtede sprog, der er relateret til opslagsordet *steinn* 'sten' (IOW).

I figur 5 kan man se hvordan ordet *steinn* 'sten' er knyttet til låneordet *sterín* 'stearin' igennem det græske ord *stéar*. Ordet *sterín* stammer fra det græske ord som har oprindelse i den samme indoeuropæiske rod som man finder i arveordet *steinn*. Den slags visualiseringer kan være gavnlige for at se hvordan ord bliver koblet sammen igennem de oplysninger der forekommer i artiklerne i IOW. På den måde kan man spore forbindelser som er svære at finde frem til i den trykte bog.

Det er dog ikke kun etymologiske forbindelser der diskuteres i IO, men også andre typer relationer mellem ord. For eksempel bliver ordet *heimur* også knyttet til personnavnet *Valdimar* igen gennem det russiske ord *mír*. Grunden til at ordet *mír* bliver nævnt i forbindelse med *heimur*, er dog ikke at de stammer fra samme rod, men derimod at det russiske ord udviser en semantisk udvidelse som er parallel til den man finder i ordet *heimur*, dvs. fra betydningen 'hjem' til 'verden', dvs. *mír* 'bygd, samfund, verden'. *Valdimar* (fra det russiske *Vladimir*) er et kompositum hvor andet led er det russiske ord *mír*, og dette er grunden til koblingen mellem *heimur* og personnavnet *Valdimar* selvom disse ord ikke er etymologisk relateret.

3.4 Yderligere oplysninger

En af begrænsningerne i den første online udgave af IO var at man kun kunne søge på enkelte opslagsord. Der manglede en del supplerende materiale som fandtes i den trykte udgave, fx registre og liste over forkortelser, liste over henvisninger og bibliografi. På IOW har man forsøgt at organisere de supplerende oplysninger og gøre dem tilgængelige på en overskuelig måde. På figur 6 kan man se hvordan oplysningerne er organiseret, og hvordan undersiden *Skammstafanir* og *flokkar* 'forkortelser og grupperinger' er struktureret under fire faner.

Tungumálaheiti	Skammstafanir	Rit	Nafnaflokkar
Tungumálaheiti og lýsingarorð af þeim dregin			
afinn.	austur-finnska		
afrank.	austur-frankíska		
afrísn.	austur-frísneska		
agerm.	austur-germanska		
agotn.	austur-gotneska		
akkad.	akkadíska		

Figur 6: Fanen *Tungumálaheiti* på undersiden *Skammstafanir* og *flokkar* på IOW.

På denne underside findes fire faner: *Tungumálaheiti* 'sprogtermer', *Skammstafanir* 'forkortelser', *Rit* 'værk der henvises til' og *Nafnaflokkar* 'navnekategorier'. Fanen *Tungumálaheiti* vises i figur 6. Den indeholder forkortelser over sprog og afledte adjektiver, fx *gríska* 'græsk' (sprog) og *grísk-rómversk* 'græsk-romersk' (adj.).

Under fanen *Skammstafanir* findes alle andre forkortelser der forekommer i IO. Det er bl.a. symboler der har særlig betydning,

fx ☉ der betyder at ordet kun forekommer i et begrænset geografisk område i talesprog. Fanen *Rit* indeholder en liste over citerede værker, ofte islandske middelalderværker, men også kendte skriftlige kilder fra andre lande. Nogle gange refereres der til bestemte udgaver af det pågældende værk eller visse forfattere (jf. figur 7).

Bhagavad-gita	Fornindverskt heilræðaljóð
Bibl.	Biblían
Bisk.	Biskupa sögur 1–2. Khöfn 1858, 1878
Bjólfskviða	Fornenskt kvæði frá 8. öld
Bólu-Hjálmar	Hjálmar Jónsson frá Bólu. Ritsafn II, V. Rvík 1949
Bósa s.	Bósa saga (Fornaldars.)

Figur 7: Uddrag af listen over henvisninger.

I figur 7 kan man se et tilfældigt uddrag af referencelisten, der viser forskellige kildetyper: *Bhagavadgita*: et oldindisk digt, *Bibelen*, *Biskupasögur*: islandske fortællinger om biskopper fra middelalderen, *Beowulf*: et oldengelsk digt, *Bólu-Hjálmar* (1796-1875): en islandsk digter og *Bósa saga*: en oldtidssaga fra det 14. århundrede. Til venstre ses forkortelsen som den forekommer i IOW, og til højre den fulde titel. Man kan trykke på hver enkelt reference og se de pågældende artikler hvori referencen bliver nævnt.

Den sidste fane *Nafnaflokkar* har ikke et tilsvarende afsnit i den trykte udgave. Her har man tilføjet en mulighed for at slå forskellige navnetyper op der forekommer i IO. Navnetyper bemærkes nogle gange i artikelteksten (fx *nafn á x* eller *heiti á x* 'navn på x'), og her har man samlet de 15 hyppigst forekommende grupper, bl.a. *árheiti* 'flodnavne', *dvergsheiti* 'dværgenavne', *eyjarheiti* 'ønavne', *goðsöguleg heiti* 'mytologiske navne', *hestsheiti* 'hestenavne'. IOW gør det muligt for brugeren at tilgå disse oplysninger samlet på ét sted og klikke videre til de relevante opslag.

3.5 Udfordringer

Ved gennemgang af materialet har vi fundet frem til nogle udfordringer som kræver en særlig behandling. Det drejer sig både om problematiske forhold i det oprindelige materiale og tekniske problemer.

Nogle af forkortelserne kan være tvetydige. For eksempel er forkortelsen ”no.” både brugt for *nafnorð* ’navneord, substantiv’ og *norska* ’norsk’. Det samme gælder forkortelsen ”sk.” *skylt* ’relateret’ som også bliver brugt som forkortelse for *skoska* ’skotsk’. Tvetydige forkortelser som disse skulle behandles på en særlig måde hvor man forsøgte at definere i hvilken sammenhæng de forekommer, og lave regler der kunne skelne mellem dem. Det virker i de fleste tilfælde. Dog er det muligt at unøjagtigheder stadig kan forekomme, især islandske substantiver der bliver tolket som norske. Sådanne fejl bliver rettet når de opdages.

Som tidligere nævnt er IO et forfatterværk og under ophavsret. Det betyder at man ikke kan ændre IO’s indhold selvom man støder på fejl og mangler. Ved gennemgangen af den trykte udgave er det klart at oplysninger vedrørende ordenes alder i sproget eller tidspunkt for første belæg i skriftlige kilder i mange tilfælde ikke er korrekte. Disse oplysninger er ikke knyttet til alle opslagsord, men derimod næsten udelukkende til de ord der ikke forekommer i middelalderkilder, se opslaget for adjektivet *fallegur* ’smuk, køn’ i figur 8.

fallegur l. (16. öld) ‘fagur, snotur, laglegur’; sbr. [fær.](#) *falligur* ‘fagur’, [nno.](#) *fallig* ‘liðugur’, [jó.](#) *falle* ao. ‘all-, mjög’; e.t.v. af so. *falla*, sbr. [nhp.](#) *gefällig* ‘geðþekkur, snotur’, [nno.](#) *felleleg* ‘hæfandi’. Aðrir ætla að orðið sé < *far-ligr* (í [fisl.](#) skáldam.) ‘laglegur, lögulegur’ (m.a. um skip). Vafasamt.

Figur 8: Opslaget *fallegur* ’smuk, køn’ fra IOW.

Ifølge oplysningerne i IO er ordet *fallegur* først belagt i det 16. århundrede. En nærmere undersøgelse viser dog at ordet findes allerede i skriftlige kilder fra 14. og 15. århundrede, som det fremgår af opslag i *Ordbog over det norrøne prosasprog* (ONP), en online digital ressource over middelaldersproget i Island og Norge. Der er mange flere lignende tilfælde hvor der findes ældre belæg for opslagsord end det fremgår af IO.

Dette tilfælde og andre lignende understreger den problematik der kan opstå når videnskabelige værker, såsom ordbøger skrevet af en afdød forfatter, ikke kan opdateres på grund af deres statiske tilstand. Når den oprindelige forfatter ikke længere er i live, overgår ophavsretten normalt til dennes efterkommere. Hvis arvingerne ikke er interesseret i at vedligeholde eller videreudvikle det videnskabelige værk, skaber det problemer med at holde materialet relevant og opdateret. Denne situation begrænser den akademiske udvikling, da nye indsigter eller data ikke kan inkorporeres i et værk der stadig er beskyttet af ophavsretten, på trods af dets potentiale for videre forskningsmæssig relevans. De eneste muligheder kan være at lave en slags kommentar til det oprindelige værk, lave et nyt supplerende værk eller starte forfra, hvilket kan være både tids- og ressourcekrævende, samtidig med at det kan være en udfordring at opnå den samme autoritet som det oprindelige værk har.

3.6 Fremtidsperspektiver

Som vi har set, er det muligt at visualisere mange af relationerne mellem ord og sprog der er præsenteret i den trykte IO. Det er dog ikke altid klart af hvilken art relationerne er. En af udfordringerne i fremtiden vil blive at forbedre disse oplysninger og gøre det klarere hvilken relation der eksisterer mellem ordene, fx direkte kognater eller lignende semantisk udvikling.

Selvom ordbogens indhold ikke kan opdateres eller omstruk-

tureres, er der dog mulighed for at udarbejde supplerende materiale, måske i form af noter eller kommentarer til de relevante artikler hvor man enten er stødt på deciderede fejl eller gerne vil tilføje supplerende oplysninger. Det kunne være en god midlertidig løsning og mindre krævende end en gennemgribende revidering.

En mulig forbedring kunne være linkning til andre etymologiske ordbøger eller etymologiske oplysninger fra andre sprog. På den måde kunne brugeren få endnu mere omfattende diagrammer over ordenes relationer og muligvis finde frem til nyttige etymologiske oplysninger der ikke er præsenteret i IOW.

En anden måde at forbedre ordbogens brugbarhed på kunne være at indføre semantisk opmærkning hvor opslagsordene blev markeret med semantiske etiketter eller tildelt semantiske kategorier. Mange islandske ordbogsressourcer har allerede integreret oplysninger om betydningsområder i deres databaser (jf. Jóhannsson & Úlfarsdóttir 2024). Det ville være forholdsvis nemt at overføre disse semantiske informationer til hovedparten af opslagsordene i IOW.

Når vi tænker længere frem, ville det være ønskeligt om man kunne lave en aftale med ophavsrethaverne og få lov til at revidere IOW og opdatere den med både tilføjelse af nye opslagsord og supplerende informationer til de eksisterende artikler. Der er også behov for redaktionel revidering, fordi der er en del ord i IO som hører til sjældent eller specialiseret sprogbrug i moderne islandsk og enten skulle udgå eller have en særlig etikette. Drastiske ændringer er dog ikke realistiske i den nærmeste fremtid.

4. Afsluttende bemærkninger

Íslensk orðsifjabók indeholder mange nyttige informationer som det er værd at gøre tilgængelige på en digital platform. En statisk ordbog kan revitaliseres ved hjælp af forskellige søgefunktioner

og en ny bearbejdelse af materialet. Revitaliseringsprocessen og udviklingen af IOW har markant forbedret tilgængeligheden og brugervenligheden af IO, blandt andet gennem forbedrede søgemuligheder, visuelle diagrammer der viser ords relationer, og hyperlinks som gør det muligt hurtigt at navigere mellem opslag.

Der tilbagestår dog endnu nogle udfordringer i forbindelse med dataene og visningen af dem på IOW. Der kan forekomme fejlagtige henvisninger, og yderligere fejl skal også rettes. Regler om ophavsret begrænser udviklingen, men yderligere forbedringer er dog mulige selvom det oprindelige materiale ikke bliver ændret.

Litteratur

Ordbøger, korpusser og digitale resurser

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Árni Magnússon-instituttet for islandske studier. <<https://bin.arnastofnun.is>> (marts 2025).

ÍOB = *Íðorðabankinn*. Ágústa Þorbergsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <<https://idordabanki.arnastofnun.is/>> (marts 2025).

IO = *Íslensk orðsifjabók*.

IOW = *Íslensk orðsifjabók*. [En digital webplatform] <<https://ord-sifjabok.arnastofnun.is>> (marts 2025).

Magnússon, Ásgeir Blöndal (1989): *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.

málið.is = Webportalen *málið.is*. Árni Magnússon-instituttet for islandske studier. <<https://malid.is>> (marts 2025).

OMI = *Íslensk nútímamálsorðabók* [Ordbog over moderne islandsk]. Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (red.). Árni Magnússon-instituttet for islandske studier. <<https://islenskordabok.is>> (marts 2025).

- ONP = *Ordbog over det norrøne prosasprog / A Dictionary of Old Norse Prose Online*. Københavns Universitet. <<https://onp.ku.dk>> (april 2025).
- SOB = *Íslensk stafsetningarorðabók*. Jóhannes B. Sigtryggsson (ritstj.). Árni Magnússon-instituttet for íslandske studier. <<https://stafsetning.arnastofnun.is>> (marts 2025).

Anden litteratur

- Haugen, Einar (1976): *The Scandinavian Languages: An Introduction to their History*. London: Faber and Faber.
- Harbert, Wayne (2006): *The Germanic Languages*. Cambridge: Cambridge University Press.
- Johannsson, Ellert Thor & Thórdís Úlfarsdóttir (2024): The role of semantic fields in contemporary Icelandic dictionaries. I: Kristina Š. Despot, Ana Ostroški Anić & Ivana Brač (red.): *Proceedings of the XXI EURALEX International Congress 8–12 October 2024*. Zagreb: Institut za hrvatski jezik. Cop. 75–85.
- Kvaran, Guðrún (2005): *Íslensk tunga II. Orð. Handbók um beygingar- og orðmyndunarfræði*. Reykjavík: Almenna bókafélagið.
- Malkiel, Yakov (1976): *Etymological Dictionaries. A Tentative Typology*. Chicago & London: The University of Chicago Press.
- Olander, Thomas (red.) (2022): *The Indo-European Language Family. A Phylogenetic Perspective*. Cambridge: Cambridge University Press. <<https://doi.org/10.1017/9781108758666>>.
- Steingrímsson, Steinþór, Ari Páll Kristinsson & Halldóra Jónsdóttir (2018): Málið.is: A Web Portal for Information on the Icelandic Language. I: Ásta Svavarsdóttir, Halldóra Jónsdóttir, Helga Hilmiðsdóttir & Þórdís Úlfarsdóttir (red.): *Nordiske Studier i Leksikografi* 14. Reykjavík: Nordisk Forening for Leksikografi. 255–261.

Svensén, Bo (1987): *Handbok i Lexikografi. Principer och metoder i ordboksarbete*. Stockholm: Esselte Studium & Tekniska nomenklaturcentralen.

Svensén, Bo (2004): *Handbok i Lexikografi. Ordböcker och ordboksarbete i teori og praktik*. Stockholm: Nordstedts Akademiska Förlag.

Trausti Dagsson
projektleder og udvikler
Árni Magnússon-instituttet for
íslandske studier
Arngrímshgata 5
IS-107 Reykjavík
trausti.dagsson@arnastofnun.is

Ellert Þór Jóhannsson
forskningslektor, ph.d.
Árni Magnússon-instituttet for
íslandske studier
Arngrímshgata 5
IS-107 Reykjavík
ellert.thor.johannsson@arnastofnun.is

Litteraturen i den diakrone dokumentasjonsordboken – en beskrivelse med utgangspunkt i *Det Norske Akademis ordbok* (NAOB)

Hanne Lauvstad

Authentic quotations play an important role in a historical dictionary documenting the development of language in a certain period, based on literature of different types. This article concerns the experience from working with the *Norwegian Academy Dictionary* (NAOB). This type of dictionary offers specific challenges concerning the selection and treatment of the source material. The lexicographic functions of the authentic quotations are discussed, e.g. whether the quotations demonstrate typical use of words and collocations, whether they reflect semantic differences and changes, and whether they exemplify morphological and orthographical variations. Additional lexicographical functions of the authentic quotations are presented. These may serve as windows to the literary heritage, cultural changes or various authors' use of words and collocations. Assertions are made that the sources should represent a balanced selection of authors, the oeuvres of certain authors, genres and literary epochs, with an even distribution between male and female writers, and between fiction and non-fiction. The relationship between NAOB and its literary sources is described and discussed. The methods used for the compilation of relevant literary examples are presented. The importance of textual criticism is emphasized.

1. NAOB som ordboktype og litteraturens rolle

Sitateksempler spiller en sentral rolle i en dokumentasjonsordbok. En diakron og litterær dokumentasjonsordbok står i et interessant symbiotisk forhold til den litteraturen den bygger på, i og med at de dokumenterende sitateksemplene viser variasjonen i språk-

bruk og at ordbokens lemmatisering, grammatiske opplysninger, bruksmarkører og definisjoner forklarer og belyser sitatmaterialet. Arbeidet med å oppdatere og videreutvikle ordboken med sitateksempler står derfor sentralt i en slik ordbok. Det samme gjelder tilgjengeliggjøring av både eksisterende og nytt litterært stoff i ordboken. Det er derfor nyttig å se nærmere på forholdet mellom *Det Norske Akademis ordbok* (NAOB) og dens litterære kildemateriale. NAOB dekker perioden fra ca. 1814 til i dag. Ordboken bygger på litterære kilder. Betegnelsen *litteratur* må her forstås i sin videste betydning 'publiserte tekster'; sitatene i NAOB er hentet både fra skjønnlitteratur og sakprosa, inkludert avistekster.

Sitateksemplene har flere forskjellige funksjoner i NAOB. For det første bidrar litterære sitater til å vise frem og forklare ord og uttrykk, for det andre gir de autentiske eksempler på bruk av ord og uttrykk, og for det tredje blir de en slags vinduer inn til litterære verk, idet de gir smakebiter på forskjellige forfatters ordforråd og stilistiske særegenheter.

NAOB bygger på – og er en kraftig utvidelse og modernisering av – *Norsk riksmålsordbok* (NRO), som utkom i fire bind i perioden 1937–57, med to tilleggsbind i 1995. Dette har betydning for sitatmaterialet på flere plan. NRO bygde på et seddelarkiv med ekserpter, hovedsakelig fra norsk skjønnlitteratur, men også fra aviser og noe fra tidsskrifter og annen sakprosa. Skjønnlitteraturen var dominert av forfatterskap som tilhørte den litterære kanon. NAOBs materiale er utvidet til å omfatte et bredere utvalg av skjønnlitteratur og ikke minst sakprosa. Hovedsakelig foreligger dette materialet som flere digitale tekstkorpus. Innsamlingen av ordtilfang og sitateksempler foregår først og fremst ved digital søking og sanking, men suppleres med manuell ekserpering. Det litterære grunnlagsmaterialet for hhv. NRO og NAOB, litteratursynet i de to ordbokverkene og metoder for innhenting av sitateksempler vil bli nærmere omtalt nedenfor.

2. Sitateksemplenes leksikografiske funksjon

Sitateksemplene i NAOB har forskjellige funksjoner.

2.1. Hovedfunksjon 1: Språklig eksempel materiale

Som Svensén (2004) påpeker, bør sitateksemplene vise ord og uttrykk i typisk bruk, de bør reflektere forskjeller i betydninger og betydningsutvikling, og de bør vise variasjoner i det språklige formverket.

Mens NRO var en utpreget *resepsjonsordbok*, dvs. en ordbok som man kan slå opp i for å få forklaring og eksempler på bruk av ord eller uttrykk som man ikke kjenner til eller forstår, er NAOB i større grad multifunksjonell, en *flerbruksordbok*. Både resepsjons- og produksjonsaspektet påvirker sitateksemplenes språklige funksjon.

2.1.1. NAOB som resepsjonsordbok

Resepsjonsordbokfunksjonen er selvsagt fortsatt til stede; NAOB vil være nyttig å slå opp i for alle som leser litteratur, eller som på annen måte kommer i kontakt med det som for dem er nye deler av ordforrådet. En resepsjonsordboks ordforklaringer minner om dem man finner i verbalkommentarene i tekstkritiske utgaver av litterære verk, hvor de står som noter til teksten. Forskjellen er at ordforklaringene i slike utgaver er knyttet til én spesiell litterær kontekst, mens definisjonene i en ordbok er mer generelle og dekker bruk i flere typer sammenhenger, bl.a. både i skjønnlitteratur og sakprosa. Et eksempel på et fremmedord som ordbokbrukere kan komme til å slå opp på for forklaring, er *invektiv* 'skjellsord'. Ordet blir illustrert med følgende sitateksempler i NAOB:

invektivene blir ikke engang forsøkt begrunnet (Hans Heiberg *Så stort et hjerte* 59 1972)

Undset sparer ikke på invective i sin omtale av Luther
(Liv Blikrud *Sigrid Undset* 118 1997)

hun ble frosset ut av bedriften, mobbet og baktalt, for til
slutt å bli sparket ut i mørket med et tonefølge av grove in-
vectiver (Christian Borch *Sannhetens kår* 218 2009)

Betydningsutvikling blir også illustrert ved hjelp av sitateksempler i NAOB, f.eks. har substantivet *kjole* endret betydning gjennom tidene. NAOBs betydning 1 'knelangt ytterplagg for menn, med oppslag på ermene, knapper foran og vide skjøter' er illustrert med bl.a. følgende sitat fra en novelle av Maurits Hansen fra 1829: «[magister Blom] bar en staselig paryk, en broderet silkevest med taskelommer og en rød stoffes kjole derover» (M.C. Hansen *Noveller* 57–58), et sitat som tydeliggjør den historiske forankringen av denne betydningen. En av underbetydningene (1.4.), 'snippkjole', er også kjent i uttrykket *kjole og hvitt*. Det vises bl.a. i et sitateksempel som «damer i silke og herrer i kjole og hvitt» (*Tidens Tegn* 1935/93/6/5), en anskueliggjøring av at det er tale om et festplagg. Betydning 2 av *kjole*, som er den nå alminnelige, 'kvinnedrakt (med eller uten ermer), med sammenhørende (nå som regel sammensydd) liv og skjørt', er illustrert bl.a. med sitateksempelen «der kom [småpikene] over plassen, alle sammen, noen med store hvite sløyfer i håret og blomstrete kjole og korte puffermer» (Gerd Brantenberg *Sangen om St. Croix* 10 1979).

Selv om NAOB-materialet dekker en lang periode, formidles materialet ut fra en moderne synsvinkel. Dette gir seg utslag i utformingen av definisjoner og bruksmarkører, som er knyttet til enkeltbetydninger eller direkte til sitatene når disse avviker stilistisk, grammatisk eller semantisk fra den betydningen de er knyttet til. Sitateksempler med ordformer som ikke tilhører den moderne bokmålsnormen, blir likeledes markert med bruksmarkører, bl.a. stilmarkører. Også hele lemmaer eller underbetydninger av lem-

maer markeres med bruksmarkører, f.eks. stilmarkørene «foreldet» og «dialektalt». Noen av bruksmarkørene er de samme som i NRO, men systemet er revidert. Dette er et ledd i tilgjengeliggjøringen av sitatmaterialet i ordboken. Et eksempel finner man under lemmaet *ekorn*, hvor vi finner to sitater med markering.

Et sitateksempel har stilmarkøren «foreldet» direkte foran sitatet pga. den danske formen *egern*:

egernet pudser sin snude (Henrik Wergeland *Samlede Skrifter I*, 3 68)

Et annet sitateksempel har stilmarkøren «dialektalt» pga. formen *ikorn*:

der [var] kommet et lidet ikorn med op imellem nødderne,
og det sad på hug i fanget og spiste dem op allesammen
(*Bjørnstjerne Bjørnsons fortællinger* 44)

Også grammatiske avvik, f.eks. uvanlig genus, kan markeres ved sitater. Et eksempel er et sitat med nøytrumsform av ordet *benrad* 'skjelett', som vanligvis er maskulinum, og som har fått markøren «i nøytrum» før sitatet:

de fingre, som Weydahl greb om glasset med, var knoklede
og tørre som et benrads (Tryggve Andersen *Samlede fortællinger I* 210)

Sitater med en spesiell bibetydning får en utfyllende forklaring. Et eksempel finnes under uttrykket *høre etter*, i betydningen '(oppreksomt) lytte etter eller til (noe, noen)'

[han fant] at det ofte lønnet sig at høre efter hende (Hans Aanrud *Fortællinger III* 24 1923) | følge hennes råd

Eldre rettskrivningsformer (f.eks. preteritumsformen *skulde* for *skulle*) forekommer i store mengder i NAOB og markeres ikke dersom de representerer en regulær ortografisk utvikling. Ellers ville mesteparten av NAOBs eldre sitatmateriale måtte markeres, noe som både ville føre til at viktigere markører ville drukne i mengden, og som i tillegg ville være uhyre arbeidskrevende.

Også ord som kunne kvalifisere som *hapax legomena*, og som altså viser atypisk språkbruk, har fått stå i NAOB, i og med at ordboken er en resepsjonsordbok. Det er her snakk om sjeldne ord, brukt av kjente forfattere – ord som ordbokbrukerne kunne tenkes å slå opp på, fordi de rett og slett ikke forstår dem når de støter på dem. Ordet *hybensparlak* er brukt metaforisk av Henrik Wergeland i betydningen 'sengeomheng av nyperoser'. Men i og med at både *hyben* (dansk for 'nype') og det arkaiserende *sparlaken* (med betydningen 'sengeomheng') sannsynligvis er uforståelige for mange av våre ordbokbrukere, har vi valgt å ta med ordet og forklare det med lenker til begge leddene. Formen *sparlak* avviker dessuten fra NAOBs oppslagsform *sparlaken*.

NAOB har lenker fra de aller fleste sitateksemplene til det norske Nasjonalbibliotekets digitaliserte utgaver i Nettbiblioteket (unntakene er litteratur som av rettighetsgrunner ikke er allment tilgjengelig). Dette bør bidra til tilgjengeliggjøringen av sitatmaterialet, ikke minst fordi ordbokbrukerne ved å følge lenkene kan se sitateksemplene i kontekst.

2.1.2. NAOB som produksjonsordbok

Et annet aspekt ved sitateksemplenes språklige funksjon er at de hjelper den som skal uttrykke seg selv skriftlig eller muntlig. NAOB kan dermed også brukes som *produksjonsordbok*. Sitatene utfyller ordbokartiklene med autentiske eksempler på språket i bruk. I denne sammenheng er det viktig at sitatene viser en typisk språkbruk, som kan tjene som mønster, og at man luker bort *hapax legomena* og andre språklige rariteter. Samtidig bør sitatene

vise variasjon, f.eks. i ortografi (som ofte gjengir uttale) og i bøyning. Også en viss veksling mellom hjemlige former, utviklet av arveord, og danske former i bokmålet kan synliggjøres i enkelte ordbokartikler i NAOB slik det gjøres under lemmaet *kullsvart*. Sitatene viser, i tillegg til oppslagordet, former som «kulsort», «kullsort», «kulsvart», «kølsvart», «kølasvart». Sitatene kan altså underbygge og supplere definisjonene, ved at de viser ordene i autentisk bruk med variasjon i stavemåtene. I ordbokartikkelen *abscess* i NAOB viser sitatene semantiske aspekter ved lemmaet; hvordan en abscess kan oppfattes (fæl), hvor den kan vise seg (ved hælen), hvordan den kan føles (hard som en kastanje), og hva den kan inneholde (puss, materie, som kan sprute). Slik kan sitateksemplene være et nyttig supplement til ordbokdefinisjonene. Også en rekke kollokasjoner blir eksemplifisert med litterære sitater. F.eks. i artikkelen *god* finner vi uttrykket *ha det så godt* ironisk brukt med betydning 'få som fortjent', illustrert med både et eldre og et nyere sitateksempel:

så kan de ha det så godt (Henrik Ibsen *En folkefiende* 186
1882)

hadde hun valgt å bli bondekjerring, så fikk hun vel ha det
så godt! (Per Qvale *Tante Amerika* 55 2016)

I alle de nevnte eksemplene bør sitateksemplene kunne være til hjelp for den som selv skal skrive, ved at de viser en bredde i form, bruk og betydning.

2.2. Hovedfunksjon 2: Sitater som vinduer til litteratur og kultur

Sitatene blir dessuten en tekst i teksten; en slags vinduer inn til litteraturen og til den kulturelle og tekniske utviklingen. Denne

multifunksjonaliteten finnes i alle lignende historiske dokumentasjonsordbøker og er bl.a. omtalt i forbindelse med *Ordbog over det norrøne prosasprog* (se Wills 2022:184–185).

Et eksempel er substantivet *telefon* (her i NAOBs betydning 1, 'telefonapparat'). Et utvalg av sitateksemplene der viser eldre og nyere bruk og betydning av ordet. Disse reflekterer utviklingen av den gjenstanden som ordet refererer til:

jeg har pr. telefon bestilt lunch på Bådshaug (Bjørnstjerne Bjørnson *Samlede digter-verker* VI 379)

[telefonapparatet], der tillader telefonisk korrespondance ... indeholder: en telefon, en mikrofon, to kontaktskykker, en elektrisk klokke, en umschalter for to retninger og endelig batterielementer (*Norske Intelligenssedler* 16.08.1889/4/2)

«Den hvite telefon»s epoke, med sine oppkonstruerte teatraliske intriger i elegant luksusmiljø (Oddvar Foss *Antonio* 15 1972) | om filmsjanger i mellomkrigstiden hvor hvite telefoner var et statussymbol

om kveldene satt hun lenge i telefonen (Marie Aubert *Kan jeg bli med deg hjem* 47 2016)

jeg tar noen bilder med telefonen (Simon Stranger *Leksikon om lys og mørke* 191 2018)

Det siste sitatet viser ordet brukt i betydningen 'smarttelefon', dvs. en mobiltelefon som også har datamaskinfunksjoner. Det er imidlertid vanskelig å skille ut denne som en egen underbetydning, da det i mange tilfeller er umulig å fastslå ut fra konteksten hva slags telefon det er tale om, i andre tilfeller er det mindre relevant.

Dette er et viktig poeng for leksikografen når nye sitateksempler skal innpasses i ordboken. Løsningen i dette tilfellet har vært å endre definisjonen av ordet telefon, slik at den dekker både eldre og nyere typer telefoner. Dette er et godt eksempel på hvordan moderniseringer i en historisk dokumentasjonsordbok ikke bare innebærer suppleringer av sitatmaterialet, men at slike fornyelser også kan gjøre det nødvendig å endre selve definisjonene.

3. Utvalg av kildemateriale

I nyere tid er det kommet en økt bevissthet om hvor viktig representativitet er i utvalget av kildematerialet for en ordbok. Både forskjellige forfatterskap, sjangre og litterære perioder bør være representert i en diakron ordbok. Man bør også ha en balanse mellom mannlige og kvinnelige forfattere og mellom skjønnlitteratur og sakprosa, inkludert tekster fra aviser og tidsskrifter (jf. Bergenholtz 1996).

Det er nyttig å få med sitater med ord og uttrykk fra spesielle fagspråklige domener i ordboken. Et eksempel er oppslagsordet *primat*, som i den zoologiske betydningen 'dyr i ordenen primater' har følgende sitateksempler i NAOB:

disse primatene utvikler seg til menneskeaper (Tor Ulven
Avløsning 55 1993)

det finnes ikke noe avgjørende sprang mellom menneskets
sjeleliv og sjelelivet hos primatene, mener [Darwin] (Trond
Berg Eriksen *Undringens labyrinter* 476 1994)

det var som å bli angrepet av en ukjent jungelskapning, en
primitiv og livsfarlig primat (Bernt Roughthvedt *Tårn* 139
2013)

fysisk er vi [mennesker] definitivt natur, en primat blant andre primater (Dag O. Hessen *Natur* 20 2008)

Betydningen er med andre ord belyst med sitateksempler både fra en idéhistoriker (Trond Berg Eriksen) og en biolog (Dag O. Hessen), i tillegg til skjønnlitterære eksempler (fra Tor Ulven og Bernt Roughtvedt). Med sitateksempler både fra relevant sakprosa og fra skjønnlitteratur får man vist hvordan oppslagsordet anvendes i forskjellig type kontekst.

Selv om vi ønsker at en rekke sjangre skal være belagt i NAOBs kildemateriale, er det forskjell på hvor egnede de forskjellige sjangrene er. Jeg nevnte tidligere det spesielle ordet *hybensparlak*, som er med i NAOB av resepsjonsgrunner, men generelt er vi selektive i bruk av sitater fra lyrikk i ordboken. I lyrikk har språket også en poetisk funksjon, i den forstand at språket peker på seg selv og bl.a. tematiserer sin lydlige side (f.eks. med rim). Dermed kan det lyriske språket i en del tilfeller passe dårligere til å belyse typisk språkbruk. Når vi velger ut sitateksempler til NAOB, tar vi hensyn til dette og siterer bare dikt dersom språkbruken er nonlunde typisk. Fortellende lyrikk (bl.a. viselyrikk) er den typen lyrikk som best er egnet som kilde for sitateksempler, mens lyrikk med kreativ bruk av språklige og lydlige virkemidler kan være mindre egnet som språklig eksempelmateriale i ordboken, siden språket der er tydeligere preget av den poetiske funksjonen.

Det diakrone aspektet står sentralt i NAOB. I likhet med NRO er ikke NAOB en strengt diakron ordbok, slik som f.eks. SAOB, i den forstand at ordboken ikke har en systematisk registrering av når et ord første gang opptrer i kildematerialet. Vi arbeider allikevel kontinuerlig for å få med eldste mulige belegg på ord og med en tidsmessig spredning av sitatene.

En viktig del av arbeidet med NAOB er det å få inn sitater fra den aller nyeste litteraturen. Vi har derfor som løpende prosjekt å supplere kildematerialet med ny skjønnlitteratur og sakprosa. Vi

velger da bl.a. bøker av kjente forfattere, bøker som er blitt tildelt eller er blitt nominert til større litterære priser, og bøker som får stor utbredelse. Når det gjelder sakprosa, er vi opptatt av å dekke forskjellige domener, ikke minst domener som har vært dårlig representert fra før i ordboken (som kunst, design og arkitektur, matlagning), og domener hvor det har vært en stor utvikling av terminologien (som medisin). Dette arbeidet er til dels organisert som delprosjekter som er finansiert av private fond og stiftelser. NAOB-redaksjonen er vel å merke opptatt av å dekke fagtermer som opptrer i allmennspråklig kontekst, f.eks. i avisartikler og skjønnlitteratur. Det rent terminologiske arbeidet overlater vi til andre.

Ikke bare ny litteratur har det vært viktig å få dekket i NAOB. Også litteraturen fra 1930- til 1970-årene var nokså sparsomt representert i NRO. Vi har derfor drevet med målrettet ekserpering av eldre og nyere litteratur for å fylle lakunene. Ikke minst har det vært viktig å øke andelen sitater fra tekster av kvinnelige forfattere. Vi har i tillegg vært opptatt av å sitere fra barne- og ungdomslitteratur og sakprosa av forskjellig slag for denne perioden. Også kriminallitteratur av noen kjente forfattere er innlemmet i NAOBs korpus og har gitt sitatbelegg.

Et prosjekt NAOB-leksikografene nylig har startet på, er å innlemme viselyrikk mer målrettet i NAOB. Viselyrikken er som nevnt en type lyrikk som er egnet som kildemateriale for ordboken, i og med at den ofte er holdt i en fortellende stil. Den har i noen grad vært representert med sitater fra kjente forfattere som Alf Prøysen, men en rekke andre viseforfatteres tekster (f.eks. tekster av Lillebjørn Nilsen, Anne Grete Preus, Ole Paus og Odd Børretzen) manglet. Disse kildene gir nyttig tilfang av belegg, ikke minst på dagligspråket, men de kan også inneholde interessante eksempler på poetisk språkbruk, bl.a. bibelallusjoner. Fra dagligspråket kjenner vi uttrykket *bli tiden ut* 'bli hele den planlagte perioden'. Det uttrykket har vi innlemmet i NAOB, belagt med følgende sitat:

opptil åtte uker [på feriekoloni] var vårt tilbud fra Kommunen. Og selvsagt måtte vi bli tida ut (Lillebjørn Nilsen Tekst og musikk: *Lillebjørn Nilsen* (2004) 97) | fra visen «Stilleste gutt på sovesal 1»

En tydelig bibelallusjon finner vi i dette sitateksempel:

Fader vår, komme ditt rike. Fader vår, vær her i kveld. Fader vår, la ondskapen vike (Ole Paus *Pauspoeten* 377 2020) | i visen «Pyro»

I NAOB tar vi bare inn tekster fra trykte utgaver. Rene nettekster og materiale fra sosiale medier, blogger, chatter o.l. benyttes ikke, da disse oppfattes som mer flyktige og uformelle, i en slags gråsoner mellom skriftspråk og talespråk.

Det er viktig å understreke at NAOB, på samme måte som NRO, ikke har med sitateksempler fra oversatt litteratur, med noen få unntak, nemlig bibeloversettelser og oversettelser av norrøn litteratur. Årsaken til utelatelsen av oversatt litteratur er dels at det kan oppstå interferens fra opphavsspråket, dels at en utvidelse i den retningen av korpusmaterialet ville bli vanskelig å avgrense.

Som leksikograf ønsker man å bruke litterære sitater både som eksempler på typisk språkbruk og som eksempler på litteratur som blir høyt vurdert av kritikere og lesere. Man ønsker videre å sitere litteratur som når vidt ut. Disse faktorene er ofte, men ikke nødvendigvis, sammenfallende. Det finnes verk som regnes som verdifulle litterært sett, men som er lite egnet til ekserpering, f.eks. fordi språket er idiosynkratisk. Det kan også være vanskelig å trekke ut sitateksempler fordi det ikke er mulig å hente ut korte sitater uten at den kontekstuelle meningen blir uklar. Andre litterære verk byr umiddelbart på fyndige belegg på typisk språkbruk. Noe av det som gjør skjønnlitterære forfattere ekserperingsvennlige, er kontakten med hverdagslivet og det dagligdagse språket.

Dette gjør tekstene interessante i ordboksammenheng. Det er også viktig å kunne vise frem spennvidden i den enkelte forfatters ordforråd og uttrykksmåter. Allerede i innledningen til de fire første bindene av NRO nevnes det at Henrik Ibsens forfatterskap er sitert bl.a. fordi språket hans spenner «fra de historiske dramaers arkaisering til nutidsdramaenes levende og nyanserike dagligtale». Derfor har man i NRO også vært spesielt interessert i hverdags-språket hos forfattere som Oskar Braaten, Harald Meltzer, Sigrid Undset, Gunnar Heiberg, Helge Krog mfl. (jf. Knudsen et al. 1983 [1937]:V). Denne linjen er fortsatt i NAOB med sitater fra forfattere som Karl Ove Knausgård og Tove Nilsen, det omtalte viselyriske materialet m.m.

4. Litteratursyn

Som nevnt er NAOB en videreutvikling av NRO. Der la man vekt på å få med mange sitater fra den store litteraturen, den litterære kanon. Derfor finnes det en enorm mengde sitater av 1800-tallsforfattere som Henrik Wergeland, J.S. Welhaven, Bjørnstjerne Bjørnson, Amalie Skram, Henrik Ibsen, Alexander L. Kielland og Jonas Lie. I sankingen av sitater fra Ibsens forfatterskap og Bibelen har man kunnet støtte seg på allerede eksisterende seddelarkiver, og på den måten har man fått inn en mengde sitater (jf. innledningene til NRO, til bind I, Knudsen et al. 1983 [1937] og til supplementsbind V, Noreng 1995). I NROs supplementsbind V og VI fra 1995 ble hele Wergeland-samlingen til Det norske litterære ordboksverk (en tidligere leksikografisk enhet ved Universitetet i Oslo) benyttet. I NRO er riktignok også mindre kjente forfattere ekserpert og sitert, og til tilleggsbindene ble det ekserpert noe nyere litteratur, også barne- og ungdomsbøker, men allikevel er det klassikerne som dominerer kildematerialet. Ifølge innledningen til NRO har man dessuten lagt vekt på geografisk spredning av

de skjønnlitterære forfatterskapene som er sitert, ikke minst for å få frem variasjoner i skriftspråket som springer ut fra dialektene. Avisene er i NRO sitert først og fremst som kilder til talespråkspregede ord.

NROs vektlegging av geografisk spredning blir videreført av leksikografene i NAOB. Dagsaktuelle ord, ikke minst nyord, og ord fra talespråket fanger vi også opp. Vi venter likevel med å ta inn nyord til de har fått en viss utbredelse, slik at rene døgnfluer og rariteter unngås. Avistekster brukes fortsatt flittig som kilde, og utover det typisk talespråkspregede, er de særlig nyttige ressurser for ord knyttet til politikk, samfunnsliv, nyere teknikk og sport.

Det har allikevel skjedd et skifte i synet på det litterære tekstgrunnlaget. Dette har bl.a. sammenheng med at den gruppen av ordbokbrukere som man henvender seg til, er blitt større og mer sammensatt. Dette skyldes at den digitale ordboken NAOB når videre ut enn NRO. NAOB er gratis tilgjengelig på internett og enklere å lese enn NRO, som enten måtte kjøpes eller letes opp i et bibliotek, og som hadde et grafisk kompakt utseende som kunne være vanskelig å tyde. Inntrykket av at NAOB-brukerne er en heterogen gruppe, forsterkes av at NAOB-redaksjonen jevnlig blir kontaktet via e-post med ordforslag, kommentarer og spørsmål fra så vel skoleelever, studenter, personer som tar norskkurs, ivrige lesere av skjønnlitteratur, som lærere, forskere, profesjonelle oversettere og forfattere.

Noe som skiller praksis i NRO fra praksis i NAOB, er at man i NRO la stor vekt på å dekke enkelte store, kanoniske forfatterskap, inkludert særheter og rariteter, som vi alt har sett. I NAOB har vi vært mer opptatt av å få belagt typisk språkbruk. Ikke minst er *representativitet* et viktig aspekt for oss. I tillegg til å dekke kjente forfatterskap legger NAOB-redaksjonen vekt på å representere flere grupper av forfattere og temaer. Sakprosa blir like høyt verdsatt som kilde som skjønnlitteratur. Dette er i tråd med utviklingen i leksikografifaget generelt, hvor sakprosa, inkludert avistekster,

spiller en viktig rolle i nyoppbygde korpus som grunnlag for ordbøker (jf. Heinonen & Sandström 2024:89 og Forsberg & Holmer 2024:64–71). NAOB vektlegger skjønnlitteraturen i grunnlagsmaterialet sterkere enn en del andre nordiske ordbøker som dokumenterer samtidsspråket. Dette henger trolig sammen med at NAOB samtidig er en historisk ordbok og en samtidsordbok, og at ordboken utelukkende dokumenterer skriftspråket.

I tilfanget av skjønnlitteratur har vi i NAOB som nevnt økt andelen av kvinnelige forfattere, og vi har lagt til en god del barne- og ungdomslitteratur i de digitale tekstkorpus som ordboken bygger på. Vi henter ikke sitater bare fra fremragende litteratur, men er også opptatt av at kildematerialet – ikke bare sakprosaen, men også skjønnlitteraturen – dekker forskjellige domener, noe også andre nordiske leksikografer arbeider med (jf. Heinonen & Sandström 2024:96). Slik er f.eks. en roman med handling fra bankvesen eller sykehusmiljø et nyttig tilskudd. Forfattere som skriver om etniske minoritetsmiljøer eller spesielle erfaringer (f.eks. med funksjonshemming), kan berike ordboken med viktig ordstoff. Samfunnets behov for språklig dokumentasjon og behovet for at ordbøker er oppdaterte og samtidig representerer flere grupper i samfunnet, blir også vektlagt. Kort sagt har vi i NAOB en bredere oppfatning av hva som regnes som litteratur, et syn som er i samsvar med det som er rådende i våre dager.

5. Tekstgrunnlaget – tekstkritiske betraktninger

Sitateksempler kalles også *autentiske eksempler*. Autentisiteten bør medføre en bevissthet, både om hva man bygger på, og om hvordan man gjengir litteraturen i ordboken. Dette er et stort felt, som jeg bare kort skal berøre her. I den grenen av norsk edisjonsfilologi som gjelder utgivelse av litteratur basert på trykte utgaver fra moderne tid, er det nå alminnelig praksis, især i tradisjonen ba-

sert på tysk edisjonsfilologisk metode, å vektlegge førsteutgaver av skjønnlitterære verk (jf. Kraft 1990 og Kondrup 2011). Årsaken er at førsteutgavene har dannet grunnlaget for bruks- og virkningshistorien til et litterært verk. I norsk leksikografi har den tekstkritiske bevisstheten vært lav; i NRO kan det virke som man har eksperimentert de utgaver man har hatt for hånden. Valget av utgave kan imidlertid få konsekvenser for datering av ord. Et kjent og talende eksempel, som ikke er hentet fra NAOB, men fra nynorsk litteratur, er dateringen av det særnorske (og noe gammeldagse) ordet *pjølter* ('drink av mineralvann (som selters, soda) tilsatt konjakk eller whisky'). På 1970-tallet ble ordet hevdet å være tidlig belagt i Arne Garborgs roman *Bondestudentar*, et sentralt verk i norsk litteratur. Romanen utkom som føljetong i avisen Fedraheimen i 1882 og i bokform i 1883. Oddvar Nes (1982:22) har påvist at dateringen ikke stemmer; ordet forekommer i sammensetningen *raudvin-pjølter* ('drink av mineralvann og rødvin', for øvrig en atypisk *pjølter*-type) først i tredje utgave av verket fra 1908, mens drikken er omtalt som «selters med rødvin [sic] i» i 1882 og 1883. Riktignok har man tidlige avisbelegg for simplekset *pjølter* (med stavemåten «pjaalter») fra Oslo-avisen Dagen 1882 og fra Morgenbladet 1885, men tilknytningen til Garborgs *Bondestudentar* på 1880-tallet er altså misvisende. (Derimot bruker Garborg ordet i formen «pjølter» i romanen *Mannfolk* fra 1886.) Saken illustrerer hvor viktig det er med bevissthet om tekstgrunnlaget for datering av ord. Litterære verk er revidert og modernisert, også av forfatterne selv, slik at teksten kan være sterkt endret i senere utgaver av samme verk.

I NAOB er målet derfor å sitere fra førsteutgaver. Vi arbeider også fortløpende med å erstatte teksten i adskillige sitater fra NRO-tiden med tekst fra førsteutgaver, men dette er et langt lærer å bleke. Et særlig fenomen er kildehenvisninger fra NRO hvor selve sitatet er utelatt, hovedsakelig for å spare plass i papirutgaven. Slike sitater arbeider vi også med å fylle ut – også da i størst mulig grad med førsteutgaver som kilder.

Kravet om en nøyaktig gjengivelse av originalteksten er et sentralt edisjonsfilologisk prinsipp. I en ordbok kan dette prinsippet komme i konflikt med hensynet til at sitateksemplene skal presenteres helhetlig for brukerne. Dette er sannsynligvis årsaken til at man har valgt å erstatte versaler med minuskler i begynnelsen av substantiver i sitater fra eldre tekster i NRO. Denne praksisen har det vært nødvendig å videreføre i NAOB, ellers ville man måtte skrive inn samtlige NRO-sitater på nytt, noe som ville medføre store kostnader og være svært tidkrevende. En formildende omstendighet er det at forskjellen mellom versaler og minuskler i substantiver ikke er meningsbærende. Dessuten er sitatene lenket til digitaliserte versjoner av verkene, som finnes i det norske Nasjonalbibliotekets digitale tekstsamling Nettbiblioteket, slik at ordbokbrukerne kan se originaltekstene der.

6. Metoder for ekserpering, søk og sanking

Tidligere var som kjent ordbokkorpus fysiske; man ekserperte ved å lese på papir, samlet ekserptene i et seddelarkiv og baserte deretter redigeringen av ordbokartiklene på disse sedlene. I våre dager har man i tillegg mulighet for både å søke i og sanke fra store digitale tekstkorpus. En del av sitattilveksten i NAOB har foregått ved hjelp av god gammeldags ekserpering. En slik ekserpering er et nyttig supplement til de moderne metodene, dels fordi ikke alle tekster er tilgjengelige digitalt, dels fordi man får frem spesielle betydninger, stilistiske særegenheter og idiomatiske uttrykk, som kan være vanskelige å identifisere med de digitale metodene. Et nytt belegg for substantivet *sak* i betydningen 'spisevare' ble f.eks. funnet under ekserpering, og den ville vært vanskelig å finne gjennom et digitalt søk i korpus:

fine pannekaker sånne ordentlige gode saker som smaker både sukker og kanel! (Lillebjørn Nilsen Tekst og musikk: *Lillebjørn Nilsen* (2004)¹⁶¹) | fra visen «Gutten med lutten»

Et eksempel på en spesiell bruk av verbet *blusse*, som vanligvis brukes med personlig subjekt, er følgende sitat, markert «sjelden, om ting, ytring» (alternativt kunne det ha vært markert med stilmarkøren «besjeling»):

han takker mig i et brev, som blusser af godhet (*Knut Hamsuns brev I* 100)

Det er gjort målrettede søk i flere tekstkorpus for å finne belegg for ord som enten har stått uten sitat i vår ordbokbase, eller som har hatt få eller bare eldre belegg. Flere forskjellige korpus er brukt i undersøkelsen av frekvens og bruksmåter for ord eller uttrykk, inkludert nyord. Viktige her er Nettbiblioteket, aviskorpuset A-tekst/Retriever og NAOBs eget interne korpus, som består av en samling av skjønn- og faglitterære tekster. Leksikografisk bokmålskorpus fra Universitetet i Oslo og Google Books ble tidligere brukt, men er nå i stor grad erstattet av Nettbiblioteket. Mange av tekstene vi ønsker å sitere, er så nye at vi av rettighetsgrunner ikke har tilgang til dem via Nettbiblioteket, og det er derfor praktisk at NAOB-redaksjonen har fått en avtale med en rekke forlag om å få tilsendt pdf-filer av nye utgivelser jevnlig. Pdf-filene utgjør et eget, internt tekstkorpus. Dette kan vi søke i med programmet DocFetcher.

En annen metode som har vist seg å være nyttig, er sankning av «nøkkelord» fra en pdf-fil som inneholder et litterært verk. Til dette bruker vi også NAOBs interne tekstkorpus. Ved hjelp av dataprogrammet Sketch Engine kan en slik fil sammenlignes med et større bokmålskorpus som programmet har tilgang til, og man får skilt ut de mest frekvente ordene i det aktuelle verket. Mens man ved ekserpering har en tendens til å plukke ut det spesielle, kan en

systematisk digital sanking gi et mer typisk bilde av språkbruken. En kombinasjon av forskjellige metoder er nyttig, siden de fører til forskjellige resultater og derfor er egnet til å utfylle hverandre.

7. Sammenfatning

Sitateksempler spiller en sentral rolle i en dokumentasjonsordbok som NAOB. De har flere forskjellige funksjoner i ordboken. For det første bidrar litterære sitater til å vise frem og forklare ord og uttrykk, for det andre gir de autentiske eksempler på bruk av ord og uttrykk, og for det tredje bidrar de med smakebiter på forskjellige forfatters ordforråd og stilistiske særegenheter. En historisk dokumentasjonsordbok kan ha både en resepsjons- og en produksjonsfunksjon. Sitateksemplene er nyttige for begge disse funksjonene.

Arbeidet med å oppdatere og videreutvikle ordboken med sitateksempler og tilgjengeliggjøringen av den informasjonen sitateksemplene bidrar med i ordboken, står sentralt og berører flere aspekter. Representativitet er viktig i valg av kildemateriale for ordboken. Både forskjellige forfatterskap, sjangre og litterære perioder bør være representert i en diakron ordbok. Man bør også ha en balanse mellom mannlige og kvinnelige forfattere, og mellom skjønnlitteratur og sakprosa. Dette vil ikke minst bidra til stilistisk bredde. Kildene bør dekke både hverdagsspråket og forskjellige domener. Særegenheter ved visse sitater kan fremheves, og dermed tilgjengeliggjøres bedre, med bruksmarkører og utdypende forklaringer.

Til grunn for kildeutvalget ligger et bestemt litteratursyn. NAOB er en videreutvikling og modernisering av NRO. Der siterte man gjerne fra den litterære kanon, mens man i NAOB har et bredere litteratursyn, hvor f.eks. sakprosa og barne- og ungdomslitteratur spiller en viktigere rolle. Både i NRO og i NAOB

er man opptatt av litteratur som dekker dagligspråket og som har en viss geografisk spredning. Litteraturutvalget må gjøres ut fra en bevissthet om disse faktorene, men også ut fra kjennskap til litteraturhistorien og til samtidslitteraturen.

Sitateksempler eller autentiske eksempler krever en tekstkritisk bevissthet hos leksikografen om hva som er tekstgrunnlaget for sitatene. Dette har bl.a. betydning for dateringen av ord. På linje med moderne tysk edisjonsfilologisk praksis legger man i NAOB vekt på å sitere fra førsteutgaver av litterære verk, fordi disse har dannet grunnlaget for verkets bruks- og virkningshistorie. Kravet om nøyaktig gjengivelse av originalteksten er et viktig edisjonsfilologisk prinsipp, som imidlertid kan komme i konflikt med en helhetlig grafisk presentasjon av sitatene i ordboken. Arbeidet med NAOB har tydeliggjort at litteraturvitenskapelig og tekstkritisk kompetanse er uunnværlig når man arbeider med ordbøker som dokumenterer språket basert på litterære tekster.

I våre dager finnes det flere metoder for innhenting av sitateksempler i ordboken. Man kan ekserpere manuelt, slik man gjorde tidligere, men man kan også søke og sanke materiale digitalt. En kombinasjon av forskjellige metoder er nyttig, da de fører til forskjellige resultater og derfor er egnet til å supplere hverandre.

Forhåpentlig vil refleksjonene omkring og erfaringene fra arbeidet med sitatmaterialet i NAOB kunne være til nytte for andre dokumentasjonsordbøker. Ikke minst bør man vurdere fordelingen mellom forskjellige typer litteratur og hvordan tekstutvalget bør være i et grunnlagskorpus. Kanskje bør skjønnlitteratur i større grad innlemmes i korpus, siden man slik får mer variert tilfang av ord og uttrykk. Det er likeledes nødvendig med en tekstkritisk bevissthet når man velger tekstene som man bygger ordbokpresentasjonen på.

Litteratur

Ordbøker

- NAOB = *Det Norske Akademis ordbok*. Det Norske Akademi for Språk og Litteratur. <<https://naob.no/>> (juni 2025).
- NRO = Knudsen, Trygve et al. (1983 [1937]) og Noreng, Harald (1995): *Norsk riksmålsordbok*, bind I–VI. Oslo: Kunnskapsforlaget / Det Norske Akademi for Sprog og Litteratur.

Annen litteratur

- Bergenholtz, Henning (1996): Korpusbasert leksikografi. I: *LexicoNordica* 3, 5–18.
- Forsberg, Markus & Louise Holmer (2024): Datatillgang, metodutveckling och lexicografiskt arbete vid Språkbanken Text. I: *LexicoNordica* 31, 61–79.
- Heinonen, Tarja Riitta & Caroline Sandström (2024): Skevheter och utmaningar i ordboksbasen för fyra enspråkiga finländska ordböcker. I: *LexicoNordica* 31, 81–101.
- Kondrup, Johnny (2011): *Editionsfilologi*. København: Museum Tusculanums Forlag.
- Kraft, Herbert (1990): *Editionsphilologie. Mit Beiträgen von Jürgen Gregolin, Wilhelm Ott und Gert Vönhoff. Unter Mitarbeit von Michael Billmann*. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Knudsen, Trygve et al. (1983 [1937]): Innledning. I: *Norsk riksmålsordbok*, bind I. Oslo: Kunnskapsforlaget / Det Norske Akademi for Sprog og Litteratur. III–IX.
- Nes, Oddvar (1982): Ablegøy(er), gøy og pjølter. Ein rapport om arbeid med opp- og avklaring av ordgåter. I: *Skriftserie nr. 23, serie B*. Bergen: Universitetet i Bergen, Institutt for fonetikk og lingvistikk. 1–29.

- Noreng, Harald (1995): Innledning. I: *Norsk riksmålsordbok*, bind V. Oslo: Kunnskapsforlaget / Det Norske Akademi for Sprog og Litteratur. VII–XII.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2. utgave. Stockholm: Norstedts Akademiska Förlag. 50–95.
- Wills, Tarrin (2022): En historisk ordbogs digitale fremtid. I: *LexicoNordica* 29, 175–194.

Hanne Lauvstad
leksikograf og hovedredaktør, dr.art.
Det Norske Akademis ordbok
Grensen 3
NO-0159 Oslo
hanne.lauvstad@naob.no

Vetenskaplig akribi under tidspress – och andra utmaningar att hantera vid revideringen av en stor historisk ordbok

Pär Nilsson

The first edition of the comprehensive diachronic-contemporary Swedish Academy Dictionary (SAOB) was completed in 2023 after more than 130 years of work. In 2024, a revision project was initiated with the aim of creating a more balanced dictionary by modernizing the content and addressing the major shortcomings. This paper discusses the revision process and the challenges encountered along the way, which are also relevant to other revision projects of large historical dictionaries of the SAOB type. Special attention is given to the difficulty of achieving precision and accuracy within tight time frames and to decision-making related to the combination of old and new parts within one and the same dictionary. The conclusion is that the workflow and review procedure need to be adjusted, and that clarity and transparency towards the user is crucial regarding both new and older versions of the lexicographic descriptions and the extent to which updating is carried out.

1. Inledning

Under 2023 publicerades band 39 av *Svenska Akademiens ordbok* (SAOB), och med det fullbordades Nordens största ordboksverk med beskrivningar av ord från A till Ö efter mer än 130 års arbete. Den utdragna publiceringstiden har lett till att stora delar av innehållet i ordboken var otidsenliga eller omoderna redan när den första upplagan färdigställdes, och för att komma till rätta med det bedrivs sedan 2024 ett revideringsarbete vid redaktionen, vilket beräknas pågå under ca 7 år. I den här artikeln redogörs för detta revideringsprojekt, med särskilt fokus på områden som innebär utmaningar för redaktionen, inte minst tidsramarnas påverkan

på kvaliteten och uppgiften att kombinera gamla och nya beskrivningar i en och samma lexikografiska resurs. Syftet är att rikta ljuset mot utmaningar och svårigheter, och att diskutera möjliga lösningar, som är relevanta också vid andra revideringsprojekt av större dokumentationsordböcker av SAOB:s typ.

2. Bakgrund

Svenska Akademiens ordbok, eller *Ordbok över svenska språket utgiven av Svenska Akademien*, som den fullständiga titeln lyder, är en historisk ordbok som beskriver det svenska språket med startpunkt i 1520-talets nysvenska språkperiod. Det sista bandet av ordboken färdigställdes 2023, men den allra första delen av ordboken publicerades redan 1893, vilket betyder att slutpunkten för SAOB:s beskrivningsperiod varierar mycket beroende på vilken del av verket man tittar i. I de första banden av ordboken omfattar perioden ca 370 år, medan den i de sista banden sträcker sig över lite drygt 500 år (1521–2023).

Ordboksprojektet inleddes egentligen redan i slutet av 1700-talet då kung Gustav III instiftade Svenska Akademien och i stadgarna slog fast att det ”äligger äfven Academien att utarbete en Svensk ordabok” (1786, § 23). Men trots att en del inledande försök gjordes tidigt var det inte förrän mot slutet av 1800-talet som ordboken fick den form som den (än så länge, jfr nedan) har idag. Under denna tid tillsattes en fast ordboksredaktion, med lexikografen K. F. Söderwall som huvudredaktör. Redaktionen har sedan dess huserat i Lund (se Svensén 2015 för en fyligare beskrivning av projektets initiala skeden).

Liksom andra stora historiska ordböcker som såg dagens ljus under 1800-talet är SAOB en omfattande dokumentationsordbok vars första upplaga bygger på ett insamlat excerptmaterial och innehåller utförliga och detaljerade beskrivningar av ordens for-

mer, betydelser och användningar. Materialet samlades in löpande från andra halvan av 1800-talet fram till slutet av 1900-talet och finns förvarat i redaktionens arkiv; sedan 2022 finns det även digitalt tillgängligt i en skannad version. Totalt innehåller ordboken mer än 500 000 uppslagsord fördelade på drygt 33 000 sidor. Förutom i den tryckta versionen, som omfattar 39 tjocka volymer plus källförteckning och efterord, föreligger ordboken i digitalt format, dels på den egna hemsidan (saob.se), dels ihop med de renodlade samtidsordböckerna *Svenska Akademiens ordlista* (SAOL) och *Svensk ordbok utgiven av Svenska Akademien* (SO) på ordboksportalen svenska.se (för en utförligare typologisering av SAOB som ordboksverk se t.ex. Nilsson 2023:24f.).

2.1. Behov och planering av revidering

Redan långt innan publiceringen av det sista bandet av SAOB var det uppenbart att ordboken trots (eller snarare: på grund av) sin långa lemmalista och sina utförliga beskrivningar vid färdigställandet skulle innehålla uppenbara luckor och vara i tydligt behov av revidering. En stor mängd ord som är centrala i det svenska allmänspråket idag saknas, till följd av att de myntades eller etablerades efter det att det aktuella alfabetiska spannet behandlades i ordboken. Många av de beskrivningar som finns med i ordboken behöver uppdateras, antingen eftersom definitionsspråket är föråldrat eller för att företeelsen som beskrivs har ändrat karaktär eller upphört att finnas till osv. Därför inleddes ett planeringsarbete inför en kommande revidering flera år i förväg (och i mer formaliserad form omkring 2016).

I detta arbete tillsattes en mängd olika interna arbetsgrupper som inventerade revideringsbehovet inom en lång rad områden. Bl.a. undersöktes behovet av att modernisera definitionsspråk och språkvetenskaplig terminologi, av makro- och mikrostrukturella omstöpningar och av ett nytt material och vilket det skulle tän-

kas vara. Också beskrivningen av uttryck och konstruktioner och hanteringen av t.ex. förkortningar och av finlandssvenska ord och uttryck sågs över. Arbetsgruppernas arbete dryftades på redaktionen och lades fram och diskuterades i olika akademiska sammanhang. I vad som på redaktionen kallades för en revideringsturné besöktes universitet och forskningsinstitut i Sverige och Finland. Under ordinarie forskarseminarier presenterades idéer och förslag vid bl.a. Lunds, Göteborgs, Uppsala och Umeå universitet, då deltagare bjöds in med möjligheten att påverka den framtida utformningen av SAOB. Dessutom gjordes anhalt hos Språkrådet i Stockholm och hos Institutet för de inhemska språken i Helsingfors (Kotus). Presentationer hölls även i samband med konferenser och symposier arrangerade av Nordiska föreningen för lexikografi, och under konferensen Svenska språkets historia 16, 2021, ägnades ett halvdagsseminarium åt den framtida revideringen av SAOB (se t.ex. Rosqvist & Wendt 2017, Nilsson & Rosqvist 2022, Nilsson 2024a).

Turnén och de olika seminarierna resulterade i en lång rad synpunkter och inspel av både praktisk och mer teoretisk karaktär av vilka flera fick direkt påverkan på utformningen av den reviderade versionen av ordboken. Många kommentarer gällde svårtolkade och problematiska beskrivningssätt och formuleringar och bruket av språkvetenskaplig terminologi, både i den första upplagan och i de presenterade nya förslagen, vilket bl.a. bidrog till beslutet att lösa upp förkortningar i den redaktionella texten med urskillning, och inte konsekvent välja utskrivna form.¹ Ett annat exempel på

1 För en icke-avancerad användare är t.ex. en förkortning som *interr.* inte nödvändigtvis mindre till hjälp än utskrivna *interrogativ*, vilken troligtvis ändå måste slås upp. I sin utskrivna form tar den däremot en större del av definitionstextens totala utrymme i anspråk och hamnar lätt i fokus. Etablerade allmänspråkliga förkortningar som exempelvis *el.* för *eller* skrivs inte heller ut i full form eftersom de är oproblematiska att tolka och eftersom de i utskrivna form i flera fall skulle dominera innehållet i definitionerna. Förkortningen *l.* från SAOB:s första upplaga ersätts därför av etablerade *el.*

feedback som lett till förändring gäller SAOB:s tämligen kryptiska system för uttalsangivelser i den första upplagan. Mot bakgrund av det i hög grad komplicerade företaget att implicera ett modernt system i samtliga artiklar över ord i bruk (obrukliga ord saknade redan tidigare uttal) – och att uttal anges både i skrift och som ljudfil i SAOB:s systerordbok SO på plattformen svenska.se – anges i den reviderade versionen tills vidare inte längre uttal alls. Därmed tas steget fullt ut mot att bli en ordbok med beskrivningar endast över skriftspråket. Efter det inledande planeringsarbetet har redaktionen även inlett ett formaliserat samarbete med Kotus i Helsingfors för hantering av finlandssvenska ord och betydelser i ordboken.

När planeringsarbetet avslutats stod SAOB-redaktionen vetenskapligt på en tämligen stadig grund inför en kommande revidering av ordboken; frågan var bara inom vilka ekonomiska ramar och i vilken regi arbetet skulle utföras. Svenska Akademien som de senaste åren finansierat ordboksprojektet bl.a. med hjälp av ett avtal med svenska staten, meddelade plötsligt 2021 att den inte längre själv kunde finansiera ett revideringsprojekt, sedan nämnda avtal sagts upp av staten, vilket väckte starka reaktioner i mediasverige och det språkvetenskapliga forskarsamhället.² Ett hot om nedläggning avväjades när Svenska Akademien tillsammans med Riksbankens Jubileumsfond, Helge Ax:son Johnsons Stiftelse, Kungliga Vitterhetsakademien och Svenska litteratursällskapet i Finland bildade ett konsortium som garanterade revideringsprojektet 100 miljoner kronor fördelade över maximalt 10 år.

Med dessa ramar givna har ett revideringsarbete påbörjats som avser att åstadkomma ett mer balanserat ordboksverk genom att modernisera innehållet och avhjälpa de värsta bristerna. Den reviderade versionen av ordboken kommer endast att publiceras i

2 Jfr t.ex. följande nyhetsartiklar i Svenska Dagbladet och på Sveriges televisions nyhetssajt: <https://www.svd.se/a/pWlGVE/lat-inte-den-svenska-ordskatten-ga-forlorad>, <https://www.svt.se/nyheter/lokalt/skane/svenska-akademiens-ordbok-finansieringen-oviss-efter-bokstaven-o>.

digital form. Med bibehållen personalstyrka (redaktionen består i skrivande stund av 18 personer i olika tjänstgöringsgrader) och efter en ökad inflation de senaste åren beräknas finansieringen vara i ca 7 år. Uppdateringar kommer att göras löpande under projektperioden med en beräknad första nypublicering under 2026.

2.2. Utblick

I det följande ges en kort beskrivning av revideringsstatus för några av SAOB:s historiska systerordböcker med avsikten att sätta SAOB:s revideringsprojekt i perspektiv. För SAOB är nu inte den första historiska ordbok som står inför att revideras och uppdateras – det förhåller sig i det närmaste tvärtom. I den nordiska kontexten föreligger samtliga systerordböcker, *Ordbog over det danske Sprog* (ODS), *Norsk Ordbok – Ordbok over det norske folkemålet og det nynorske skriftmålet* (NO) och *Det Norske Akademis ordbok* (NAOB), i någon utsträckning i reviderad form. Det 28 band stora verket ODS (färdigställt 1956) försågs med ett supplement på fem band mellan 1956 och 2005, vilket sedermera läggs till i den digitala versionen av ordboken (se t.ex. Trap-Jensen 2005). Det tolfte och sista bandet av den första upplagan av NO gavs ut 2015. Den officiella³ digitala versionen av ordboken omfattar dock i nuläget endast det alfabetiska spannet I till Å, vilket föranledde det pågående uppdaterings- och digitaliseringsprojektet *Norsk Ordbok a–h* (se Myking, Grønvik & Helset 2020) vari även redan digitalt tillgängliga artiklar moderniseras och uppdateras. NAOB är resultatet av en omfattande revidering och utvidgning av *Norsk Riksmålsordbok*, som gavs ut i fyra band mellan 1937 och 1957 med ett supplement på två band som kom 1995. Sedan 2017 finns ordboken digitalt tillgänglig i en version som uppdateras löpande (och enligt hemsidan motsvarar en tryckt upplaga om 12 band; se t.ex. Nilstun 2016, jfr Lauvstad 2025).

3 På sidan <https://alfa.norsk-ordbok.no/> finns emellertid redan nu en betaversion av den reviderade ordboken.

Översikten ovan antyder en diversitet i sätten att ta sig an och utföra revidering och sätter fingret på hur avgörande tidpunkten är då den första upplagan färdigställdes. Både ODS och *Norsk Riksmålsordbok* avslutades i en tid då pappersordböcker var det rådande mediet och tryckta supplement var den givna vägen framåt. För NO var situationen en annan, och den första upplagan var ännu inte komplett innan digitaliseringsarbetet och planeringen av en följande revidering inleddes. Det ska också tilläggas att ODS efter att första upplagan och supplementet färdigställdes övergick från att vara en diakronisk samtidsordbok till en renodlat historisk ordbok, med ett slutår omkring 1950. Redan i arbetet med supplementet var denna tidsgräns dragen. I de övriga projekten hålls kontakten med samtidspråken levande, och det är också den väg som SAOB väljer. SAOB är i förhållande till de ovannämnda ordböckerna ett betydligt mer omfattande verk, vilket säkert delvis kan förklara projektets tidsutdräkt och senfärdighet, men inte är den enda påverkande faktorn. Avgörande är förstås också vilka resurser och tekniska hjälpmedel som funnits till förfogande och utformningen av arbetsrutiner osv., varför rättvisa jämförelser sinsemellan är svåra göra. Men även i förhållande till storleksmässigt och strukturellt mer likvärdiga verk är SAOB sist ut att revideras och uppdateras.

Oxford English Dictionary (OED, ursprungligen *A New English Dictionary*, NED) färdigställdes i sin första upplaga 1928, och har sedan dess försetts med supplement i olika omgångar, vilka sedan har integrerats med originaltexten i en ny upplaga. Sedan 1990-talet bedrivs ett omfattande revideringsarbete på redaktionen där varje artikel ses över och uppdateras. Under 2020 beräknades ca 40 % av ordboken i detta projekt vara bearbetad (Se Gilliver 2016, Brewer 2022).

Det sista bandet av den första upplagan av *Deutsches Wörterbuch von Jacob und Wilhelm Grimm* (DWB) publicerades 1961, och i samband med det påbörjades arbetet med supplementet av

det alfabetiska spannet A–F (vilket färdigställdes 2018, se Harm 2019). *Woordenboek der Nederlandsche Taal* (WNT), som med sina 40 tryckta ordboksband räknas som den största ordboken i världen, färdigställdes 1998. Redan innan den första upplagan var klar påbörjades arbetet med ett supplement (som dock endast hann komma till bokstaven A innan det lades ned i sin dåvarande form). 2001 publicerades ytterligare tillägg i tre tryckta volymer (se Moerdijk 1994, 2001).

I förhållande till samtliga ovan nämnda verk är alltså SAOB en tydlig efterslänrare. I det här sammanhanget är det dock inte enbart ofördelaktigt att komma sist, det finns lärdomar att dra av andra projekt, och dagens tekniska förutsättningar är gynnsammare än de var för bara ett par decennier sedan. SAOB slipper t.ex. ta steget via ett separat supplement, vilket ur användarsynpunkt aldrig varit en optimal lösning. Genom kontinuerlig digital uppdatering finns också bättre förutsättningar för att hålla ordbokens samtidsspråkliga perspektiv levande.

3. Revideringens huvudspår

Sedan 2024 pågår alltså ett reviderings- och uppdateringsprojekt på SAOB-redaktionen, och de arbetsuppgifter som utförs inom ramen för detta projekt kan samtliga sorteras in i någon av följande fyra kategorier: 1) anpassning av ett nytt redigeringsystem och design av ny artikelstruktur, 2) tillägg av nya artiklar, 3) automatiska och systematiska förändringar av redan publicerat material, eller 4) manuell revidering. I föreliggande avsnitt sammanfattas huvudprinciperna bakom dessa kategorier mer i detalj, och några principiellt viktiga ställningstaganden och överväganden och praktiska omständigheter att förhålla sig till diskuteras (jfr äv. Nilsson (2024a) för en mer detaljerad beskrivning med delvis parallell utgångspunkt).

3.1. Nytt redigeringsverktyg, ny artikelstruktur

SAOB har, som flera andra historiska ordböcker, en väldigt komplex artikelstruktur med djup hierarki och många sinsemellan underordnade artikeltyper och betydelsemoment i olika nivåer. På makronivå, under en huvudartikel, förekommer artiklar med sammansättningar och avledningar som i sin tur inte sällan har egna avledningar och sammansättningar. På mikronivå, inom en och samma artikel, innehåller den maximala strukturen moment betecknade med romersk siffra, följt av sådana betecknade med arabisk siffra, med kursiv latinsk bokstav, kursiv grekisk bokstav samt kursiv grekisk bokstav tillsammans med primtecken ('): I. ↓ I. 1 ↓ I. 1 a ↓ I. 1 a α ↓ I. 1 a α α'. Redovisningen av olika böjningsformer och paradigm är också ofta i hög grad komplex, och genom åren har dessutom flera olika system tillämpats för sådan redovisning. Detta sammantaget – ihop med det faktum att själva textmängden som ska hanteras alltså är tämligen diger – har inneburit att det länge var svårt att hitta ett lämpligt redigeringsverktyg att använda. Efter ett par misslyckade försök (från omkring millennieskiftet) med olika mer avancerade verktyg skrevs därför den första upplagan av ordboken färdigt i en enkel mall i ett ordbehandlingsprogram. Innehållet i artiklarna taggades sedan i efterhand och gjordes tillgängligt för användarna i form av en digital version av den tryckta ordboken.

I den reviderade versionen av ordboken används i stället verktyget *Digital Publishing System* (DPS), från det franska företaget IDM. Detta redigeringsverktyg är etablerat och välkänt och används av flera andra ordboksredaktioner i Europa. Avgörande för SAOB-redaktionen var att även redaktionen för OED använder en version av DPS (jfr Elliott & Williams 2006, Abel 2012). Det faktum att dessa båda ordböcker är väldigt lika med avseende på innehåll och uppbyggnad – ihop med att verktyget bevisligen kan hantera mycket stora textmängder – innebär att det

redan från början har goda förutsättningar att vara lämpligt för SAOB.

Innehållet från den första upplagan av SAOB har nu förts över till det nya redigeringsystemet och ordboken stöps om till att bli en digital databas, en helt digital ordbok och inte bara en digital *version* av en ordbok. Syftet med detta är förstås att automatisera och underlätta redigeringen, att möjliggöra utökade sökmöjligheter för användarna (och för redaktörerna i det lexikografiska arbetet) och att kunna visa upp innehållet i användargränssnittet på ett tydligare och mer överskådligt sätt. För att åstadkomma ökad tydlighet och överskådlighet behöver dock den befintliga strukturen ändras. Målet är att separera olika informationskategorier i artiklarna på ett tydligare sätt och att i större utsträckning placera vissa typer av information i in- och utfällbara rutor som inte stjäl uppmärksamhet från presentationen av mer centrala delar av artiklarna. Exempelvis bör längre utläggningar under rubriken *Anmärkning* och utrymmeskrävande böjningsredovisningar kunna hanteras på detta sätt. Dessutom ska den underordnade kategorin *sammansättningar* få en mer framträdande plats och bli mer självbärande i förhållande till sina överordnade huvudord (bl.a. genom att hela uppslagsformen presenteras, inte bara efterleden, att de ingående sammansättningslederna identifieras på lemma- och betydelsenivå och att information relevant för den enskilda sammansättningen presenteras inom sammansättningsartikeln och inte bara på en övergripande position som en allmän kommentar till samtliga sammansättningar).

Modificeringar och omställningar av det här slaget kräver förstås anpassningar i redigeringsverktyget, och för detta ändamål finns både en ansvarig språkteknolog i redaktionen och ett tätt samarbete bedrivs med utvecklarna på IDM. Dessutom är övriga redaktionsmedlemmar delaktiga i praktiskt utformnings- och taggningsarbete.

3.2. Tillägg av nya artiklar

En stor anledning till att mycket av innehållet i SAOB upplevs som föråldrat och i behov av revidering är att många ord som är centrala i det svenska ordförrådet idag (2025) saknas i ordbokens lemmalista. De senaste decenniernas tekniska ord och termer som *AI*, *app*, *dator*, *fax* och *miniräknare* saknas till följd av att de aktuella alfabetiska spannen bearbetades i SAOB innan respektive fenomen hade hunnit sätta sitt avtryck i språket. Men också sådana (idag) vardagsnära ord som *allergi*, *bacon*, *dammsuga* och *empati* lyser med sin frånvaro. För att råda bot på detta ska 10 000 nya artiklar läggas till i ordboken under projektperioden. Vad som utgör ett rimligt antal ord att lägga till vid revideringen går förstås att diskutera, men i det här fallet har det beslutade antalet bedömts utgöra en tillräcklig mängd för att användarna ska uppleva skillnad och tydlig förbättring, men samtidigt vara så begränsat att det i praktiken är möjligt att hantera inom de givna tidsramarna. Antalet motsvarar ungefär innehållet i ett genomsnittligt tryckt band i den första upplagan.

För att sätta SAOB:s antal tillägg i perspektiv kan här nämnas att OED-redaktionen i arbetet med sin andra upplaga av ordboken lade till ca 5 000 artiklar (se Gilliver 2016), medan revideringen av de norska samtidsordböckerna *Bokmålsordboka* (BOB) och *Nynorskordboka* (NOB; se t.ex. Selback et al. 2025) resulterade i över 20 000 tillagda artiklar sammanlagt i de båda ordböckerna. BOB och NOB är förvisso strukturellt sett ganska annorlunda SAOB, vilken ofta innehåller betydligt mer omfattande beskrivningar, men detta norska revideringsprojekt omfattade en tidsmässigt nästan identisk projektperiod som SAOB:s (nämligen från 2018–2024) varför en jämförelse trots allt är relevant.

Urvalet av ord baseras på ett antal grundläggande kriterier (se Nilsson 2024b). Bl.a. bör orden som kommer med ha (eller ha haft) hög frekvens och stor spridning i allmänspråket. Vidare har

sammansatta ord lägre prioritet än simplexord. Simplexordet *curling* saknades t.ex. i den första upplagan av ordboken och det bedöms ingå i det centrala ordförrådet och rimligt att lägga till i den reviderade versionen. Däremot förbigås t.ex. *curlingbana* trots stor spridning och hög etableringsgrad, eftersom ordbildningsmönstret och semantiken är förutsägbar. Ordens semantik är alltså också avgörande, och en parameter att förhålla sig till är i vilken utsträckning ordet är i behov av förklaring. Av denna anledning har ordet *curlingförälder*, 'förälder som röjer hinder för barn på sätt som riskerar att göra det överbeskyddat o. bortskämt', tagits med trots att ordet är en sammansättning. Förklaringsbehovet innebär även bl.a. att ord som tidigare varit frekventa i språket, men sedan minskat i bruk, är extra intressanta för en historisk ordbok som SAOB. Ord som riskerar att hamna utanför svenska renodlade samtidsordböckers lemmalistor av denna anledning utgör därmed nyordskandidater för SAOB.

Redaktionen arbetar med olika metoder för att sammanställa listor över potentiella kandidater för nya artiklar. Det initiala arbetet bygger till stor del på en jämförelse som gjorts mellan lemmalistorna för SAOB, SAOL och SO. Ett urval av ord som saknas i SAOB men finns med i systerordböckerna har valts ut för att täppa till akuta luckor i SAOB:s lista.

3.3. Automatiska och systematiska revideringsinsatser

Mycket av innehållet i den första upplagan av SAOB är skrivet för lång tid sedan. Spannet A–R, dvs. mer än hälften av alfabetet, behandlades för mer än 60 år sedan, och detta märks i ordboken både i fråga om stilen hos det redaktionella metaspråket och i aktualiteten för många av beskrivningarna. Ordboken innehåller många ålderdomliga formuleringar och stavningar (fram till bokstaven D förekommer t.ex. gammal stavning med *hv-*, *-fv-* o. *-dt* i den redaktionella texten, och plural verbböjning påträffas i definitionstext

ända fram i alfabetet till S)⁴, men också artiklar som är i behov av uppdatering eftersom den beskrivna företeelsen inte längre förekommer eller avsevärt har ändrat karaktär (*chaufför* beskrivs t.ex. som en ”eldare på en automobil”, och i artikeln *ettöring* påträffas den idag märkliga formuleringen: ”ej om äldre förhållanden”).

Otidsenligheter av det här slaget behöver också tas itu med i revideringen, och flera av dem går att åtgärda automatiskt. I projektet uppdateras och åtgärdas på automatisk och maskinell väg olika så kallade *aspekter* inom samtliga eller väldigt många artiklar samtidigt. I aspektrevideringen (som den internt benämns) har t.ex. kryptiska och föråldrade förkortningar lösts upp och entydiga gammelstavningar ändrats. Men även principiellt mer långtgående förändringar har genomförts enligt denna metod. Bl.a. har systemet för att ange bruklighet ersatts (vilket t.ex. har fått följden att det tidigare använda korstecknet, †, på de flesta ställen har ersatts av formuleringen ”ej belagt i nusvenska”), och det väldigt intrikata och för läsaren tämligen svårtolkade genusbeteckningssystemet (som omfattade både sexuellt och grammatiskt genus och den ålderdomliga termen *reale*, för utrala icke-personbetecknande substantiv) har fått ge plats åt ett mer binärt system (*utrum* el. *neutrum* + angivande av äldre genus, *mask.* el. *fem.*). (Se Nilsson 2024a för en beskrivning av det nya och gamla systemet att ange genus i SAOB).

3.4. Manuella uppdateringar

Trots att en hel del av SAOB:s brister kan åtgärdas systematiskt på automatisk väg finns det förstås mycket som måste ses över och

4 I Sverige avskaffades de aktuella stavningarna officiellt i och med den senaste stora stavningsreformen 1906, och på 1940-talet övergavs allmänt de plurala verbformerna i skrift (se t.ex. Pettersson 1996), men i SAOB (och även i SAOL) levde de vidare ännu längre: de sista artiklarna i SAOB på bokstaven D publicerades på 1920-talet och utgivningen av ord på bokstaven S påbörjades under 1960-talet.

behandlas manuellt och individuellt. Av den anledningen bearbetas även innehållet i ordboken artikel för artikel av redaktionen. Fokus i detta arbete är det alfabetiska spannet A–R vars artiklar, mot bakgrund av vad som sagts ovan, bedöms vara i störst behov av uppdatering (de sista artiklarna under bokstaven R publicerades 1962). Huvuddelen av arbetet är uppdelad i två nivåer (jfr dock avsnitt 4.1 nedan) och hårt styrt av projektets tidsramar. Målet är att med lätt hand åtgärda och uppdatera så mycket som möjligt inom den begränsade tiden. Artiklarna har grupperats i olika partier omfattande ett visst alfabetiskt spann motsvarande ett visst ungefärligt antal spalter i den tryckta ordboken (ett exempel är det i skrivande stund just avslutade sjunde partiet av bokstaven B, som sträckte sig mellan *batalj* och *bedärva*, motsvarande ca 100 spalter i den tryckta versionen). Först bearbetas varje parti av en så kallad yttredaktör som fokuserar på de lägst hängande frukterna, vilket bl.a. innebär att ta hand om rester av ålderdomlig stavning, åtgärda feltaggningar och föra ut förkortningar som inte kunnat behandlas automatiskt. Detta arbete går relativt fort framåt och är tänkt att generera artiklar av en kvalitet som är anständig att visa upp för användarna. I nästa steg bearbetas partierna av djupredaktörer som uppdaterar och moderniserar definitionerna, lägger till nya språkexempel (från SAOB:s befintliga excerptmaterial eller i form av nyinsamlade⁵ citat från digitalt tillgängliga källor) och ser över att de automatiskt genererade nya bruklighetsangivelserna är adekvata. Även omstruktureringar av innehållet görs i detta steg, och nya betydelser läggs till vid uppenbara behov och där insatsen är försvarbar ur ett tidsperspektiv. Djupredaktörens del av arbetet är på det hela taget relativt tidskrävande, men trots det finns inom det nuvarande projektet inte möjligheten att konsekvent åtgärda och uppdatera allt som kan eller bör revideras, utan ambitionen

5 De korpusverktyg som utnyttjas i detta arbete utgörs i huvudsak av tidningsdatabaserna KB *Svenska Tidningar*, Retrievers *Mediearkivet* samt Språkbanken texts sökverktyg *Korp* (Jfr Nilsson 2024b).

måste i stället vara att förbättra så mycket som tiden tillåter. Ambitionernas relation till tiden utvecklas vidare i nästa avsnitt.

4. Särskilda utmaningar – och möjliga åtgärder

Trots att redaktionen var väl förberedd inför starten av revideringsprojektet har många utmaningar redan nu dykt upp längs vägen, som det gäller att förhålla sig till eller hitta lösningar för. Några av dem har antytts i det föregående, och ett par andra med mer genomgripande inverkan på ordboksverkets karaktär beskrivs närmare här nedan.

4.1. Akribi under tidspress

Den första upplagan av SAOB är präglad av grundliga analyser och beskrivningar, av långa utredningar – och av pålitlighet. Innehållet är noggrant granskat. Arbetsgången involverade förutom ren korrekturläsning dubbla kollegiala granskningar av manus i flera olika omgångar. Det fanns även en särskild kontrollavdelning bland ordboksassistenterna som hade till uppgift att granska hänvisningar i texten och att kontrollera citat och källangivelser mot originalen. Dessutom hade redaktionen regelbunden kontakt med externa sakkunniga granskare och experter inom omkring 200 olika ämnen. Allt detta sammantaget bidrog till att arbetet gick långsamt framåt. Trots att det går att diskutera den första upplagans utgivningstakt, var det alltså inte för intet som ordboken tog mer än 130 år att färdigställa. I relation till det här ser revideringsprojektets tidshorisont på mellan 7 och 10 år helt annorlunda ut, även om de övergripande principerna för arbetet till stor del är desamma. Tid är numera en bristvara, men målet är fortfarande att åstadkomma noggranna och pålitliga beskrivningar, dessutom med ett modernare uttryckssätt och en mer tillgänglig struktur

för användaren. För att nå målet att både utöka lemmalistan med 10 000 artiklar och att uppdatera de äldre delarna av ordboken har därför arbetsgången och granskningsproceduren uppdaterats.

Mest noggrant nagelfars de nya artiklar som läggs till i ordboken. Fortfarande granskas manusen av flera par ögon, läsningar genomförs i olika omgångar, men de är färre till antalet och arbetet har förenklats av digitala hjälpmedel. En gynnsam omständighet i sammanhanget är att de ord som lemmalistan kompletteras med överlag har en kortare historia och mindre komplicerade förhållanden för redaktören att reda ut än många av orden i den första upplagan av SAOB. Det handlar ofta om ord som har bildats eller lånats in i modernare tid och som av den anledningen inte har hunnit utveckla och etablera komplexa nätverk av olika betydelser och specialanvändningar. Analys och beskrivning går alltså ofta fortare på grund av ordens delvis annorlunda karaktär.

Vid revideringen av tidigare publicerade artiklar har granskningsproceduren förenklats ytterligare. Det är helt enkelt inte möjligt att granska alla artiklar mellan A–R lika noggrant, och eftersom detta material redan står på en stabil grund och håller god kvalitet är det inte heller nödvändigt. Revideringen går snabbt fram för att så många artiklar som möjligt ska hinnas med, och det innebär att det ofta är smärre förändringar som görs, vilket i sin tur innebär mindre resurskrävande granskning. Väldigt omfattande artiklar (t.ex. vissa stora verb- eller prepositionsartiklar) förbigås vid manuell revidering i den här omgången eftersom uppdatering till anständig nivå vore för tidskrävande för dem. På sikt är förstås förhoppningen att även denna typ av artiklar ska uppdateras.

Det bör understrykas att den manuella revideringen alltså går alfabetiskt framåt artikel efter artikel i den hastighet som är möjlig att uppbringa. Eftersom nypubliceringen av SAOB:s innehåll planeras ske löpande under projektperioden kommer ordboken vid de inledande publiceringstillfällena innehålla ett stort antal

artiklar som inte är manuellt reviderade överhuvudtaget. Vissa artiklar är dock i mer akut behov av uppdatering än andra, inte minst sådana som utgörs av känsliga eller kontroversiella ord eller innehåller kontroversiella redaktionella beskrivningar. Därför ingår i den manuella artikelrevideringen ett snabbspår för översyn och bearbetning av prioriterade ord. Ett visst antal partier av ord har identifierats som behandlas separat. Dessa partier är tematiskt indelade och innehåller ord som har att göra med exempelvis etnicitet, funktionsnedsättningar och -variationer, genus och sexualitet. Ord som *zigenare*, *idiot* och *dvärg* är exempel som behandlas i denna del av revideringsarbetet. Uppgiften är dels att se över och uppdatera de artiklar där dessa ord utgör huvuduppslagsord och i aktuella fall se till att (efterhand utvecklad) värdeladdning och nedsättande stilprägel beskrivs och förklaras. Dels ses andra artiklar i ordboken över som innehåller dessa ord, varpå olämpliga formuleringar i definitionstext eller andra redaktionella kommentarer ändras; en följd av denna metod blev t.ex. att synonymdefinitionen ”tattare l. zigenare” i artikeln *kittelbotare* ersattes av ”rom el. resande” ihop med en övergripande bruklighetskommentar. Vid den första publiceringen kommer de högst prioriterade partierna att vara reviderade (se Nilsson 2024a för en beskrivning av SAOB:s principer bakom hanteringen av kontroversiella ord).

En ytterligare utmaning relaterad till akribi med snäva tidsramar är möjligheten att åstadkomma ett verk i balans. Den första upplagan av SAOB ser ganska annorlunda ut i början jämfört med i slutet av alfabetet, och ett mål med revideringen är att jämna ut skillnaderna. I vissa avseenden går det att komma en bra bit på vägen; både tillägg av nya artiklar och metoden att revidera ett så stort antal artiklar som möjligt i stället för att fokusera på grundligare omarbetning av ett litet antal förväntas ha detta resultat. Men att åstadkomma fullständig konsekvens i sättet att beskriva orden är i det närmaste omöjligt. Föråldrad språkvetenskaplig teori är exempelvis något som har uppmärksamrats internt på redaktio-

nen som ett område med god förbättringspotential. En modernare terminologisk apparat vore önskvärd, åtminstone kunde man utgå ifrån de termer som används i Svenska Akademiens grammatik (SAG), som är ett standardverk i Sverige. Men det är i praktiken ingen enkel sak att ersätta termerna fullt ut, även om det förstås går att göra punktsatser (så har t.ex. det något ålderdomliga *imperfekt* konsekvent ersatts av *preteritum* i den reviderade versionen). Harm (2024) som diskuterar revideringen av DWB, är inne på samma linje och menar att det inte bara är svårt utan också orimligt att förvänta sig att stora historiska ordböcker ska hålla jämna steg med utvecklingen inom språkvetenskapliga teorier. För SAOB:s del verkar det rimligt att anta ett pragmatiskt förhållningssätt och sträva efter ett så bra, modernt och konsekvent resultat som möjligt, och inte efter ett komplett balanserat verk.

4.2. Gammalt och nytt sida vid sida

En konsekvens av SAOB-redaktionens sätt att uppdatera ordboken är att gamla och nya element kommer att kombineras och samförekomma i större utsträckning än tidigare. Ett grundvillkor för en diakronisk samtidsordbok är förstås att både gamla och nya ord och betydelser tas upp och beskrivs.⁶ Men genom tillägget av nya ord kommer dessutom nyskrivna och över 100 år gamla artiklar stå sida vid sida i den reviderade versionen av SAOB, och nya sammansättningar och avledningar kommer att underordnas huvudord vars beskrivningar publicerades i den första upplagan. Till följd av den manuella uppdateringen av ordbokens äldre innehåll kommer också moderna betydelser läggas till, och nya definitioner kommer därmed att kombineras med äldre inom en och samma artikel. Den automatiska så kallade aspektrevideringen gör

6 Jonsson (1983) visar att denna förutsättning i och för sig inte är okomplicerad, redaktörernas samtidsperspektiv riskerar t.ex. att grumla sikten vid historiska analyser och leda till anakronistiska slutsatser.

även att det i vissa artiklar kommer att finnas både äldre och moderna formuleringar inom en och samma definition. Hanteringen av de plurala verbböjningar som förekommer i den första upplagan av SAOB, och som nämndes i avsnitt 3.3 ovan, kan illustrera detta faktum. Entydiga sådana former kan nämligen sökas fram och ersättas på ett oproblematiskt sätt. T.ex. kan presensformen *äro* av *vara* ersättas med *är*, eller preteritumformer av starka verb, t.ex. *gingo* och *kommo* med *gick* resp. *kom*. I revideringsprojektet har denna typ av ändringar genomförts. Den vanligaste tempusformen i definitionstexten är emellertid presens, och för de allra flesta verb gäller att pluralformen i presens sammanfaller med infinitiven, vilket innebär att de inte kan ersättas automatiskt utan problematiska följdverkningar (infinitivformen är dessutom betydligt mer frekvent i SAOB:s beskrivningar än de plurala verbböjningarna). Förekomster av dessa former har därför inte ersatts automatiskt i revideringen, utan ändras efterhand vid manuell bearbetning. En följd av denna metod blir att i en definition som den i betydelsemoment 2 hos substantivet *kråka*: ”om föremål l. figurer som äro avsedda att föreställa l. som i ngt avseende erinra om en kråka l. en kråknäbb”, har plurala *äro* ändrats till *är* och den otidsenliga förkortningen *l.* har ändrats till mer gängse *el.*, men plurala *erinra*, som sammanfaller med infinitivformen, har inte automatiskt kunnat ändras till *erinrar*, vilket ordboksanvändaren tvingas stå ut med tills de aktuella artiklarna har hunnit bearbetas manuellt. Ett särskilt föremål för den manuella revideringen blir därmed att jämma ut stilmässiga skillnader mellan gammalt och nytt.

Brewer (2022), som med utgångspunkt i OED diskuterar framtiden för historiska ordböcker, sätter fingret just på negativa följdverkningar och det förvirrande intryck som kombination och sammanfogning av moderna och äldre komponenter i ordboksartiklarna riskerar att ge upphov till hos användarna. Hon menar att det finns ett frustrationselement i det faktum att definitionerna i den digitala tidsåldern inte är stabila på samma sätt som förut

– med löpande uppdateringar riskerar innehållet i en och samma artikel ändras på ett subtilt sätt, utan att man som användare är medveten om det. Brewer menar att det historiska inslaget hos historiska ordböcker inte bara avser de äldre språkskeden som beskrivs, utan att ordböckerna själva också utgör historiska dokument. Det är därför viktigt att ge tillgång till äldre versioner av ordbokstexten och att vara tydlig med vad som är gammalt och vad som är nytt.

Vid revideringen av SAOB eftersträvas så stor tydlighet som möjligt i detta avseende. För det första kommer den första upplagan av ordboken att bevaras intakt, förstås i tryckt form, men även digitalt, vilket gör det möjligt att jämföra uppdaterade delar med den ursprungliga versionen. Redigeringsverktyget som används gör det också möjligt att spara och återskapa samtliga tidigare versioner av en artikel. På längre sikt går det därmed att publicera inte bara första och senaste versionen, utan även t.ex. långlivade tidigare utföranden längs vägen. Dessutom graderar redaktionen nivån av revidering för varje artikel som bearbetas; för tillfället används internt en skala från 1–3. I användargränssnittet kommer denna gradering att översättas till formuleringar av typen ”endast ytligt bearbetad”, ”delvis uppdaterad”, ”reviderad i sin helhet” el. dyl. I varje artikel kommer även fortsättningsvis ursprungligt publiceringsår att anges.

5. Sammanfattande avslutning

Att revidera ett stort historiskt ordboksverk vars första upplaga tog mer än 100 år att färdigställa är såklart inte en enkel uppgift. Målet med uppdateringen av SAOB är att modernisera ordboken och att öka balansen innehållsmässigt. Detta görs med olika metoder och angreppssätt, med tillägg av nytt och med automatiska och manuella uppdateringar av äldre innehåll. Men som visats verkar

det inte rimligt att sträva mot en ordbok helt i balans eller mot fullständig systematik i beskrivningarna, i stället bör riktmärket vara så systematiskt och konsekvent och så tydligt för användaren som möjligt. Revideringsprojektets tidsramar formar arbetet. Arbetsgången måste effektiviseras för att tydliga förbättringar ska bli möjliga att genomföra. Samtidigt bör hög kvalitet och noggrannhet fortfarande vara rättesnöret. Den allra mest överordnade utmaningen blir därmed för redaktionen att hinna uppdatera så mycket av ordbokens innehåll som är möjligt givet kraven på akribi och tidlig punktlighet. Samtidigt bör man inte blunda för det faktum att hög kvalitet kräver tid, och även om målet inom det pågående projektet är en tydligt förbättrad version under de givna förutsättningarna, så är förhoppningen att projektet på sikt får en fortsättning i någon form så att även kvarvarande brister kan ses över och åtgärdas.

Den reviderade versionen av SAOB kommer att innehålla både gamla och nya beskrivningar ihop och sida vid sida. För att hantera detta faktum krävs en viss finess. Läsaren måste kunna förstå vad i en artikel som har ändrats eller tillkommit och vad som ännu inte bearbetats. SAOB-redaktionen strävar i detta arbete efter transparens och tydlighet. Det ska för en ordboksanvändare vara möjligt att notera när en beskrivning först publicerades, i vilken utsträckning den har uppdaterats och hur den aktuella versionen skiljer sig ifrån originalversionen.

I den här artikeln har några viktiga aspekter av revideringsprojektet ringats in och diskuterats. Fortfarande befinner sig projektet i en inledande fas och ännu återstår mycket arbete att utföra, men redan nu upplever redaktionen sig ha kommit en god bit på väg. Varje liten förändring utgör ett steg mot en förbättrad version – och genom att identifiera och hantera stora utmaningar tidigt underlättas förhoppningsvis det efterföljande arbetet.

Litteratur

Ordböcker och grammatik

BOB = *Bokmålsordboka*. Språkrådet og Universitetet i Bergen.
<<https://ordbokene.no>> (mars 2025).

DWB = *Deutsches Wörterbuch, von Jacob Grimm und Wilhelm Grimm*, hrsg. von der Deutschen Akademie der Wissenschaften zu Berlin, Band 1–16 + Quellenverzeichnis, Leipzig (1852–1971).

NAOB = *Det Norske Akademis Ordbok*. Det Norske Akademi for Språk og Litteratur. <<https://naob.no>> (mars 2025).

NO = *Norsk Ordbok: Ordbok over det norske folkemålet og det nynorske skriftmålet* (1966–2015). Oslo: Det Norske Samlaget.

NOB = *Nynorskordboka*. Språkrådet og Universitetet i Bergen.
<<https://ordbokene.no>> (mars 2025).

Norsk Riksmålsordbok (1937–1957). Oslo: Aschehoug.

ODS = *Ordbog over det danske Sprog*, bind 1–28 (1918–1956). København: Det Danske Sprog- og Litteraturselskab.

OED = *The Oxford English Dictionary*, second edition, vol. I–XX (of the corrected reissue (1933) with an Introduction, Supplement, and Bibliography of *A New English Dictionary on Historical Principles*, vol. I–X, ed. by James A. H. Murray et al. (1884–1928) and Supplement, vol. I–IV, ed. by R. W. Burchfield (1972–1986)), Oxford 1989.

SAG = Teleman, Ulf, Staffan Hellberg & Erik Andersson. *Svenska Akademiens grammatik*, Band 1–4 (1999). Stockholm: Norstedts.

SAOB = *Svenska Akademiens ordbok* (1898–2023). Lund.

SAOL = *Svenska Akademiens ordlista* (2015). 14 uppl.

SO = *Svensk ordbok utgiven av Svenska Akademien* (2021). 2 uppl.

svenska.se = Svenska Akademiens ordboksportal <<https://svenska.se>> (mars 2025).

WNT = *Woordenboek der Nederlandsche Taal*, deel I–XXIX (1882–1998) + Bronnenlijst (1943, 1966) + Supplement I A–Azuursteen (1956). 's-Gravenhage & Leiden.

Annan litteratur

- Abel, Andrea (2012): Dictionary Writing Systems and Beyond. I: S. Granger & M. Paquot (eds.): *Electronic Lexicography*. Oxford: Oxford University Press. 83–106.
- Brewer, Charlotte (2022): The Future of Historical Dictionaries. I: Howard Jackson (ed.): *The Bloomsbury Handbook of Lexicography*. London: Bloomsbury Academic. 375–388.
- Elliott, L. & S. Williams (2006): Pasadena: A New Editing System for the Oxford English Dictionary. I: E. Conno, C. Marcilo & C. Onesti (eds.): *Proceedings of the XII EURALEX International Congress*. 257–64.
- Gilliver, Peter (2016): *The Making of the Oxford English Dictionary*. Oxford: Oxford University Press.
- Harm, Volker (2019): Between Science and Romanticism: The Deutsches Wörterbuch of the Brothers Grimm. I: S. J. Ogilvie-Thomson & G. Safran (eds.): *The Whole World in a Book: Dictionaries in the Nineteenth Century*. New York, NY: Oxford University Press. 73–92.
- Harm, Volker (2024): The Deutsches Wörterbuch as a Resource for the Investigation of Loan Vocabulary. I: Ph. Stöckle & S. Wah (eds.): *Lexicography and Language Variation*. Wien: Vienna University Press. 139–154.
- Jonsson, Hans (1983): Om historisk lexikografi av typ SAOB. I: *Nysvenska studier* 63, 127–158.
- Lauvstad, Hanne (2025): Kildegrunnlag og presentasjon i *Det Norske Akademis ordbok* (NAOB). I: *LexicoNordica* 32 (detta nummer).

- Moerdijk, A. (1994): Supplement. I: *Handleiding bij het Woordenboek der Nederlandsche Taal (WNT)*. 's-Gravenhage: Sdu Uitgeverij. 260–263.
- Moerdijk, A. (2001): *Inleiding uit de folioutgave van WNT, Aanvullingen, deel I*. <https://ivdnt.org/wp-content/uploads/2020/08/wnt_inleiding_aanvullingen_moerdijk.pdf> (mars 2023).
- Myking, Johan, Oddrun Grønvik & Stig Jarle Helset (2020): Netbrukaren i fokus: digitalisering og revisjon av Norsk Ordbok a–h. I: C. Sandström, U.-M. Forsberg, C. af Hällström-Reijonen, M. Lehtonen & K. Ruppel (red.): *Nordiska studier i lexicografi 15. Rapport från Konferensen om lexicografi i Norden*. Helsingfors: Nordiska föreningen för lexicografi. 235–244.
- Nilsson, Pär (2023): *Lexikal betydelseutveckling i teori och praktik: en analys av fem definitionsformler i Svenska Akademiens ordbok och de semantiska förändringsmekanismer som de beskriver*. PhD diss. Lunds universitet.
- Nilsson, Pär (2024a): Glancing back, looking forward: Swedish Academy Dictionary completed after 130 years – now time for revision. I: *Lexicographica* 40, 5–28.
- Nilsson, Pär (2024b): Report on the Revision of the Swedish Academy Dictionary – and the Search for “Old Neologisms. I: K. Despot, A. Ostroški Anić & I. Brač (eds.): *Lexicography and Semantics: Proceedings of the XXI EURALEX International Congress 8–12 October 2024 Cavtat, Croatia*. Cavtat: Institut za hrvatski jezik. 507–522.
- Nilsson, Pär & Bodil Rosqvist (2022): Nya tider, nya möjligheter: inför en reviderad version av SAOB i en helt ny tid. I: *Lexiconordica* 29, 53–72.
- Nilstun, Carina (2016): NAOB: På vei mot en nasjonalordbok for bokmålet: Nyordprosjektet ”2014 nyord i 2014” I: A. Gudiksen & H. Hovmark (red.): *Nordiske Studier i Leksikografi. Rapport fra 13. Konferanse om Leksikografi i Norden, København 19.–22. maj 2015*. København: Nordisk Forening for Leksikografi. 395–403.

- Pettersson, Gertrud (1996): *Svenska språket under sjuhundra år: en historia om svenskan och dess utforskande*. Lund: Studentlitteratur.
- Rosqvist, Bodil & Bo-A. Wendt (2017): Gammalt projekt och nya resurser – SAOB:s hantering av nytilkomna elektroniska textmängder. I: Å. Svavarsdóttir, H. Jónsdóttir, H. Hilmsdóttir & Þ. Úlfarsdóttir (red.): *Nordiske Studier i Leksikografi 14. Rapport fra Konference om Leksikografi i Norden*. Reykjavík: Nordisk Forening for Leksikografi. 214–221.
- Selback, Bente, Klara Sjo, Sturla Berg-Olsen & Knut E. Karlsen (2025): Revisjon av *Bokmålsordboka* og *Nynorskordboka*. I: *LexicoNordica* 32 (detta nummer).
- Svensén, Bo (2015): *Snille och smak. Svenska Akademien förr och nu*. Stockholm: Norstedts.
- Svenska Akademiens stadgar (1786). Digital resurs: <https://sa-admin.lb.se/assets/cfab13da-2fca-4353-aof3-8b51d27c3774.pdf/1._stadgar_for_svenska_akademien_2019_o.pdf> (mars 2023).
- Trap-Jensen, Lars (2005): Virtuelle perspektiver for ordbogsredigering: muligheder, strategier og virkelighedens begrænsning. I: *LexicoNordica* 12, 109–121.

Pär Nilsson
 Huvudredaktör, fil. dr
 Svenska Akademiens ordboksredaktion
 Dalbyvägen 3
 SE-224 60 Lund
 par.nilsson@svenskaakademien.se

Den Danske Begrebsordbog – opdatering, udvikling og tilgængeliggørelse

Sanni Nimb & Agnes Aggergaard Mikkelsen

We present *The Danish Thesaurus* (DDB) and the ways in which it is updated and distributed. A recently launched website targeted at researchers makes it possible to select a thesaurus section or chapter, combine the semantically ordered vocabulary with lexical information from *Den Danske Ordbog* (DDO) and *COR.SEM*, and download it for further studies. We describe how the thesaurus manuscript has been updated since the first edition and discuss three focal points: the addition of new words, changes in category designations, and increased focus on labelling.

1. Indledning

Den Danske Begrebsordbog (DDB, Nimb et al. 2014) er en omfattende tesaurus for moderne dansk hvis ordforråd og ordbetydninger bygger direkte på *Den Danske Ordbog* (DDO; 2003-05 og online), der dækker dansk sprog fra ca. 1950 til i dag. Både DDO og DDB udvides løbende med nye ord som en del af det leksikografiske arbejde i Det Danske Sprog- og Litteraturselskab (DSL), og i den seneste opdatering af DDO i marts 2025 indgik også de nyeste data fra DDB i elementet “Ord i nærheden”, idet begrebsordbogen danner grundlag for de grupper af synonymmer og nærsynonymer der vises.

Første udgave af DDB udkom på tryk i 2014. Emneinddelingen består af 22 kapitler og 888 afsnit der er inspireret af inddelingen i en tysk tesaurus, *Der Deutsche Wortschatz nach Fachgruppen* (Dornseiff 2004, herefter Dornseiff), samt en ældre dansk tesaurus, *Dansk Begrebsordbog* (Andersen 1945), der byggede på datidens udgave af samme tesaurus. Emneinddelingen og afsnits-

betegnelser blev i DDB både moderniseret og tilrettet danske samfundsforhold.

Den trykte udgave fra 2014 indeholdt 192.000 ord og udtryk fra DDO, heraf 112.000 unikke. De to tal indikerer at ét ord eller udtryk godt kan stå i flere afsnit, det er eksempelvis tilfældet med ordet *pølsesnak* der tæller som ét (unikt) ord og optræder i tre afsnit: [11.16 Sludder], [12.13 Intetsigende] og [12.27 Utydelig, uforståelig]. Siden da er manuskriptet til stadighed blevet suppleret med flere ord og betydninger fra DDO, herunder også med nye ord optaget i DDO efter 2014. Desuden er der løbende foretaget revideringer og moderniseringer af visse afsnit, men den grundlæggende kapitel- og afsnitsstruktur fra den trykte udgave er bevaret.

I 2024, 10 år efter førsteudgaven, udkom DDB som selvstændigt digitalt værk i form af en særudgave målrettet forskere (Nimb, Sørensen & Jensen 2024). Udgaven indeholder 15 % flere ord og udtryk end den trykte førsteudgave, nemlig i alt 222.000, heraf 124.000 unikke. Ved henvendelse til DSL (ddb@dsl.dk) og udlevering af password kan man som forsker få adgang til at søge efter ordforrådet i afsnit og kapitler efter eget valg og downloade det emneinddelte og semantisk ordnede indhold til videre sproglige undersøgelser og anvendelser, fx til undervisning eller inden for sprogteknologisk forskning.

På DSL har vi siden udgivelsen af den trykte DDB selv anvendt begrebsordbogens data i både leksikografisk arbejde og til sprogteknologisk forskning og udvikling. I denne artikel vil vi indlede med at beskrive de forskellige anvendelser inden vi præsenterer den nye forskningsudgave. I den afsluttende del af artiklen fokuserer vi på selve det redaktionelle arbejde med at udbygge DDB-manuskriptet. Vi vil gå i dybden med de mange suppleringer og revideringer af det leksikalske indhold der er foretaget i 2024-udgaven sammenlignet med den trykte udgave fra 2014, og bl.a. beskrive hvordan vi i langt højere grad end tidligere anvender sprogbrugsmarkører i DDB.

Det redaktionelle arbejde er ikke afsluttet, men fortsætter i takt med videreudviklingen af DDO. Målet er at de to værker følges ad, og at så godt som alle lemmaer og betydninger af substantiver, verber, adjektiver og adverbier i DDO er repræsenteret i DDB i mindst et afsnit. På længere sigt er det planen at udgive DDB i sin helhed endnu en gang som opslagsværk for ordbogsbrugere, enten i trykt form eller som en offentligt tilgængelig online begrebsordbog.

2. Begrebsordbøger over andre sprog

DDB's inddeling af det danske ordforråd i 22 kapitler og 888 afsnit er som nævnt inspireret af strukturen i 2004-udgaven af Dornseiff. Men generelt er det strukturen i den engelske tesaurus *Roget's Thesaurus* der har dannet udgangspunkt for tesaurusprojekter gennem årene. Dette værk, der udkom første gang i 1852, er inddelt i seks overordnede kategorier, indledt med abstrakte relationer, og under disse seks i alt 1.000 afsnit. I alle opdaterede udgaver er den originale klassificering fra førsteudgaven bevaret og kun ordforrådet udbygget (*Wikipedia*: *Roget's Thesaurus*, *Merriam-Webster*, omtækster: *Roget's Thesaurus*).

Den svenske tesaurus *Svenskt ordförråd ordnat i begreppsklasser* (Bring 1930) bygger fx på samtidens nyeste udgave af *Roget's Thesaurus*. I Zechner & Borin (2020) udtrykkes der et behov for en større, moderniseret og sprogteknologisk anvendelig begrebsordbog for svensk, og artiklen beskriver eksperimenter med at digitalisere *Svenskt ordförråd ordnat i begreppsklasser* og kombinere data med det sprogteknologiske leksikon *SALDO*, hvor betydninger af svenske lemmaer er koblet til hinanden i en wordnet-lignende struktur (Borin et al. 2013).

Også den senere norske tesaurus *Ord og begreper: Norsk tesaurus* (Rosbach 2001) er inspireret af *Roget's Thesaurus*, både hvad

angår inddeling og den semantisk betingede rækkefølge af ordene, men ordforrådet er langt mindre end det man finder i nutidens udgave af *Roget's Thesaurus*.

Den tyske thesaurus Dornseiff er ikke opbygget som *Roget's Thesaurus*. I stedet inddeles ordforrådet i 22 overordnede kapitler der starter med den konkrete verden (universet, naturen og levende væsener), og først derefter behandles abstrakte relationer. Den præsenterer derefter kapitler om handlinger, følelser, tænkning, socialt liv mv. før der til sidst afsluttes med økonomi, retssystem og religion. Inden for hvert kapitel optræder en række navngivne afsnit. Undergrupper i afsnittene indledes med et centralt ord der fungerer som en slags overskrift til gruppen. Ordene i gruppen er derefter alfabetiseret og ikke ordnet i semantisk rækkefølge (dvs. at to synonymer sandsynligvis står langt fra hinanden, fx hvis det ene starter med a og det andet med s). Dette er modsat *Roget's Thesaurus*, der udelukkende præsenterer ordforrådet i semantisk orden, også i gruppering af ordene på laveste niveau i et afsnit.

Dansk Begrebsordbog lod sig i sin tid inspirere af Dornseiff. Det samme gjorde redaktionen bag DDB ved projektets start. Inddelingen i 22 navngivne kapitler forekom umiddelbart tydeligere emneorienteret og velegnet som udgangspunkt end inddelingen i *Roget's Thesaurus*.

Den nyere engelske, videnskabeligt udarbejdede *The Historical Thesaurus for English*, der er baseret på *Oxford English Dictionary*, er interessant nok ikke baseret på kapitel- og afsnitsinddelingen i *Roget's Thesaurus*, hvor udgangspunktet som nævnt er abstrakte kategorier som relation og lighed. I stedet tager *The Historical Thesaurus for English* udgangspunkt først i den ydre verden (universet, naturen, levende væsener og deres behov), dernæst i den mentale verden (tanker, følelser) og endelig i den sociale verden (sociale grupperinger og systemer, lov, moral, kommunikation og fornøjelser) (se Alexander & Kay 2025). Inddelingen i den minder dermed meget mere om den inddeling man finder hos Dornseiff

og DDB. Men generelt er det som man i engelsk leksikografi forstår ved en *thesaurus*, blot grupper af synonymer og nært beslægtede ord der ikke er indordnet i en overordnet afsnits- og kapitelstruktur, dette gælder fx *Merriam-Webster Thesaurus* (se også *Merriam-Webster*, omtekster).

Som nævnt baserer DDB sig i hovedtræk på kapitel- og afsnitsstrukturen i Dornseiff. Den indeholder 888 nummererede og navngivne afsnit, fx [10.31 Vrede] og [17.20 Vintersport], indplaceret under 22 navngivne kapitler, fx [10. Følelser] og [17. Sport og fritid]. Dog svarer rækkefølgen af ord inden for de enkelte afsnit til den man finder i *Roget's Thesaurus*, idet ordene udelukkende præsenteres i semantisk orden. Når ordforrådet i et afsnit først er inddelt i hovedtyper som handling, egenskab, person, genstand mv., og et velegnet nøgleord til hver type er udpeget af redaktøren, tilstræbes det som et gennemgående redigeringsprincip at de mest centrale, prototypiske og neutrale ord inden for de enkelte typer altid står før de mindre centrale og sprogbrugs- eller frekvensmarkerede ord, og at synonymer altid står umiddelbart efter hinanden i undergruppen.

DDB opererer som den eneste af de omtalte begrebsordbøger med to niveauer af indledende (grafisk fremhævede) nøgleord til de enkelte undergrupper i et afsnit, et overordnet og et underordnet nøgleord, idet det overordnede har skopus over det underordnede, se figur 1. Både den semantiske rækkefølge og hierarkiet af nøgleord udnyttes i elementet "Ord i nærheden" i onlineudgaven af DDO, ligesom begge dele spiller en afgørende rolle i anvendelsen af DDB-data til evaluering af sprogmodeller, se afsnit 3.

The Historical Thesaurus for English er betydningsopmærket, idet alle ord er koblet til *Oxford English Dictionary's* onlineudgave, så man kan veksle mellem opslag i de to online ressourcer. På samme måde er alle ord i DDB-manuskriptet koblet til en betydning i DDO ved hjælp af et fælles underliggende id-nummer. Dette har muliggjort "Ord i nærheden"-funktionen i DDO, ligesom det har

drik, drikkelser, drikkevare, væske, nationaldrik, aftapning • forfriskning, tørstslukker • pjask ^{neds.}, sjask ^{neds.}, sprøjt ^{neds.}, bryg ^{neds.} • **sjat**, skvat, slat, bundskjuler, dråbe, stænk • tår, slurk, nip, sip, drag •

vand, postevand, gåsevin ^{spøg.}, isvand, drikkevand, brøndvand • vandkvalitet • kildevand, mineralvand, danskvand • danskvand, apollinaris ^{sj.}, karrierevand ^{spøg.}, mineralvand, kurvand® • **isterning**, isklump, is, knust is •

Figur 1: Der er to niveauer af nøgleord i DDB. Under hvert nøgleord afgrænser undergrupper (indledt af prikker) en række synonyme og/eller nærsynonyme ord. *Drik* og *vand* er begge overordnede nøgleord, og det første, *drik*, har skopus over alle undergrupper frem til det andet, *vand*. *Sjat* er underordnet under *drik* og har skopus over de to undergrupper frem til det næste nøgleord (uanset type), i dette tilfælde det overordnede nøgleord *vand* (dvs. skopus over ordene *skvat*, *slat*, *bundskjuler*, *dråbe*, *stænk* og også *tår*, *slurk*, *nip*, *sip*, *drag*).

været forudsætningen for de anvendelser af data vi vil beskrive nærmere i afsnit 3.

3. Sprogteknologiske og leksikografiske anvendelser af DDB-data

DDB indgår i dag i et større netværk af ordbøger og semantiske ressourcer der alle er koblet til id-nummeret i DDO. Netværket er udviklet i samarbejde med Center for Sprogteknologi, Københavns Universitet, indledt med etableringen af *DanNet*, et dansk WordNet baseret på DDO's ordforråd og betydningsinventar (Pedersen et al. 2009). DDB-arbejdet tog efterfølgende udgangspunkt i *DanNets* betydningshierarkier inden for emner som fx møbler, tøj og sygdomme. I DDB blev ordforrådet indplaceret i et afsnit, sorteret i indbyrdes semantisk rækkefølge og suppleret med endnu flere ord fra DDO.

Efterfølgende kunne sammenkoblingen mellem DDB og DDO, mere specifikt koblingen mellem verber og verbalsubstantiver fra afsnit i DDB og valensmønstre for samme ord i DDO, udnyttes til at udvikle et framenet-leksikon for dansk på kort tid (*Det Danske FrameNet-leksikon*, Nimb et al. 2017). Semantisk beslægtede ord kunne med fordel behandles samtidig, og valg af semantisk ramme i *Berkeley FrameNet* blev understøttet af information om valensmønstret.

Endnu en semantisk ressource blev udviklet på basis af DDB, nemlig *Det Danske Sentimentleksikon* (Nimb et al. 2022). En opmærkning af negativt, henholdsvis positivt ladede afsnitsnavne i DDB (fx [10.31 Vrede], henholdsvis [10.20 Begejstring, glæde, lykke] i kapitel [10. Følelser]) kombineret med ord med valøroplevelsningen “nedsættende” i DDO muliggjorde en effektiv opmærkningsmetode hvor ordforråd først blev automatisk forhåndsopmærket og efterfølgende manuelt valideret. Informationen om polaritet blev også integreret i *DanNet*, der i samme projekt blev udvidet semiautomatisk med store grupper af adjektiver, også baseret på DDB’s semantiske inddeling (Nimb et al. 2021).

Den seneste ordbogsressource der er udviklet på baggrund af udtræk af leksikalsk information fra alle de forskellige ordbøger i netværket, er det sprogteknologiske leksikon *COR.SEM* (Nimb et al. 2024). DDB’s 17.500 nøgleord, der repræsenterer 10.500 forskellige lemmer, dannede bl.a. basis for lemmaselektionen så man sikrede sig at centrale ord inden for alle emner var blandt *COR.SEM*’s i alt 34.000 lemmer.

Den seneste sprogteknologiske udnyttelse af DDB er som datagrundlag for udviklingen af datasæt til automatisk evaluering af kvaliteten af sprogmodellers håndtering af dansk (Pedersen et al. 2024a og b). Her er både information om kapitel-, afsnits- og undergruppestruktur samt over- og underordnet nøgleord relevant, og sammenkoblingen på betydningsniveau mellem DDB og DDO udnyttes også. Sprogmodellerne testes i deres evne til fx at

udpege synonymet til et givet inputord blandt fire forslag. Et eksempel er ordgruppen [*søluft*, *anticyklon*, *klimabælte*, *nyvurdering*] hvor sprogmodellen får inputordet *havluft* og bliver bedt om at udpege synonymet blandt de fire ord. Fra oplysningen om *havlufts* synonym i DDO ved vi at det korrekte svar er *søluft*, og fordi to af de øvrige tre ord er udtrukket fra samme afsnit ([1.5 Vejr, luft]) i DDB (*anticyklon* og *klimabælte*), sikres det at spørgsmålet har en vis sværhedsgrad.

Sprogmodellen bliver også testet i sin evne til at udpege det betydningsmæssigt afvigende ord blandt de fire, og her ved vi fra strukturen i DDB at det rigtige svar er *nyvurdering* fordi ordet står i et helt andet kapitel ([15. Socialt liv]) end de andre tre ([1. Natur og miljø]).

I arbejdet med praktisk ordbogsudgivelse på DSL har sammenkoblingen mellem DDB og DDO som allerede nævnt muliggjort visning af grupper af nærsynonymer for ordbogsbrugerne (Nimb et al. 2018). I den nyeste onlineudgave af DDO (marts 2025) kan nærsynonymer ses ved 90 % af lemmaer og betydninger på grund af udvidelsen af DDB-manuskriptet siden den trykte udgave. Når den pågældende DDO-betydning er repræsenteret under flere forskellige emner i DDB, vises der mere end én boks med nærsynonymer, se eksemplet *nordkyst* i figur 2, der både viser ord fra afsnittet [1.13 Verdenshjørner] og fra afsnittet [1.17 Kystområder, ø].

4. Forskningsressourcen webddb.dsl.dk

Med udgivelsen af forskningsressourcen webddb.dsl.dk, finansieret af Carlsbergfondet (Nimb, Sørensen & Jensen 2024), bliver det nu muligt også for udefrakommende forskere at arbejde med struktur og ordforråd i DDB og få gavn af sammenkoblingen mellem ord og betydninger i netværket af ordbøger. Hjemmesiden præsenterer 90.472 lemmaer fra DDO, og 130.827 forskellige

nordkyst substantiv, fælleskøn
 BØJNING -en, -er, -erne
 UDTALE  

Betydninger

kyst der har havet på sin nordlige side

ORD I NÆRHEDEN

VERDENSHJØRNER
nordside, nordfløj, nordende • nordgrænse, **nordkyst** •
 nordenvind, nordlig vind
 fra Den Danske Begrebsordbog, kapitel.1

KYSTOMRÅDE, Ø
kyst, kystzone, kystland, kystområde • landside,
 vandside • forreven kyst, klippekyst • kyststrækning,
 riviera, **nordkyst**, sydkyst, vestkyst, østkyst •
 vadehavskyst, østersøkyst • svenskekyst •
 middelhavskyst, solkyst, den spanske solkyst, den franske
 riviera • stillehavskyst
 fra Den Danske Begrebsordbog, kapitel.1

Figur 2: “Ord i nærheden” ved opslag på *nordkyst* i DDO.

DDO-betydninger er nævnt i et eller flere afsnit. I alt omfatter ressourcen 221.445 ord og udtryk, hvoraf 124.336 er unikke. Desuden er det gjort muligt at tilkoble en lang række udvalgte oplysninger fra de øvrige ordbogsressourcer i netværket af ordbøger. Dermed faciliteres semantisk baserede undersøgelser af det danske ordforråd. Når et bestemt navngivet afsnit (eller et fuldt kapitel) er udvalgt, kan man fx tilføje etymologi, valens, orddannelse, grad af polysemi og/eller oplysning om metaforisk betydning fra DDO. Man kan også supplere med information om ontologisk type, polaritet og semantisk ramme fra *COR.SEM*. Data kan efterfølgende downloades i et regneark og anvendes til egne undersøgelser, opmærkning og sprogteknologiske eksperimenter.

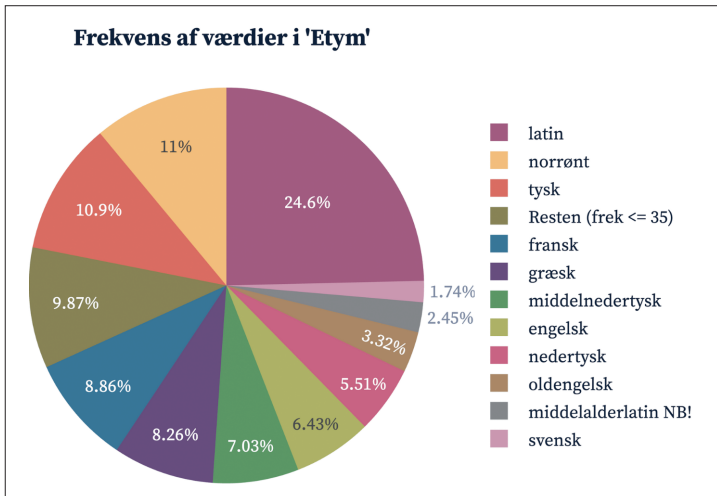
Hvis man fx er interesseret i at studere metaforer i dansk, kan man udvælge et afsnit og undersøge om der er overførte betydninger blandt ordene, og om de udgør et mønster fordi de stammer fra samme semantiske felt. I figur 3 ses fx ordforråd fra af-

Vælg kapitel og/eller afsnit Vælg kapitel 09 Vilje og handling Vælg afsnit 09.050 Svær, vanskelig Vis data Ryd alt Dataværker Vælg Ordklasse Se muligheder Vælg Nøgleord 1 Se muligheder Vælg Nøgleord 2 Se muligheder Tiltvælg flere kolonner Vælg kolonner Overført		grums	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		støj	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		snirkel	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		flimmer	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		mudder	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		uforståelighed	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		sværhedsgrad	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		en hel videnskab	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		hovedbrud	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		gåde	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		mysterium	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		rebus	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		labyrint	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		puslespil	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		kabale	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		irgang	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		spejlkabinet	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		komplikation	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		stof til eftertanke	09 Vilje og handling	09.050 Svær, vanskelig	sværhed
		afgrænsningsproblem	09 Vilje og handling	09.050 Svær, vanskelig	sværhed

Figur 3: På hjemmesiden webddb.dsl.dk kan man udtrække ordforråd fra afsnit [9.50 Svær, vanskelig] og supplere det med oplysning om hvorvidt betydningen er markeret som “overført” i DDO.

snit [9.50 Svær, vanskelig] suppleret med oplysningen “overført” i DDO. Det ses fx at spil der kræver tankevirksomhed (*gåde, rebus, labyrint, puslespil, kabale*), sammen indgår i et metaforisk mønster og repræsenterer ‘noget der er svært eller vanskeligt’.

Hvis man i stedet er interesseret i at undersøge hvilke sprog der særligt har påvirket dansk inden for et bestemt semantisk eller tematisk område, er det muligt at supplere ord i et afsnit eller kapitel med etymologiske oplysninger fra DDO. I figur 4 ses fx fordelingen af lånsprog (eller sprog med lignende etymologi) der nævnes i DDO ved betydninger i det samlede kapitel [11. Tænkning]. Det fremgår at næsten $\frac{1}{4}$ af de ord i kapitlet der indeholder en etymologisk oplysning i DDO, stammer fra latin. Til sammenligning vil den tilsvarende undersøgelse for kapitel [10. Følelser] vise at det i dette tilfælde kun drejer sig om 15 % af ordene. På samme måde viser en undersøgelse af afsnit [19.18 Computer] (ikke over-



Figur 4: Kapitel [11. Tænkning]: 24,6 % af ordene er beskrevet i DDO med en etymologisk oplysning der indeholder sprogkategorien latin.

raskende) at størstedelen af ordene, nemlig 55 %, er indlånt fra engelsk.

En begrebsordbog afspejler sin samtid mere direkte end en betydningsordbog ud fra de kategoriseringer der udgør de enkelte afsnit, og ud fra de ord der fremhæves som centrale (og neutrale) kategorioverskrifter i et afsnit. Selvom selve kapitel- og afsnitsstrukturen i det store og hele er bevaret i webddb.dsl.dk siden værket kom på tryk i 2014 (jf. også redigeringsprincippet i de løbende udgivelser af *Roget's Thesaurus*), er manuskriptet i den tid der er gået, blevet udsat for et omfattende revideringsarbejde hvad angår ordforråd og kategorisering inden for de enkelte afsnit. Mange nye ord er kommet til i dansk i løbet af de 10 år, og der er kommet et større fokus på diskriminerende og politisk ukorrekt sprogbrug i samfundet. I næste afsnit peger vi på hvilke særlige udfordringer

det medfører i redaktionsarbejdet. Vi vil komme ind på nogle af de suppleringer der er foretaget, herunder også hvordan arbejdet med at angive valøroplysninger er grebet an på en ny måde.

5. Ændringer og tilføjelser siden 1. udgave 2014

5.1. Flere ord

Siden førsteudgaven udkom, er DDB vokset med mere end 25.000 ord og udtryk. Det gælder både ord der er integreret i DDO i samme periode, "gamle" ord fra DDO der ikke kom med i førsteudgaven af DDB, og ord der endnu ikke er beskrevet i DDO. Sidstnævnte tilføjes i DDB med samme krav til korpusfrekvens som anvendes i DDO-redigeringen.

Så godt som alle afsnit har fået nye tilføjelser, men særligt de afsnit der dækker samfundsaktuelle temaer, har fået mange nye ord og udtryk. Det gælder fx afsnittene [1.20 Økologi, miljøbeskyttelse], [2.30 Graviditet], [2.38 Psykiske lidelser] og [18.19 Ligestilling, diskrimination], se tabel 1 for eksempler.

Afsnitsnavn	Eksempler på tilføjelser
1.20 Økologi, miljøbeskyttelse	<i>klimaforandring, bæredygtighed, biodiversitet, vild natur, miljøansvarlighed</i>
2.30 Graviditet	<i>morgenkvalme, tegnblødning, fertilitetsbehandling, donorregister, rugemoderskab</i>
2.38 Psykiske lidelser	<i>mistrivsel, social angst, personlighedsforstyrrelse, selvskade, autismspektrum</i>
18.19 Ligestilling, diskrimination	<i>kønsbalance, hyggeracisme, kvindejob, heteronormativitet, pride</i>

Tabel 1: Eksempler på tilføjelser i fire afsnit i DDB.

5.2 Ændringer i kategorisering

For at bevare deres aktualitet må opslagsværker følge med tiden, og det gælder også for DDB – ikke alene i forhold til ordforråd, men især også i forhold til kategorisering, som er centralt i en begrebsordbog. Kigger man på tidligere tiders begrebsordbøger, er det tydeligt at hvad der passerede ubemærket som afsnitsnavne før i tiden, ikke er gangbare kategoriseringer af verden i dag. Fra *Dansk Begrebsordbog* kan fx nævnes afsnitsnavne som “Grimhed – Mishag”, “Vanvid”, “Konkubinat” og “Skøge”. I førsteudgaven af DDB var det nødvendigt at foretage mange ændringer i forhold til disse afsnitsbetegnelser. Et ældre afsnit med titlen “sprogfejl” blev fx ændret til “sprognormering” ([12.26 Sprognormering]) i DDB, og ord inden for emnet arbejdsløshed blev placeret i et nyt selvstændigt afsnit [20.5 Arbejdsløshed] fremfor som i *Dansk Begrebsordbog* at blive nævnt i selskab med ord for ‘dovenskab’: “*Arbejdsløshed, Dorskhed, Dovenskab, Driveri, .., Ledighed, .. Ugedelighed*” (i afsnittet “Uvirksomhed”, der i DDB har fået titlen [9.38 Dovenskab]).

En række nye afsnit blev oprettet, fx [18.16 Samfundsdebat], [18.12 Institution], [2.26 Menneskets udseende], [2.31 Prævention og abort] samt [2.38 Psykiske lidelser]. Tager man netop samfundets opfattelse af psykisk sygdom som eksempel, er det interessant at se under hvilke afsnit lemmaer som *depression* og *psykose* optræder i tidligere begrebsordbøger: *Depression* findes i afsnittene “Tungsind” (*Dansk Begrebsordbog*), “Misnøye” og “Nedslåthed” (*Ord og begreper: Norsk tesaurus*), “Krankheit” og “Trübsinn” (Dornseiff), og *psykose*/*Psychose* findes under “Vanvid” (*Dansk Begrebsordbog*), “Galskap” (*Ord og begreper: Norsk tesaurus*), “Krankheit” og “Verrückt” (Dornseiff). Til sammenligning findes *depression* og *psykose* under [2.38 Psykiske lidelser] i DDB – her ser man altså for første gang psykisk sygdom kategoriseret som netop sin egen kategori. Eksemplet viser at kategoriseringer

afspejler samtiden meget direkte, og måske derfor har det været nødvendigt at ændre en række nøgleord og afsnitsbetegnelser i DDB der allerede efter 10 år forekom utidssvarende. Her skal man også tage i betragtning at der siden 2014 er kommet et større fokus på diskriminerende og politisk ukorrekt sprogbrug i samfundet samt en øget opmærksomhed på forholdet mellem sprog og magt, se fx Peterson & Sköldberg (2020-2021). Endelig baserer strukturen, og dermed også kategoriseringerne, i DDB sig på ældre begrebsordbøgers struktur (se afsnit 2), og selvom der blev foretaget mange ændringer i første udgave af DDB, har der stadig været et behov for ændring af visse tidligere betegnelser.

Eksempler på nøgleord vi har ændret siden førsteudgaven, er fx *den hvide race* og *den sorte race* i afsnit [15.2 Fremmede, udlandet] som nu i stedet står under nøgleordene *hvid*, henholdsvis *sort*, og som nu er markeret med sprogbrugsmarkøren “gl.” for ‘gammeldags’, se figur 5 og 6.

nesker • **de fremmede**, andre folkeslag, fremmede folkeslag, udenlandske statsborgere • **den hvide race**, den hvide mand, den ariske race, gringo^{uform.}, blegansigt^{uform.} • albino • **den sorte race**, sort, neger^{neds.}, nigger^{neds.} • **ikkehvid**,

Figur 5: I den trykte 2014-udgave af DDB optræder *den hvide race* og *den sorte race* som nøgleord i afsnit [15.2 Fremmede, udlandet].

INDBYGGER, BEFOLKNING

hvid², den hvide mand, gringo^{uform.}, blegansigt^{uform.},
kartoffel, yankee^{uform.} • den hvide race^{gl.}, arier^{gl.} •
albino

fra Den Danske Begrebsordbog, kapitel 15

FREMMEDE, UDLANDET

sort, den sorte race^{gl.}, neger^{neds.}, nigger^{neds.}, farvet,
mulat • albino

fra Den Danske Begrebsordbog, kapitel 15

Figur 6: I den nyeste udgave af DDB (2025) er *den hvide race* og *den sorte race* ikke selv nøgleord i afsnit [15.2 Fremmede, udlandet], men i stedet anført under nøgleordene *hvid* og *sort* samt forsynet med sprogbrugsmarkøren “gl.” for ‘gammeldags’.

Et andet eksempel er at nøgleordet *smuk pige* i afsnit [2.15 Kvinde] i det reviderede manuskript er erstattet af *smuk kvinde*. I afsnit [2.38 Psykiske lidelser] er nøgleordene *miste forstanden* og *tvangs-indlægge* rykket et niveau ned, så de nu figurerer som nøgleord på niveauet under de nye, overordnede nøgleord *være psykisk syg* og *behandle*. I afsnittet [18.19 Ligestilling, diskrimination] har de mange nye tilføjelser medvirket til at ordene *racediskrimination* og *kønsdiskrimination* nu er fremhævet som nøgleord, hvor de før blot stod under nøgleordet *diskrimination*.

Manuskriptet er i høj grad dynamisk, og tilføjelsen af nye ord og udtryk samt udviklinger i synet på kategorier giver ofte anledning til ændringer. Således også med selve navnene på afsnittene, og et par eksempler på det ses i tabel 2.

	DDB 2014	DDB 2025
2.29	Sex og begær	Seksualitet og begær
15.6	Ugift, enlig	Enlig, single
18.19	Ligestilling	Ligestilling, diskrimination

Tabel 2: Eksempler på ændringer af afsnitsnavne i DDB.

5.3 Sprogbrugsmarkører

Vi har også haft behov for at ændre den måde vi arbejder med sprogbrugsmarkører på i DDB, altså om et ord markeres som fx “uformelt”, “nedsættende”, “sjældent” etc. Det skyldes både samfundets ændrede syn på kontroversielle ord i forlængelse af den øgede opmærksomhed på forholdet mellem sprog og magt, men også den måde som DDB nu udkommer på. Førsteudgaven kom på tryk, mens de nye, opdaterede udgaver som nævnt lever på nettet, primært i form af elementet “Ord i nærheden” i DDO. Dertil kommer at sprogbrugsmarkering er et område af leksikografien hvor der er “more work to be done” (Atkins & Rundell 2005:405), idet det er “vanskeligt at have en systematisk og for brugeren

gennemskuelig praksis med sprogbrugsmarkering” (Mikkelsen 2024:15).

I forordet til DDB (2014) kan man blandt andet læse følgende om sprogbrugsmarkering:

“For eksempel har et uformelt ord i en gruppe af i øvrigt neutrale ord typisk bevaret sprogbrugsmarkøren fra *Den Danske Ordbog*, hvorimod en større gruppe af udelukkende uformelle ord og udtryk ikke er forsynet med kommentarer.” (DDB 2014:9)

Da førsteudgaven blev til, arbejdede vi altså ud fra reglen om at en gruppe ord med samme valør ikke alle skulle markeres, mens et enkelt ord der med sin valør skilte sig ud fra de øvrige i en gruppe, skulle markeres. Dette skyldtes både behovet for at spare plads i den trykte bog, men også ønsket om ikke at gentage den samme sprogbrugsmarkør for hvert ord i en gruppe.

Som følge af en del brugerhenvendelser har vi imidlertid indset at vi ved grupper af ord der fx oftest bruges nedsættende, ikke kan antage at “sprogbrugen i disse tilfælde kan udledes af den samlede gruppes eller endda hele afsnittets betydning” (DDB 2014:9). Der er behov for tydeligt at indikere hvilke konnotationer der er på spil. Det ses fx i figur 7, et udsnit af afsnit [10.58 Ikke bryde sig om, antipati], hvor en lang række ord er forsynet med samme sprogbrugsmarkør. Denne nye praksis, der kan ses i visningen “Ord i nærheden” på ordnet.dk/ddo (men ikke er en del af data i webddb.dsl.dk), er måske ikke videre elegant, men den har vist sig nødvendig, ikke mindst fordi visningen “Ord i nærheden” kun viser mindre udsnit af afsnittenes indhold.

En anden ændring i arbejdet med sprogbrugsmarkører er at vi har bevæget os væk fra at følge sprogbrugsmarkeringen i DDO én til én. En betydningsordbog som DDO har flere muligheder for at beskrive konnotativ betydning end en begrebsordbog har, fx ved

hadeobjekt, yndlingsaversion, idiosynkrasi, pestilens, pinagtighed, misfoster, lort ^{uformelt}, bras, skidt, juks, møg ^{uformelt}, gøgemøg ^{uformelt}, noget værre lort, noget værre møg, noget skideværk, noget hø, vanskabning, pestbyld, byld, en hørn af .. • lortefilm ^{uformelt}, lortejob, lorteland • lortevejr ^{uformelt}, møgvejr ^{uformelt} • **ragelse**, gammelt lort, billigt lort, billigt skidt, en masse skidt, hakkelse, hø og hakkelse • øjebæ ^{uformelt}, ørebæ ^{uformelt} • **idiot** ^{neds.}, tåbe ^{neds.}, narrov ^{neds.}, kæmpeidiot ^{neds.}, skid ^{neds.}, dum skid ^{neds.}, kold skid ^{neds.} • paphoved ^{neds.}, båtnakke ^{neds.}, nørð ^{neds.} • læs ^{neds.}, store læs ^{neds.}, brilleabe ^{neds.}, skidespræller ^{neds.}, fjollehoved ^{neds.}, snøbel ^{neds.}, bavian ^{neds.}, bondetamp ^{neds.}, jysk tranlampe ^{neds.}, haleneger ^{neds.}, abe ^{neds.} • gnom ^{neds.}, noksagt ^{neds.}, en køn plante ^{neds.} • lus ^{neds.}, maddike ^{neds.}, mide ^{neds.}, fnatmide ^{neds.}, orm ^{neds.} • røvpuler ^{neds.} • trans ^{neds.} • stodder ^{neds.}, koter ^{neds.}, møghund ^{neds.}, møgdyr ^{neds.}, skarn ^{neds.}, gespenst ^{neds.}, utyske ^{neds.}, afskum ^{neds.}, udskud ^{neds.}, stymper ^{gl.}, skumpelskud ^{gl.}, pikansjos ^{neds.} • dum kælling ^{neds.}, bitch ^{neds.}, møgkælling ^{neds.}, kælling ^{neds.}, møgludder ^{neds.}, møgfisse ^{neds.}, møgso ^{neds.}, gimpe ^{neds.}, mæfikke ^{neds.}, taske ^{neds.}, hore ^{neds.}, ludder ^{neds.} • **møgunge** ^{neds.}, snotunge ^{neds.}, snottede unge, snotas, pattebarn • **kravl**, kryb, krapyl, koter, møghund •

Figur 7: Udsnit af *Den Danske Begrebsordbog* (2025) fra afsnit [10.58 Ikke bryde sig om, antipati], hvor en lang række ord er forsynet med sprogbrugsmarkøren “neds.” for ‘nedsættende’.

at inddrage selve betydningsbeskrivelsen, ved at tilføje en længere sprogbrugskommentar til et lemma eller ved at afspejle forskellige betydningsnuancer i brugseksempler. Her har DDB kun valgt mellem enten at bringe eller ikke at bringe en sprogbrugsmarkør, og derfor vil der somme tider være behov for at afvige fra DDO. Det ses fx ved lemmaet *eskimo*, som i DDO er forsynet med sprogbrugskommentaren “kan virke stødende, idet betegnelsen *inuit* nu foretrækkes af disse folkeslag frem for *eskimo*”, mens det i DDB blot markeres som “nedsættende”. Arbejdet med sprogbrugsmarkører er en del af et større kompleks i forbindelse med behandlingen af kontroversielle ord, et emne der er leveringsdygtigt i holdninger, debat og udfordringer på samfundsniveau såvel som på ordbogsredaktionerne (Mikkelsen 2024, Nimb 2024, Petersson & Sköldberg 2020-2021). “Ordene er svære at indkredse i ordbog og korpus, og det er svært at tildele sprogbrugsmarkører fordi der kan være stor variation i holdningen til ordene blandt både sprogbrugere og ordbogsredaktører” (Nimb 2024: 170), og netop derfor har vi på redaktionen af DDB et fortsat fokus på at lokalisere kontroversielle og potentielt kontroversielle ord, ligesom vi også

ønsker at inddrage redaktører med forskellig alder og køn i beskrivelsen af dem (se også Nimb 2024).

6. Konklusion

DDB har siden værket udkom første gang i 2014, vist sig anvendeligt ikke kun som selvstændigt opslagsværk for ordbogsbrugere, men også som datagrundlag for både visningen af nærsynonyme ord i DDO og for udviklingen af en række sprogteknologiske ordbogsressourcer. Senest har de leksikalske data også vist sig anvendelige inden for et helt nyt sprogteknologisk område, nemlig i arbejdet med at udvikle datasæt til automatisk evaluering af sprogmodellers formåen på dansk.

Med udgivelsen af hjemmesiden webddb.dsl.dk får forsknings-samfundet også mulighed for at anvende de leksikalske data i sprogteknologiske sammenhænge. Hjemmesiden giver endvidere mulighed for at foretage nye former for leksikografiske og lingvistiske undersøgelser af det danske ordforråd der tager udgangspunkt i det semantiske og tematiske slægtskab mellem ord.

DDB er vokset betydeligt siden værket udkom første gang, idet nye ord og udtryk løbende indsættes fra DDO. Vi har også beskrevet hvordan vi foretager nødvendige opdateringer af både kategoribetegnelser og sprogbrugsmarkører i takt med den sproglige udvikling og holdningsændringer i sprogsamfundet.

Litteratur

Ordbøger, korpusser og digitale resurser

Berkeley FrameNet = <<https://framenet.icsi.berkeley.edu>> (marts 2025).

- COR.SEM = Det Centrale Ordregister – Semantisk Modul. Det Danske Sprog- og Litteraturselskab og Center for Sprogteknologi, Københavns Universitet. <<https://corsem.dsl.dk>> og <<https://ordregister.dk>>.
- DanNet = DanNet. Center for Sprogteknologi, Københavns Universitet, og Det Danske Sprog- og Litteraturselskab. <<https://wordnet.dk>> (marts 2025).
- Dansk Begrebsordbog* (1945). Harry Andersen. København: Munksgaard.
- DDB (2014) = Sanni Nimb, Henrik Lorentzen, Liisa Theilgaard & Thomas Troelsgård: *Den Danske Begrebsordbog*. København/Odense: Det Danske Sprog- og Litteraturselskab og Syddansk Universitetsforlag.
- DDO (2003-05) = *Den Danske Ordbog*. København: Det Danske Sprog- og Litteraturselskab og Gyldendal.
- DDO online = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. <<https://ordnet.dk/ddo>> (marts 2025).
- Det Danske FrameNet-leksikon* (2018). Det Danske Sprog- og Litteraturselskab og Københavns Universitet. <<https://korpus.dsl.dk/resources/details/framenet.html>> (marts 2025).
- Det Danske Sentimentleksikon* (2021). Det Danske Sprog- og Litteraturselskab og Center for Sprogteknologi, Københavns Universitet. <<http://github.com/dsldk/danish-sentiment-lexicon>> (marts 2025).
- Dornseiff = Franz Dornseiff (2004): *Der deutsche Wortschatz nach Sachgruppen*, 8. Auflage von Uwe Quasthoff. Berlin/New York: Walter de Gruyter.
- Merriam-Webster Thesaurus* = *Merriam-Webster Online Thesaurus* (2024). Merriam Webster, Incorporated. <<https://www.merriam-webster.com/thesaurus>> (marts 2025).
- Ord og begreper: Norsk tesaurus* (2001). Johan Hammond Rosbach. Oslo: Pax.
- Oxford English Dictionary* (2025). Oxford University Press. <<https://www.oed.com>> (marts 2025).

- Roget's Thesaurus* = Peter Mark Roget: *Roget's Thesaurus*, 150th anniversary edition edited by George Davidson 2002. London: Penguin.
- SALDO = Lars Borin, Lennart Lönngren & Markus Forsberg (2017). SALDO (updated: 2017-09-19). [Datasæt]. Språkbanken Text. <<https://spraakbanken.gu.se/en/resources/saldo>>. <<https://doi.org/10.23695/s8ow-2517>>.
- Svenskt ordförråd ordnat i begreppsklasser* (1930). Sven Casper Bring. Stockholm: Hugo Gebers Forlag.
- The Historical Thesaurus for English* = Kay, Christian, Marc Alexander, Fraser Dallachy, Jane Roberts, Michael Samuels & Irené Wotherspoon (eds.): *The Historical Thesaurus of English* (2nd ed., version 5.0). Glasgow: University of Glasgow. <<http://ht.ac.uk>> (marts 2025).

Anden litteratur

- Alexander, Marc & Christian Kay (2025): Classification. I: *About the Historical Thesaurus of English*. Glasgow: University of Glasgow. <<https://ht.ac.uk/classification>> (marts 2025).
- Atkins, B. T. Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Borin, Lars, Markus Forsberg & Lennart Lönngren (2013): SALDO: a touch of yin to WordNet's yang. I: *Language Resources and Evaluation* 47(4), 1191-1211. <<https://doi.org/10.1007/s10579-013-9233-4>>.
- Merriam-Webster, omtækster: <<https://www.merriam-webster.com/wordplay/rogets-thesaurus>> (marts 2025).
- Mikkelsen, Agnes Aggergaard (2024): *Jargon blandt leksikografer. En undersøgelse af sprogbrugsmarkering i danske betydningsordbøger*. Upubliceret speciale i lingvistik, Københavns Universitet, Institut for Nordiske Studier og Sprogvidenskab. Fremsendes efter ønske.

- Nimb, Sanni, Anna Braasch, Sussi Olsen, Bolette Sandford Pedersen & Anders Søgaard (2017): From Thesaurus to FrameNet. I: Iztok Kosem, Jelena Kallas, Carole Tiberius, Simon Krek, Miloš Jakubiček & Vít Baisa (eds.): *Electronic Lexicography in the 21st century: Proceedings of eLex 2017 conference*, Lexical Computing CZ s.r.o., Brno, Tjekkiet. 1-22. <https://ellex.link/ellex2017/proceedings/eLex_2017_Proceedings.pdf>.
- Nimb, Sanni, Nikolai Hartvig Sørensen & Thomas Troelsgård (2018): From Standalone Thesaurus to Integrated Related Words in The Danish Dictionary. I: Jaka Čibej, Vojko Gorjanc, Iztok Kosem & Simon Krek (eds.): *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts*. Ljubljana: Ljubljana University Press, Faculty of Arts. 916-923. <<https://euralex.org/publications/from-standalone-thesaurus-to-integrated-related-words-in-the-danish-dictionary>>.
- Nimb, Sanni, Bolette S. Pedersen & Sussi Olsen (2021): DanNet2: Extending the coverage of adjectives in DanNet based on thesaurus data (project presentation). I: Piek Vossen & Christiane Fellbaum (eds.): *Proceedings of the 11th Global Wordnet Conference*, University of South Africa (UNISA), Global Wordnet Association. 267-272. <<https://aclanthology.org/2021.gwc-1.31>>.
- Nimb, Sanni, Sussi Olsen, Bolette S. Pedersen & Thomas Troelsgård (2022): A Thesaurus-based Sentiment Lexicon for Danish: The Danish Sentiment Lexicon. I: Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Hélène Mazo, Jan Odijk & Stelios Piperidis (eds.): *Proceedings of the Thirteenth Language Resources and Evaluation Conference: LREC2022*. Marseille: European Language Resources Association. 2826-2832. <<https://aclanthology.org/2022.lrec-1.302>>.

- Nimb, Sanni (2024): Fra ‘sandheden om sproget’ til et opgør med stereotyper: ikke-korpusbaserede metoder til leksikografiske beskrivelser af kontroversielle ord. I: *LexicoNordica* 31, 151-174. <<https://doi.org/10.7146/ln.v1i31.151289>>.
- Nimb, Sanni, Ida Flörke, Sussi Olsen, Bolette S. Pedersen & Nathalie C. H. Sørensen (2024): COR.SEM, a New Formal Semantic Lexicon for Danish. I: Kristina Š. Despot, Ana Ostroški Anić & Ivana Brač (eds.): *Lexicography and Semantics – Proceedings of the XXI EURALEX International Congress*, 8–12 October 2024, Cavtat, Croatia. 141-153. <<https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralax-XXI-final-web.pdf>>.
- Nimb, Sanni, Nathalie C. H. Sørensen & Jonas Jensen (2024): Making Danish Thesaurus Data Available to Researchers – The WebDDB project. I: Kristina Š. Despot, Ana Ostroški Anić & Ivana Brač (eds.): *Lexicography and Semantics – Proceedings of the XXI EURALEX International Congress*, 8–12 October 2024, Cavtat, Croatia. 523-529. <<https://euralex.jezik.hr/wp-content/uploads/2021/09/Euralax-XXI-final-web.pdf>>.
- Pedersen, Bolette Sandford, Sanni Nimb, Jørg Asmussen, Nicolai Sørensen, Lars Trap-Jensen & Henrik Lorentzen (2009): DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. I: *Language resources and evaluation* 43, 269-299. <<http://doi.org/10.1007/s10579-009-9092-1>>.
- Pedersen, Bolette S., Nathalie Carmen Hau Sørensen, Sussi Olsen, Sanni Nimb & Simon Gray (2024a): Towards a Danish Semantic Reasoning Benchmark Compiled from Lexical-Semantic Resources for Assessing Selected Language Understanding Capabilities of Large Language Models. I: *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*. European Language Resources Association. 16356-16363. <<https://aclanthology.org/2024.lrec-main.1421>>.

- Pedersen, Bolette S., Nathalie Carmen Hau Sørensen, Sussi Olsen & Sanni Nimb (2024b): Evaluering af sprogforståelsen i danske sprogmodeller – med udgangspunkt i semantiske ord-bøger. I: *NyS - Nydanske Sprogstudier* 65, 8-40. <<https://doi.org/10.7146/nys.v1i65.143072>>.
- Petersson, Stellan & Emma Sköldbberg (2020-2021): To discriminate between discrimination and inclusion: a lexicographer's dilemma. I: Zoe Gavriilidou, Maria Mitsiaki & Asimakis Flia-touras (eds.): *Lexicography for Inclusion: Proceedings of the 19th EURALEX International Congress*, Alexandroupolis: Democri-tus University of Thrace. 381-386. <https://euralex.org/elx_proceedings/Euralex2020-2021/EURALEX2020-2021_ProceedingsBook-Vol1.pdf>.
- Wikipedia: Roget's Thesaurus*: <https://en.wikipedia.org/wiki/Roget%27s_Thesaurus> (februar 2025).
- Zechner, Niklas & Lars Borin (2020): Towards a Swedish Rog-et-Style Thesaurus for NLP. I: *Proceedings of the 2020 Globalex Workshop on Linked Lexicography*. Marseille: European Lan-guage Resources Association. 53-60. <<https://aclanthology.org/2020.globalex-1.9>>.

Sanni Nimb
seniorredaktør, ph.d.
Det Danske Sprog- og
Litteraturselskab
Christians Brygge 1
DK-1219 København K
sn@dsl.dk

Agnes Aggergaard Mikkelsen
assisterende redaktør, cand.mag.
Det Danske Sprog- og
Litteraturselskab
Christians Brygge 1
DK-1219 København K
aam@dsl.dk

Revisjonen av *Bokmålsordboka* og *Nynorskordboka*: oppdatering og tilpasning av ordbøkene til en ny tid

Bente Selback, Klara Sjø, Knut E. Karlsen & Sturla Berg-Olsen

This article outlines the work done between 2018 and 2024 to revise and upgrade the two Norwegian standard dictionaries. We discuss the motivation for the project and its execution, focusing on how tasks and workflow were adjusted along the way in light of unexpected challenges. The two dictionaries were revised in parallel, and unmotivated differences between them were levelled out. Practically all existing entries were adjusted and many new ones added. The project also resulted in a new and modernised website for the dictionaries, developed to meet the needs of the users, as well as a new editing platform.

1. Ordbøkene, historien og behovet for revisjon

Bokmålsordboka (BOB) og *Nynorskordboka* (NOB) er to enspråklige ordbøker som i dag begge inneholder mellom 90 000 og 100 000 artikler. De eies i fellesskap av Universitetet i Bergen (UiB) og Språkrådet. Artiklene i ordbøkene gir fullstendig oversyn over rettskrivningsnormene for bokmål og nynorsk, dvs. at de viser alle skrivemåter og bøyingsformer som er tillatt i de to offisielle normene. Utvalget av oppslagsord, betydninger og brukseksempler gjenspeiler bruken i norske allmennspråklige tekster på 2020-tallet og ca. 50 år tilbake i tid. Ordbøkene ble opprinnelig publisert på papir, men er i dag bare tilgjengelige digitalt, på nettstedet *ordbokene.no*.

Fra 2018 til 2024 gikk ordbøkene gjennom en grundig og svært påkrevd innholdsrevisjon som berørte alle artikler fra A til Å. I prosjektperioden ble ordbøkene modernisert på flere måter.

I første omgang gjaldt det selve innholdet: Mange nye oppslagsord ble tatt inn, og umotiverte forskjeller mellom de to ordbøkene jevnet ut. I tillegg besluttet en underveis i prosjektperioden å gjøre gjennomgripende endringer i brukergrensesnitt, design og funksjonalitet for å møte brukernes behov og oppfylle krav til universell utforming av nettsider. Nedenfor gjør vi greie for hvorfor det var helt nødvendig med en stor revisjon av ordbøkene. Vi skildrer hvordan revisjonen ble gjennomført, og hvilke resultater som ble oppnådd, men nevner også noen oppgaver som står igjen etter at revisjonen er fullført. Det ble tidlig klart at den opprinnelige prosjektplanen ikke tok høyde for en del endringer som var helt påkrevd dersom ordbøkene fortsatt skulle være relevante verktøy for brukerne, og det førte til at det måtte gjøres omprioriteringer underveis. Vi viser hvordan Revisjonsprosjektet ble brukt til å bygge opp et leksikografisk fagmiljø så å si fra bunnen av ved UiB. Mot slutten av artikkelen tar vi opp noen erfaringer som kan ha overføringsverdi for andre lignende prosjekter.

1.1. Historien

Dagens BOB og NOB kan spore historien sin tilbake til *Innstilling om språksaken*, som ble lagt fram i 1966 for Kirke- og undervisningsdepartementet av Vogtkomiteen. Mandatet til komiteen var «å ta opp til drøfting heile språksituasjonen i landet og gjera framlegg om tiltak som komitéen meiner kan tena til å ta vare på og utvikla vår norske språkarv» (KUD 1966:3). Bakteppet var den turbulente språkstriden i 1950- og 1960-årene, med konflikter mellom tilhengere av bokmål/riksmål og nynorsk og også innenfor hver av disse leirene. Komiteens innstilling inneholdt blant mye annet en statusbeskrivelse av det norske ordboklandskapet, og den pekte på det som ble ansett som et primærbehov: «Det som i dag trenges kanskje mere enn noe annet, er gode, praktiske ettbinds ordbøker til overkommelig pris for bokmål og nynorsk,

eventuelt for begge språkformer under ett.» (KUD 1966:13). Slike ettbinds enspråklige ordbøker med informasjon om rettskrivning, men også med definisjoner og brukseksempler, var mangelvare i Norge. En hadde på den ene side store litterære og vitenskapelige ordbokverk som *Norsk riksmålsordbok* (Knudsen & Sommerfelt 1937–1957) og *Norsk Ordbok*, og på den annen side et villnis av mindre ordlister, gjerne til skolebruk, som bare gav informasjon om rettskrivning og bøyning. På bokmålssiden stod *Tanums store rettskrivningsordbok* (Sverdrup & Sandvei 1961) i en særstilling med sine 460 sider, men heller ikke den inneholdt særlig mye mer enn rettskrivnings- og bøyningsinformasjon. I 1974 begynte arbeidet med å redigere nye ordbøker for bokmål og nynorsk ved det nylig opprettede Norsk leksikografisk institutt ved Universitetet i Oslo (UiO), med både faglig og finansiell involvering fra Norsk språkråd, som også hadde blitt etablert som resultat av Vogtkomiteens innstilling.

De to ordbøkene ble redigert av hver sin redaksjon, men arbeidet var til en viss grad koordinert: Nynorskredaksjonen begynte på begynnelsen av alfabetet, bokmålsredaksjonen midt i alfabetet, og så byttet en manus ved halvgått løp (Rauset 2019:156). Utgivelsen av disse «tvillingordbøkene» i 1986 var en storhendelse. Ordbøkene var etterspurt og måtte snart trykkes opp i flere opplag.

Det kom nye utgaver av begge ordbøkene i de påfølgende ti-årene: BOB fikk sin 2. utgave i 1993 og 3. utgave i 2005, mens NOB kom ut i 2. utgave i 1993, 3. utgave i 2001 og 4. utgave i 2006. De nye utgavene var oppdatert i henhold til de siste rettskrivningsvedtakene, og det ble også gjort visse andre endringer, som retting av feil og opptak av nye oppslagsord. I store trekk forble likevel både lemmalistene og den redaksjonelle teksten uendret.

Allerede i 1994 ble ordbøkene gjort gratis tilgjengelige i digital versjon på internett. Funksjonaliteten var enkel, men det er verdt å merke seg at en allerede fra starten av gav brukerne mulighet til å søke i begge ordbøkene på én gang. Som Hoel (2010:10) skrev i en

anmeldelse av en nyoppusset versjon av nettsidene, var det både pedagogiske og språkpolitiske grunner til dette parallellsøket: En ville vise fram begge de norske skriftnormalene og ikke minst synliggjøre nynorsken som det minst brukte av de to språkene. I tiden etter 2000 framstod nettversjonen av ordbøkene i stadig større grad som det primære produktet. De store endringene i nynorskrettskrivningen som ble vedtatt i 2012, førte f.eks. ikke til noen ny papirversjon av NOB; i stedet ble endringene implementert i nettutgaven.

1.2. Status i 2017 og prosjektsøknaden

I 2016 overtok UiB den rollen UiO inntil da hadde hatt som deleier og prosjektpartner med redaksjonelt ansvar for ordbøkene (den andre prosjektpartneren var og er Språkrådet). Da status ble gjort opp i 2017, ble det åpenbart at ordbøkene sårt trengte både en innholdsrevisjon og en ansiktsløftning. Det hadde riktig nok skjedd noen framskritt i de foregående årene: Ordbøkene var blitt koblet til Norsk ordbank – to databaser som inneholder fullstendig informasjon om normeringsstatus og bøyningsformer for et stort antall lemmaer. Det betydde at informasjon om skrivemåter og bøyning i ordbøkene ble hentet fra ordbanken. En hadde også ekspandert en del av de forkortelsene som var arvet fra papirordbøkene, og det var gjort et arbeid med å trekke ut faste uttrykk som søkbare sublemmaer. Likevel var strukturen i ordbøkene i stor grad fortsatt tilpasset papirmediet, og både brukergrensesnittet på nettsiden og redigeringsverktøyene var teknisk utdatert. Parallellvisningen av de to ordbøkene på nettsiden gjorde det dessuten lett å legge merke til at det var mange umotiverte forskjeller mellom BOB og NOB, noe mange brukere syntes var forvirrende.

For Språkrådet, som er statens forvaltningsorgan i språkspørsmål og har ansvar for forvaltningen av rettskrivningsnormene for norsk, er ordbøkene svært viktige kanaler for formidling

av normene og av godt og korrekt språk. På bakgrunn av dette la Språkrådet i budsjettsoknaden sin til Kulturdepartementet for 2018 inn en søknad til et prosjekt for en større revisjon av BOB og NOB. Prosjektsoknaden (Språkrådet 2017) listet opp hele 14 oppgaver som skulle løses, med særlig vekt på samkjøring av lemmalistene, opptak av nye oppslagsord og revisjon av definisjonstekst. Prosjektarbeidet skulle skje ved UiB, der det i løpet av prosjektet også skulle bygges opp et leksikografisk fagmiljø. Planen la opp til en tidsramme på 5,5 år. Søknaden resulterte i en bevilgning på statsbudsjettet for 2018 som gjorde det mulig for prosjektet å starte opp 1. september det året med fem fast ansatte leksikografer. Gjennom hele prosjektperioden var det Kulturdepartementet¹ som hovedsakelig stod for finansieringen, med en bevilgning på mellom 5 og 6 millioner kroner årlig (Rauset 2025:24–25).

2. Gjennomføringen av revisjonen

I prosjektsoknaden var det lagt noen føringer som fikk følger for hvordan arbeidet ble lagt opp. For det første var det bestemt at ordbøkene ikke lenger skulle komme ut på papir, men være rene nettordbøker. For det andre var det fra nå av én redaksjon som skulle jobbe med begge ordbøkene parallelt. Denne organiseringen ble valgt fordi redaktørgruppa var så liten at det i praksis ville være vanskelig å dele den i to, men like viktig var hensynet til de oppgavene som skulle løses. Et sentralt mål var å gjøre de to ordbøkene likere hverandre i lemmautvalg og artikkelstruktur, og da var det mest hensiktsmessig at de samme redaktørene arbeidet med begge ordbøkene. Arbeidet ble derfor lagt opp slik at ordbøkene ble delt inn i alfabetstrekk kalt bolker, og den samme personen redigerte den samme bolken i begge ordbøkene samtidig.

Høsten 2019, altså ett år etter at arbeidet startet opp, ble det

1 I 2022 omdøpt Kultur- og likestillingsdepartementet.

gjort opp en foreløpig status. Da viste det seg at redaksjonen ikke en gang var i nærheten av å nå arbeidsmålet for det første året. Under halvparten av de planlagte bolkene var ferdig redigert. Hva gikk galt?

Prosjektsøknaden var svært ambisiøs. Det var lagt opp til at revisjonen skulle omfatte mange oppgaver: Lemmalistene i BOB og NOB skulle samkjøres, den nye rettskrivningen fra 2012 skulle gjennomføres i hele NOB, nye oppslagsord skulle tas inn i begge ordbøkene, gamle artikler som manglet tilstrekkelig grunnlag i nyere tekst, skulle avpubliseres, definisjonstekstene skulle revideres, alle brukseksempler skulle gjennomgå og gjøres representative og tidsmessige, tomme artikler skulle få definisjon eller avpubliseres, bøyningsopplysningene for adjektiver måtte samordnes mellom de to ordbøkene, etymologiene skulle revideres, uttaleopplysninger skulle systemiseres, det skulle ryddes i visning og bruk av alternative rettskrivningsvarianter, og søket måtte forbedres slik at brukerne oftere fikk treff selv om de stavet oppslagsordet feil. I tillegg til dette skulle et fagmiljø i leksikografi bygges opp fra grunnen av ved UiB (Språkrådet 2017).

Redaksjonen bestod av flere erfarne redaktører, men noen var også nye på dette fagfeltet, og det gikk en del tid med til opplæring. Det tok tid før de nye redaktørene kom opp i det redigeringstempoet som det var lagt opp til. Det viste seg dessuten at det administrative arbeidet med å drive et så stort prosjekt og bygge opp et fagmiljø var større enn antatt, og dermed ble det mindre tid til selve bolkrevideringen.

I starten gikk det med mye tid til å trenge inn i og forstå tidligere praksis i to ordbøker som hadde hatt forskjellige redaksjoner og litt forskjellig redigeringspraksis. Den nye redaksjonen måtte utarbeide retningslinjer for hvordan den ønsket å arbeide videre, og skrive en redigeringshåndbok underveis.

Det var også en del tekniske utfordringer. Redigeringsprogrammet var gammelt og fungerte dårlig, noe som gjorde at selv

enkle arbeidsoperasjoner tok lang tid. Mye tid ble brukt på å ta omveier eller finne nødløsninger for å komme rundt de tekniske utfordringene. I tillegg viste det seg at det var nødvendig med en ny visningsside og en helt ny datainfrastruktur, og dette var det ikke tatt høyde for i søknaden (se avsnitt 3.1).

En av hovedoppgavene i revisjonen var å utjevne forskjeller mellom de to ordbøkene i de tilfellene der de ikke skyldtes faktiske forskjeller mellom bokmål og nynorsk. Ingen hadde erfaring med en slik parallellredigering av to ordbøker fra før, og tidsestimatet i søknaden, som baserte seg på testredigering av én ordbok, var for optimistisk. Da arbeidet kom i gang, viste det seg raskt at samkjøringen av ordbøkene var en svært omfattende oppgave. Forskjellene i både lemmautvalg og artikkelstruktur var større enn noen hadde forestilt seg, og flertallet av artiklene måtte endres på en eller annen måte (Selback 2020).

2.1. Prioriteringer og tiltak

Ett år inn i prosjektperioden var det tydelig at det ikke ville være mulig å gjennomføre alt som stod i søknaden, med de økonomiske ressursene og på den tiden redaksjonen hadde til disposisjon. Det var nødvendig å gjøre klare prioriteringer. Hva var viktigst, og hva kunne kuttes helt eller delvis? Det ble gjennomført flere møter både innad i redaksjonen og mellom redaksjonen og Språkrådet, og en kom i fellesskap fram til noen prinsipper for hvordan det skulle prioriteres:

- Brukerne og deres behov var det viktigste.
- De tre hovedoppgavene som var framhevet i søknaden (se avsnitt 1.2), skulle ha førsteprioritet.
- Det måtte kuttes i noe som gav en reell tidsbesparelse.
- Det var bedre å kutte noen klart definerte oppgaver som en kunne gjenoppta på et senere tidspunkt, enn å gjøre alt litt halvveis.

Med utgangspunkt i disse prioriteringene ble det satt i verk konkrete tiltak. Bolkredigeringen fikk strammere tidsstyring slik at redaktørene bare kunne gjøre det som var mulig innenfor en tilmålt tid for hver bolk. De viktigste oppgavene var å utjevne forskjeller mellom de to ordbøkene, oppdatere definisjoner og eksempler i de eksisterende artiklene og legge til nye oppslagsord. Det ble besluttet at arbeid med etymologier og uttale skulle nedprioriteres. Det ble ryddet litt i dem slik at de var rimelig like i begge ordbøkene, men ellers ble arbeidet satt på vent. Dette var smertefulle kutt, men disse oppgavene var så omfattende at det var nødvendig for at prosjektet ellers skulle kunne fullføres.

Språkrådet tok på seg alt arbeid med sluttkontroll av bolkene for å redusere arbeidsbyrden for redaksjonen (jf. avsnitt 4). Staben ble dessuten styrket med flere redaktører og vitenskapelige assistenter i perioder der det eksisterende budsjettet tillot det. Utviklingen av et nytt redigeringsprogram ble framskyndet, slik at det ble mulig å arbeide mer effektivt.

3. Tilpasning av ordbøkene

3.1. Tekniske tilpasninger

Som nevnt var det stort behov for en ny datateknisk løsning på grunn av utdatert teknologi, manglende dokumentasjon og delvis manglende kildekode. Det var derfor nødvendig å utvikle en helt ny IT-arkitektur, og dermed et nytt redigeringsgrensesnitt for leksikografene. Det var også ønske om et nytt brukergrensesnitt av flere grunner. Blant annet oppgav en del yngre brukere (Rauset 2022a:98) at den gamle visningen framstod som gammeldags, og at de heller valgte å bruke andre ordbokressurser. Det var problematisk, siden andre ressurser ikke nødvendigvis gir korrekt informasjon om rettskrivningsnormene. Stadig flere søk ble gjort

på mobiltelefoner og andre små skjermer (per 2025 er omtrent halvparten av søkene gjort fra mobiltelefon), så det var behov for et adaptivt system til bruk på mange ulike enheter. Dessuten var nettsiden meldt inn for manglende universell utforming for blinde og svaksynte til Diskrimineringsnemnda, et nøytralt forvaltningsorgan som avgjør klager som gjelder diskriminering, trakassering og gjengjeldelse. Design- og teknologibyrået Netlife (nå Tre Bergen) ble hyret inn for å lage en ny brukernettside (se Rauset 2022a:99).

For å finne ut om ordbøkene innfridde brukernes krav og forventninger, laget redaksjonen i samarbeid med Netlife en undersøkelse som ble presentert på brukersiden i november–desember 2021 og igjen etter overgang til ny side, i mars 2022 (Rauset 2022a:109ff.). De to undersøkelsene hadde ca. 700 respondenter hver gang. Målet med undersøkelsen var primært å se om brukerne fikk gjort den oppgaven de kom til nettsiden for å gjøre, men redaksjonen bad også om informasjon om hvilken gruppe respondentene tilhørte (elev, skribent, språkinteressert o.l.), og om språkbakgrunn. Det gav redaksjonen bedre forståelse av hvem brukerne var, og hvilke behov de hadde, og dette la redaksjonen til grunn i utviklingen av nettsiden og innholdet i ordbøkene.

3.2. Ny brukerside

Den nye brukersiden ble presentert i januar 2022. Siden har et mer pedagogisk oppsett med større skrift og mer luft mellom artikkeldelene. De ulike delene av artiklene har fått lett forståelige overskrifter som «Betydning og bruk», «Eksempler» og «Opphav». Hver definisjon og hvert eksempel føres opp på en ny linje (se figur 1 og figur 2). Forkortelser i definisjonsteksten er skrevet fullt ut, og grammatiske forklaringer er holdt til et minimum og basert på terminologien som brukes i skolen.

gardin **gardin m1, f1, n3** (gjennom tysk, fra latin) forheng framfor vindu
trekke for, fra gardenene / bak nedrullede gardiner / rullegardin, tyllgardin

Figur 1: Artikkelen *gardin* i *Bokmålsordboka* før revisjonen.

Brukerundersøkelsene viste at bøyningsinformasjon var svært sentral for brukerne. Normen i både bokmål og nynorsk gir svært mange valgmuligheter, og ett enkelt ord kan ha flere normerte former, både når det gjelder skrivemåte og bøyning. En av de viktigste oppgavene BOB og NOB har, er å vise alle disse mulighetene. I den gamle visningen var hvert oppslag markert med en bøyingskode, f.eks. m1, f1 og n3 (se figur 1). For å få se hele bøyingsmønsteret måtte en klikke på denne koden. Dette var lite intuitivt, og spørsmål til redaksjonen om bøyingsmønster viste også at denne funksjonaliteten i liten grad var kjent. I den nye visningen (se figur 2) var bøyningstabellene gjort større og mer pedagogiske, og de viste alle de tillatte formene ordet kan ha. Brukeren kan selv velge i visningen om hen alltid vil vise bøyningstabellene eller ikke, men dersom de velges vekk, er de lett tilgjengelige ved å klikke på «Vis bøyning».

Brukeren kan også velge om hen vil søke i oppslagsord, med bøyde former eller i fritekst, som gir treff i alt innhold i artiklene. Dersom et søk ikke gir treff i noe oppslagsord, men gir treff i en bøyde form av et ord, får brukeren beskjed om det. Hen får også forslag om «liknende oppslagsord» basert på Levenshtein-avstand, som måler forskjellen mellom søkestrengen og mulige oppslagsord og velger de oppslagsordene som har lavest avstand, altså færrest forskjeller. Mange brukere har en oppfatning av at ord som ikke står i ordbøkene, ikke er akseptable. Dersom det ikke er treff på noe som oppfattes som en sammensetning, får en derfor nå opp denne teksten: «Merk at det er talløse kombinasjonsmuligheter av

gardinSUBSTANTIV *hankjønn, hunkjønn eller intetkjønn* (m1, n., f1)

<i>kjønn</i>	entall		flertall	
	<i>ubestemt form</i>	<i>bestemt form</i>	<i>ubestemt form</i>	<i>bestemt form</i>
hankjønn	<i>en</i> gardin	gardenen	gardiner	gardinene
hunkjønn	<i>ei/en</i> gardin	gardina		
intetkjønn	<i>et</i> gardin	gardenet	gardin, gardiner	gardina, gardenene

OPPHAV gjennom tysk; fra middelalderlatin *cortina***BETYDNING OG BRUK**

forheng framfor vindu

Eksempel*trekke for gardenene;**trekke fra gardenene;**sitte bak nedrullede gardiner*som etterledd i ord som blondegardin, fortrekksgardin, rullegardinFigur 2: Artikkelen *gardin* i *Bokmålsordboka* etter revisjonen.

sammensatte ord på norsk, og mange slike ord er derfor ikke tatt med i ordbøkene.»

3.3. Revisjon av innholdet

Som nevnt var størstedelen av innholdet i ordbøkene ved starten av revisjonen minst 30–40 år gammelt. For at de skulle framstå som moderne standardordbøker, var det behov for både å oppdatere med nye lemma og legge til nye betydninger i eksisterende artikler. Dessuten ble gammeldags språkbruk i definisjoner og eksempler erstattet med mer moderne og tilgjengelige formuleringer. Rene synonymdefinisjoner i de gamle artiklene ble i stor grad utvidet. For eksempel har artikkelen *flørte*, der verbet tidligere var

definert med de lavfrekvente synonymene *kokettere* og *kurtisere*, nå realdefinisjonen *gjøre uforpliktende erotiske tilnærmelser* i tillegg til synonymene.

Flere av artiklene framstod som problematiske sett med moderne øyne, særlig artikler knyttet til etnisitet og funksjonsnedsettelse (se Lyse & Svardal 2020). Derfor arbeidet redaksjonen mer systematisk med noen av disse artikkelgruppene. En redaktør kunne for eksempel få ansvar for å arbeide med en gruppe ord som berørte de nasjonale minoritetene i Norge, eller ord som omhandlet kjønn og kjønnsinkongruens. Settet av artikler ble så lagt fram for alle redaktørene og diskutert i plenum. På den måten ble artiklene mer enhetlig behandlet, og lenkene i de ulike definisjonene fulgte samme logikk.

Det var en viktig arbeidsoppgave å omarbeide brukseksempelne slik at de ble språklig og kulturelt oppdatert. Antallet eksempler har under revisjonen økt til langt mer enn det dobbelte, og de er utvidet til å inneholde flere bøyde former og mer fullstendige setninger.

Redaksjonen var klar over at kjønnsbalansen i eksemplene ikke var optimal. I 2020 ble det utført en undersøkelse som viste at *han* var klart mer brukt enn *hun* i begge ordbøkene. I eksempler med et personlig pronomen i tredje person entall opptrådte *han* i 65 % av tilfellene i BOB og i 60 % i NOB. Ved avslutningen av prosjektet utgjorde andelen henholdsvis 54 % og 50 %. Aadlands undersøkelse (2025) viser også at ordbøkene har blitt mindre stereotype i løpet av revideringsprosessen. Fordi BOB og NOB vises parallelt, har redaktørene gjerne brukt *han* i én ordbok og *hun* i den andre, slik som i artikkelen *game*, der eksempelet i BOB er *jenter som gamer*, mens det i NOB er *mange gutar gamar kvar dag* for å vise at det å game ikke nødvendigvis er knyttet til kjønn.

3.4. Andre tilpasninger

Brukerundersøkelsene viste at ordbøkene hadde ca. 20 % brukere som ikke hadde norsk som morsmål, mens bare et fåtall oppgav at de gikk på norskopplæring. Det var overraskende for redaksjonen at antallet brukere som ikke hadde norsk som førstespråk, var så høyt, og at brukergruppa var så variert. I tillegg til fokuset på brukere under utdanning var dette styrende for en del prinsipper som ble lagt til grunn for definisjonsutformingen. Det var ønskelig med lettere tilgjengelig språk i artiklene, med færre forkortelser og mindre fagspråk. Dessuten ønsket redaksjonen å ha flere eksempler og arbeidet bevisst med å legge til flere faste uttrykk. Tidligere tomme artikler – som hovedsakelig ble brukt for å vise sammensetningsrekker – fikk definisjoner og eksempler. I tillegg arbeidet redaksjonen systematisk med noen ord i ordklassene determinativ og subjunksjoner. Disse to ordklassene er relativt nye i den norske skolegrammatikken, og de er kompliserte, særlig for innlærere. Redaksjonen valgte derfor å arbeide med dem systematisk slik at de fikk en enhetlig framstilling (se også Sjo 2025).

3.5. Funksjon for oversettelse mellom bokmål og nynorsk

I norsk skole må alle lære å skrive både bokmål og nynorsk, og i offentlig forvaltning må en også forholde seg til begge skriftspråkene. En av de funksjonene som har vært svært mye etterspurt både av lærere, elever og ansatte i offentlig sektor som skriver mye, er en oversettelsesfunksjon mellom nynorsk og bokmål. Funksjonen viser kuraterte forslag for ord i den andre ordboken dersom det ikke finnes et ord med identisk skrivemåte. Denne funksjonen ble implementert i ordbøkene i 2024 og ble raskt svært populær.

Bokmålsordboka 1

hensette

VERB (verb)

Vis bøyning ▾

BETYDNING OG BRUK
 anbringe (1) på et bestemt sted;
 plassere, sette

Eksempel
*sykler skal hensettes på anvist sted;
 føle seg hensatt til en fjern fortid;
 hensette seg i beruset tilstand*

Del Siter Artikkelside

Nynorskordboka 0

Mogleg omsetjing

Klikk på forslaga under og vurder ut frå samanhengen om nokon av dei er brukande. Forslaga er hovudsakleg basert på omsetjingsprogrammet Apertium. Hugs at omsetjingar frå slike program kan vere mangelfulle.

- plassera/plassere
- etterlata/etterlate
- setja/setje/setta/sette

Liknande oppslagsord:

- hærsette
- innsette
- resette
- anisette
- avsette

Figur 3: Oversettelsesforslag på nynorsk for ordet *hensette*.

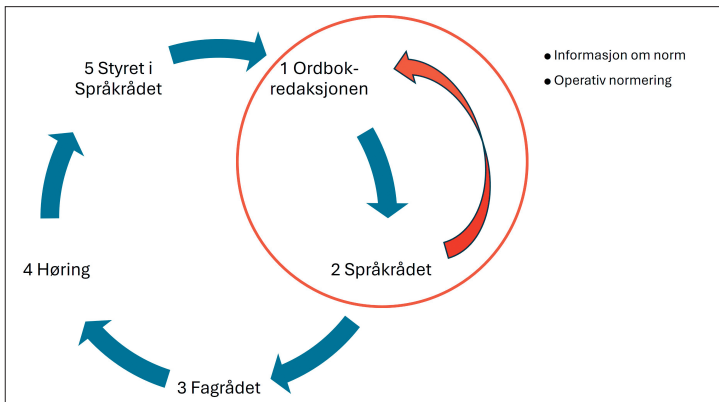
4. Språkrådets rolle

I Norge er de offisielle skriftnormalene for bokmål og nynorsk gitt av staten, og BOB og NOB er myndighetenes viktigste verktøy for å formidle informasjon om rettskrivningen i norsk. En viktig del av Språkrådets rolle i Revisjonsprosjektet har dermed vært å fastsette korrekt skrivemåte og bøyning for ord i tillegg til å veilede om gjeldende norm ved å vise til tidligere vedtak og konsekvenser av slike vedtak. Normeringsveiledningen og den utvidede sluttkontrollen av ordbokmanuset (jf. avsnitt 2.1), der tre til fire medarbeidere i Språkrådet med leksikografisk kompetanse kvalitetssikret ordbokbolkene på bred basis før publisering, gjorde at Språkrådets innsats i prosjektet ble større enn først planlagt (jf. Rauset 2025).

4.1. Normering – fra innspill til ordbok

Mange av spørsmålene ordbokredaksjonen reiste, krevde ikke normering i det hele tatt, men snarere ferdighet i å finne fram i kilder som inneholder informasjon om tidligere normering. Stør-

stedelen av sakene som faktisk krever normering, blir løst gjennom det som kalles *operativ normering*. Det betyr at normkildene ikke gir klare svar, men at det likevel er mulig å «resonnere seg fram til svaret ut fra normeringsprinsippene, parallelle tilfeller og logikken i språkssystemet» (Språkrådet 2021). Så lenge en holder seg på dette nivået (innenfor sirkelen, jf. figur 4), kan spørsmål avklares raskt. Dersom det derimot er behov for en mer prinsipiell vurdering av sakskomplekset, blir saken sendt videre til fagrådet for normering og språkbobservasjon i Språkrådet. Det er en langt mer omfattende prosedyre, for dersom en normeringssak først blir behandlet i fagrådet, impliserer det at saken skal ut på allmenn høring (sv. *remiss*) før endelig vedtak i styret i Språkrådet. Dette er en prosess som både tar tid og krever ressurser i form av saksutgreiing om normhistorie, bruk osv.



Figur 4: Gangen i normeringen av norsk skriftspråk.

Ett av de over tusen spørsmålene som ble meldt inn, gjaldt normering av ordet *kaskett/kaskjett* ‘hodeplagg med skjerm, ofte brukt som uniformslue’. Redaktøren registrerte at ordet hadde ulik normering i bokmål og nynorsk, og det skulle en ikke vente i et importord som dette. Videre viste redaktøren til at det fantes et

vedtak fra 1995 som sa at skrivemåten skulle være *kaskett* i bokmål, mens det ikke fantes noe vedtak for nynorsk. Språkrådet bekreftet at saken måtte til behandling i fagrådet med saksutgreiing, allmenn høring og nytt vedtak i styret i Språkrådet.

Ordningen med høring av rettskrivningsendringer ble innført i 2021, og praksisen er at det blir gjennomført én høringsrunde i året. Tanken bak allmenn (offentlig) høring er at forvaltningen (sv. *statsadministrationen*) skal opplyse saker så godt som mulig før avgjørelser fattes. Når saker blir sendt på høring, får innbyggere, organisasjoner, næringsliv og andre forvaltningsorganer mulighet til å si sin mening og føre kontroll med hvordan forvaltningen gjør jobben sin.

Saken om skrivemåten av *kaskett*/*kasjett* var del av høringsrunden i 2023. Med grunngivelse i bruk, uttale og regler for normering av importord var forslaget fra Språkrådet at ordet skulle ha lik normering i bokmål og nynorsk med valgfri skrivemåte *kasjett* eller *kaskett*. Etter høringsrunden ble høringssvarene vurdert. Det kom ingen vesentlige innvendinger til forslagene om ny normering av dette ordet. Forslaget gikk uendret til styret og ble vedtatt ca. 16 måneder etter at redaktøren meldte inn saken. Da kunne nynormeringen gjøres kjent for ordbokbrukerne gjennom oppdatering av nettversjonen.

5. Forskning og fagmiljøbygging

Et av elementene i prosjektsøknaden (Språkrådet 2017) var å bygge et nytt fagmiljø for leksikografi. Redaktørene har hatt sin arbeidsplass ved og vært knyttet til fagmiljøene ved Institutt for lingvistiske, litterære og estetiske studier ved UiB. Å være tilknyttet et større faglig miljø har vært en fordel og et gode for redaktørene.

Redaktørene i prosjektet hadde ulik bakgrunn og kompetans, og erfaringen med praktisk leksikografi varierte mye. De har

kunnet dra veksler på hverandres spesialkompetanse og lære av hverandre. Samarbeidet med Språkrådet, som hadde flere erfarne leksikografer som utførte sluttkontroll og svarte på normerings-spørsmål, bidro også med verdifull kompetanse.

Det har blitt avlagt to doktorgrader som er knyttet til det leksikografiske miljøet og direkte relevante for det leksikografiske arbeidet på UiB, Rausets i 2022 (Rauset 2022b) og Paulsens i 2024 (Paulsen 2024). En tredje stipendiat tilknyttet prosjektet startet i 2024 og skal etter planen være ferdig i 2028. Redaktørene har vært faglig aktive og deltatt på ulike relevante konferanser, og til sammen er det publisert 22 vitenskapelige artikler med utgangspunkt i Revisjonsprosjektet (for fullstendig oversikt se Rauset 2025). I 2020 ble faget LEX200, et emne i leksikografi, holdt for første gang, og redaktørene har også vært gjesteforelesere på flere andre kurs innen lingvistikk, nordisk og norsk som andrespråk.

Redaktørene har vært aktive formidlere av arbeidet sitt for eksempel med opptredener i radio og fjernsyn, besøk hos skoleklasser og foreninger og ellers både muntlig og skriftlig gjennom formidling i en rekke ulike kanaler. Redaksjonen ved Revisjonsprosjektet ble tildelt Det humanistiske fakultets formidlingspris 2023 for dette arbeidet.

6. Læringspunkter

Vi kan slå fast at det ble brukt for få ressurser på å holde BOB og NOB oppdatert etter førsteutgivelsen i 1986. Selv om det ble gjort visse oppdateringer i de påfølgende utgavene, klarte en ikke å holde tritt med endringene i språket og i brukergruppa. Da Revisjonsprosjektet kom i gang, var det derfor svært mange oppgaver å løse, og prosjektet fikk preg av å være et skippertak. Brukerne er helt klart bedre tjent med en løpende oppdatering av alle opplysningene i ordbøkene. I denne typen ordbøker er det særlig viktig

at innholdet framstår oppdatert og relevant. I praksis viste det seg at det bare var rundt 30 artikler i hver av ordbøkene som forble helt uendrede etter Revisjonsprosjektet. Det illustrerer behovet for fortløpende oppdateringer i ordbøkene, både med tanke på nye generasjoner ordbokbrukere og nye brukergrupper. Dersom en ordbok ikke vedlikeholdes fortløpende, må en regne med å bruke større ressurser på å gjøre innholdet tidsriktig.

Revisjonsprosjektets viktigste oppgaver var å tilføre nye oppslagsord, utjevne umotiverte forskjeller i lemmalistene mellom de to ordbøkene og oppdatere definisjonsteksten (jf. avsnitt 2.1). Disse oppgavene er fullført som planlagt. I prosjektets første fase ble det avdekket behov og ønsker som det ikke var tatt god nok høyde for i prosjektsøknaden. Det gjaldt behovet for ny redigeringsplattform for redaktørene, men ikke minst gjaldt det også krav og ønsker fra ordbokbrukerne. Ettersom oppmerksomheten rundt revisjonen økte i mediene og redaksjonen initierte brukerkontakt, ble det klart at det var nødvendig å investere større ressurser i å forbedre brukergrensesnittet både med tanke på design og funksjonalitet. I flere faser av revisjonen har brukere av ordbøkene fått mulighet til å uttale seg om ulike løsninger, og flere innspill har resultert i konkrete endringer (jf. Rauset 2022a og 2025). I framtiden bør trolig ordbokbrukerne i enda større grad få et ord med i laget, særlig når det gjelder informasjonsstruktur og visuelle skjerm-løsninger på pc og mobil. Selv om bruker-involvering er ressurskrevende og kan gå ut over tiden en har til det konkrete revisjonsarbeidet, er det trolig en god investering i alle ordbokprosjekter. Det er tross alt brukerne ordbøkene er til for.

I et så langvarig og omfattende revisjonsprosjekt vil det garantert oppstå uforutsette problemer det ikke er tatt høyde for i prosjektsøknaden, også større utfordringer som det kan kreve både tid og ressurser å løse. Generelt må det tas høyde for at sykdoms-fravær vil være merkbart i et lite og spesialisert arbeidsmiljø som

en ordbokredaksjon. Det er heller ikke enkelt å finne erstattere med riktig kompetanse ved fravær. Til dels kan en planlegge for det uforutsette ved å legge inn en buffer for merforbruk av tid og andre ressurser i prosjektp planen, men erfaringen viser at det er lett å være for optimistisk i planleggingsfasen. Som det kom fram i avsnitt 2.1, ble løsningen i dette prosjektet å skaffe seg rom ved å utsette noen av de planlagte oppgavene til driftsfasen etter prosjektslutt.

Prosjektet har vært et samarbeid mellom UiB og Språkrådet, der UiB har stått for gjennomføringen av revisjonsarbeidet, mens Språkrådet har vært prosjekteier med ansvar for normering og sluttkontroll av manus. Samarbeidet mellom UiB og Språkrådet har vært fruktbart og vist seg å fungere fint. Det betyr ikke at vi har vært enige i alle spørsmål og om alle prioriteringer, men en klar ansvarsfordeling og god rolleforståelse har likevel gjort at samarbeidet har fungert godt. For Språkrådets del har den interne leksikografiske kompetansen vært en styrke, helt fra utformingen av prosjektsøknaden. Den var også avgjørende for at Språkrådet kunne påta seg ansvaret for sluttkontrollen av manus underveis i prosjektperioden.

7. Framtidsperspektiv

Vi har sett at standardordbøker som BOB og NOB krever løpende oppdatering dersom de skal møte brukernes behov og ønsker og forbli relevante, samtidig som de skal fylle oppgaven som myndighetenes hovedkilde til informasjon til samfunnet og borgerne om gjeldende rettskrivning. Det betyr at det må settes av både menneskelige og økonomiske ressurser for å drive med feilretting, normoppdatering og lemmautvalg. Dessuten ble det underveis i prosjektet bestemt at visse oppgaver, for eksempel revisjon av etymologiske opplysninger, skulle utsettes til driftsperioden. Disse

oppgavene må nå prioriteres for at målene for Revisjonsprosjektet kan sies å være nådd.

Den trykte ordboka har utspilt sin rolle, i alle fall når det gjelder standardordbøker som NOB og BOB. I dagens samfunn er derfor en god digital infrastruktur helt avgjørende for at sluttbrukeren får et produkt som fungerer som planlagt. Det innebærer alt fra enkel, trygg og driftssikker digital tilgang til kildemateriale og funksjonell redigeringsplattform til en velfungerende pedagogisk god skjermlesning som fungerer godt på alle flater for en sammensatt brukergruppe.

Selv om nettsiden er videreutviklet i løpet av prosjektperioden, er det likevel oppgaver som gjenstår. Det gjelder blant annet videre utvikling av oversettelsesfunksjonen mellom bokmål og nynorsk. Det har også vært på tale å tilby en felles søkeportal med andre norske nettordbøker (etter mønster av for eksempel *svenska.se*), særlig når de andre ordbøkene som mottar offentlig støtte, oppfyller kravet om å vise de offisielle skriftnormalene i deres fulle bredde.

Ordbanken, de to databasene som inneholder opplysninger om skrive- og bøyningsmåter for ord i bokmål og nynorsk (jf. avsnitt 1.2), må vedlikeholdes og videreutvikles. Ordbanken er en avgjørende grunnlagsressurs, ikke bare for BOB og NOB, men også for andre ordbøker som henter informasjon om skrive- og bøyningsmåte derfra.

Avslutningsvis er det grunn til å understreke læringspunktet som gjelder brukerkontakt. Standardordbøkene BOB og NOB skal ikke dokumentere språket for framtidige generasjoner, men vise korrekt skrive- og bøyningsmåte av ord i allmennspråket med definisjoner for brukere i dag. Målet er at ordbøkene skal være nyttige ressurser for skolen, det offentlige og andre som trenger opplysninger om norsk innenfor skriftnormalene. Da må vi finne gode pedagogiske måter å presentere informasjonen på gjennom dialog med dem vi ønsker å nå.

Litteratur

Ordbøker

BOB = *Bokmålsordboka* (1986). Marit Ingebjørg Landrø & Boye Wangenstein (red.). Oslo: Universitetsforlaget. 2. utgave 1993, 3. utgave 2005. Nettversjon: Språkrådet og Universitetet i Bergen. <<https://ordbokene.no>> (mars 2025).

Knudsen, Trygve & Alf Sommerfelt (1937–57): *Norsk riksmålsordbok*. Oslo: Aschehoug.

NOB = *Nynorskordboka* (1986). Marit Hovdenak, Laurits Killingbergtrø, Arne Lauvhjell, Sigurd Nordlie, Magne Rommetveit & Dagfinn Worren (red.). Oslo: Det Norske Samlaget. 2. utgave 1993, 3. utgave 2001, 4. utgave 2006. Nettversjon: Språkrådet og Universitetet i Bergen. <<https://ordbokene.no>> (mars 2025).

Norsk Ordbok. Ordbok over det norske folkemålet og det nynorske skriftmålet (1966–2016). Oslo: Det Norske Samlaget.

Sverdrup, Jakob & Marius Sandvei (1961): *Tanums store rettskrivningsordbok*. Oslo: Johan Grundt Tanum forlag.

Annen litteratur

Hoel, Jan (2010): Edle årganger i ny tapning. I: *Språknytt* 3, 10–11.

KUD (1966) = *Innstilling om språksaken fra Komitéen til å vurdere språksituasjonen m.v. oppnevnt ved kongelig resolusjon 31. januar 1964*. Oslo: Kirke- og undervisningsdepartementet.

Lyse, Gunn Inger & Terje Svardal (2020): «Den raude rase» og «negerfolk»: kontroversielle ord i *Bokmålsordboka* og *Nynorskordboka*. I: *LexicoNordica* 27, 77–96.

Paulsen, Mikkel Ekeland (2024): *Å skille klinten fra hveten. Hvordan kan ordbokaktuelle sammensetninger identifiseres?* Doktorgradsavhandling. Bergen: Universitetet i Bergen. 178 s.

- Rauset, Margunn (2019): *Bokmålsordboka og Nynorskordboka – einægga, toegga eller siamesiske tvillinger?* I: *LexicoNordica* 26, 155–175.
- Rauset, Margunn (2022a): Brukarmedverknad i utvikling av nettsida ordbøkene.no. I: *LexicoNordica* 29, 97–118.
- Rauset, Margunn (2022b): *Frasar til besvær? Studiar av norm og bruk i norsk fraseologi*. Doktorgradsavhandling. Bergen: Universitetet i Bergen. 440 s.
- Rauset, Margunn (2025): *Sluttrapport. Revisjon av Bokmålsordboka og Nynorskordboka*. Universitetet i Bergen. <<https://hdl.handle.net/11250/3187459>> (september 2025).
- Selback, Bente (2020): «Å nei, det ordet er ikkje lov på nynorsk!» Eller ...? Om parallell redigering av to norske ordbøker. I: Caroline Sandström, Ulla-Maija Forsberg, Charlotta af Hällström-Reijonen & Klaas Ruppel (red.): *Nordiska studier i leksikografi* 15. Helsingfors: Nordisk förening för leksikografi. 297–305.
- Sjo, Klara (2025): Ny ordklasse – nye brukergrupper: Om brukertilpassing i ordbøkene.no. I: Margunn Rauset, Gyri Smørdal Losnegaard & Hanne Lauvstad (red.): *Nordiske studier i leksikografi* 17. Bergen: Nordisk forening for leksikografi.
- Språkrådet (2017): *Prosjektsøknad – revisjon av Bokmålsordboka og Nynorskordboka*. Upublisert.
- Språkrådet (2021): *Retningslinjer for normering av bokmål og nynorsk*. <<https://sprakradet.no/wp-content/uploads/retningslinjer-for-normering-2021-bokmalsversjon.pdf>> (mars 2025).
- Aadland, Emma Josefin Ölander (2025): Kjønnrepresentasjon og -fremstilling i *Bokmålsordbokas* (BOB) brukseksempel. I: *LexicoNordica* 32 (dette bindet).

SELBACK, SJO, KARLSEN & BERG-OLSEN

Bente Selback
leksikograf
Universitetet i Bergen
Postboks 7805
NO-5020 Bergen
bente.selback@uib.no

Klara Sjø
leksikograf
Universitetet i Bergen
Postboks 7805
NO-5020 Bergen
klara.sjo@uib.no

Knut E. Karlsen
seniorrådgiver, ph.d.
Språkrådet
Postboks 1573 Vika
NO-0118 Oslo
knut.e.karlsen@sprakradet.no

Sturla Berg-Olsen
seniorrådgiver, dr.art.
Språkrådet
Postboks 1573 Vika
NO-0118 Oslo
sturla.berg-olsen@sprakradet.no

Gamle historiske sprogarkiver og deres værdi i nutiden

Ásta Svavarsdóttir & Helga Hilmisdóttir

Thus far, archives of the Lexicographic Institute in Reykjavík, compiled in the 20th century, have been the only lexicographical resources available for older Icelandic. With new historical corpora in the making the situation has changed. In this article, we compare the two types of resources from different perspectives and conclude that archives and corpora have different strengths and weaknesses and thus supplement each other. Together, both the archives and corpora can continue to serve the needs of lexicographers and linguists in the future.

1. Indledning

Orðabók Háskólans (OH) blev stiftet ved Islands universitet tidligt i 1940'erne med det formål at forberede, redigere og udgive en videnskabelig, historisk ordbog over islandsk. Fokus blev sat på perioden fra midten af 1500-tallet, da de første islandske bøger blev trykt, til nutiden. De første årtier var arbejdet koncentreret omkring indsamling af materiale til ordbogen, og det resulterede i traditionelle seddelsamlinger af forskellig art og størrelse, den største af dem det såkaldte *Ritmálssafn* (ROH), et skriftsprogsarkiv med citater især fra trykte værker (se Svavarsdóttir 2024, Pind et al. 1993).

Digitalisering blev indført ganske tidligt på OH (nu en del af Árni Magnússon-Instituttet for Islandske Studier (ÁMI)), dvs. i 1980'erne. Digitaliseringen blev, i begyndelsen, først og fremmest betragtet som en måde at udføre en effektiv og systematisk redigering af materialet i OH's arkiver. Tidligt viste det sig at den almindelige digitalisering af trykkeriernes arbejds gange gav mulighed

for at få fat i en ny type materiale, dvs. sammenhængende tekster i digital form, og dermed opbygge et nyt grundlag for leksikografisk arbejde (se fx Pind et al. 1993:5). Fra trykkerierne kom der dog først og fremmest nye, samtidige tekster selv om der indimellem også fulgte enkelte genudgivelser af ældre værker. Man så derfor arkiverne fremadrettet som det vigtigste grundlag for en historisk beskrivelse af ordforrådet, mens opbygning af digitale, søgbare tekstsamlinger hovedsagelig ville erstatte traditionel excerpering når det gjaldt nyere sprog.

Nu er det blevet realistisk at digitalisere gamle tekster med scanning og OCR-behandling og dermed at opbygge store historiske sprogkorporer med tekster fra forskellige perioder og genrer. Og så er det på tide at gøre sig overvejelser om de gamle arkivers status i nutid og fremtid i forhold til nye digitale sprogresurser (jf. diskussion om forholdet mellem seddelsamlinger og korpusser i Hovmark 2012, 2023 og Trap-Jensen 2014).

Formålet med denne artikel er at sammenligne de gamle sprogarkiver med nye eller planlagte digitale sprogkorporer og diskutere deres fordele og ulemper vis-a-vis hinanden. Hovedspørgsmålene er de følgende: 1) Hvilken slags informationer findes der i de forskellige typer sprogresurser, og hvordan er forholdet mellem dem? 2) Gør udviklingen af digitale korpusser de gamle leksikalske arkiver forældede og overflødige, eller vil de gamle arkiver beholde deres værdi for fx leksikografer og sprogforskere? 3) Hvordan kan eller skal de gamle samlinger videreudvikles, fx med hensyn til (yderligere) digitalisering og analyse?

Dispositionen er som følger: I afsnit 2 beskrives ÁMI's vigtigste sprogresurser med hensyn til en udførlig, historisk beskrivelse af islandsk ordforråd, både de traditionelle sprogarkiver og nyere digitale sprogkorporer. Afsnit 3 omfatter en sammenligning af resurserne, som viser at hver og én har sine fordele og ulemper i forskellige sammenhænge og således supplerer hinanden. Afsnit 4 indeholder en kortfattet sammenfatning og konklusioner.

2. Leksikografiske sprogresurser for islandsk

2.1. Sprogarkiver

2.1.1. Seddelsamlinger

De gamle sprogarkiver fra OH indbefatter forskellige store og små samlinger. Vi fokuserer her på hovedsamlingerne, *Ritmálssafn Orðabókar Háskólans* (ROH), et skriftsprogsarkiv, og *Talmálssafn Orðabókar Háskólans* (TOH), som er et talesprogsarkiv. Begge er i udgangspunktet traditionelle seddelsamlinger, men er senere blevet digitaliseret i større eller mindre grad. Samlingerne er ellers vidt forskellige, med hensyn til deres størrelse, indhold, struktur, tidsrum osv. (se Tabel 1).

	ROH	TOH
Kilder	Trykte kilder og håndskrifter	Breve fra og samtaler med informanter
Type af information	Belæg på ord fra skriftsproget	Kommentarer om ord og ordbrug i daglig tale
Opslagsord	ca. 700.000	ca. 56.000
Omfang og indhold	ca. 2,2 mio. sedler med opslagsord, ordklasse, citat og kildehenvisning	ca. 200.000 sedler med diverse oplysninger (fx betydning, bøjning, udtale, syntaks mm.)
Tidsrum	16.-20. århundrede	20. århundrede
Formidling	Digitaliseret; søgbart på åbent website	Delvis digitaliseret; ikke alment tilgængeligt

Tabel 1: Oversigt over ROH og TOH.

ROH, skriftsprogsarkivet, er langt det største af samlingerne. Ved excerperingen var sigtet at dække så mange og varierede tekster som muligt fra midten af det 16. til slutningen af det 20. århundrede. Alle sedlerne har samme udformning og parallelle oplysninger: (1) opslagsordet samt ordklasse, fx *nýmæli*, n. ‘nyhed, inno-

vation' (n = substantiv, neutrum); (2) det exciperede citat med den pågældende ordform understreget, fx *Eg veit engin nýmæli herra!* 'Jeg kender ingen nyheder, herre'; (3) kildehenvisning, dvs. værkets faste akronym, eventuelt bind eller årgang, og sidetal, fx *StThLear 14* (en oversættelse af Kong Lear, side 14).

TOH, talesprogsarkivet, blev bygget op som resultat af et radioprogram som OH's medarbejdere styrede fra 1950'erne (jf. Ingólfsson 1988:68) til 2005. Ved siden af folkelig oplysning om sproget blev udsendelserne brugt til at komme i kontakt med lyttere som viste sig at være villige til at være sproginformanter. Dette gav leksikograferne mulighed for at få svar på spørgsmål og indsamle information, fx om bestemte ord og ordforbindelser, eller om ord inden for bestemte domæner. Svarene kom med breve eller i telefonsamtaler, og de pågældende oplysninger blev registreret på sedler med henvisning til informanten, hans eller hendes hjem- og fødeegn og alder.

I TOH består oplysningerne på en arkivseddel af fire dele: (1) opslagsord med ordklasse og køn af substantiver; (2) informantens kommentar om ordet; (3) det geografiske område som informanten stammer fra; (4) informantens navn og adresse samt brevets eller samtalens dato. Kommentarerne er vidt forskellige, afhængig af bl.a. spørgsmålet som blev stillet og informantens indstilling (jf. Hovmark 2023:166-168 om seddelsamlingen for *Ømålsordbogen*). De kan være alt fra et nej, at informanten ikke kender til ordet, til en udførlig kommentar om et ords eller en ordforbindelses semantiske, morfologiske og/eller syntaktiske egenskaber, eventuelt med brugseksempler og endda etnologisk information. Som et eksempel på en kommentar kan nævnes en informants udsagn om ordet *della* '(en type) bagværk'. I dansk oversættelse lyder kommentaren: "Dellur [plur.] (I udtalt som i Ella) = ástarpungar [dvs. samme slags boller], langt hyppigere at kalde dette for *dellur* i min egn og jeg vil fortsat gøre det." Her er der altså sagt noget om både ordets udtale, betydning og brug. En anden informants udsagn, om ordet

baukur, er derimod kun en meget kortfattet forklaring af dets betydning (dansk oversættelse): “= en beholder af metal.”

2.1.2. Digitalisering af arkivmaterialet

Digitalisering af materialet i ROH begyndte allerede i 1980'erne (jf. Svavarsdóttir et al. 1993, Prospectus 1993; se Svensén 2004:54 om digitalisering af leksikografiske seddelsamlinger alment). Det første skridt var en systematisk registrering og grov analyse af opslagsordene. Ordregistret gav en mere fleksibel adgang til materialet uafhængig af den alfabetiske organisering, som havde været den eneste tilgang til seddelsamlingen. Desuden fik man et vigtigt overblik over antallet af opslagsord og hvordan de fordelte sig på forskellige ordklasser (fx omtrent 590.000 substantiver og 17.000 verber). Således gav registret et vigtigt indblik i ordforrådets sammensætning ifølge arkivmaterialet. Samtidig blev der også oprettet et digitalt kilderegister over de skrifter som var blevet excerperede.

Næste skridt i arkivets digitalisering var indtastning af udvalgte citater så man systematisk kunne redigere materialet, bl.a. for at finde ud af hvor nyttige og fyldestgørende de foreliggende belæg viste sig at være i forbindelse med redigering af en historisk ordbog for islandsk. Mellem 1983 og 1993 blev der gjort forskellige forsøg i den retning (se Jónsson 1988), bl.a. blev der udgivet et prøvehæfte med over 100 redigerede verber (Svavarsdóttir et al. 1993; Prospectus 1993). Redigeringsarbejdet viste at selv om de gamle arkiver vanligvis gav et godt overblik over ordenes semantiske, morfologiske og syntaktiske karakter og udvikling, virkede de indimellem mangelfulde. Den digitale tekstsamling *Íslenskt textasafn* (ÍT, se afsnit 2.2.1), som var under udarbejdelse i 1990'erne, kunne bruges af redaktørerne til at bekræfte og eventuelt supplere den beskrivelse som tegnede sig på basis af skriftsprogsarkivet. Forsøg med redigering resulterede således i vigtig erfaringsopsamling og bedre forståelse af hvordan forskellige kilder kan spille sammen. Af adskillige årsager fortsatte redigeringen af det historiske ma-

teriale dog ikke på dette tidspunkt, bl.a. fordi OH fik midlertidig økonomisk støtte til at digitalisere det meste af citatsamlingen.

Selv om ROH-arkivet er stort, er strukturen forholdsvis enkel, og det kræver derfor kun en forholdsvis simpel database for at hver oplysningstype kan blive optegnet i eget felt. Opslagsord og kildeakronymer får et fast id-nummer som kobler citaterne til den relevante enhed i henholdsvis ordregistret (lemmalisten) og kilderegistret.

Skriftsprogsarkivet, ROH, havde længe været den vigtigste kilde for historiske undersøgelser af islandske ord og ordforråd, men seddelsamlingen var først og fremmest tilgængelig for forskere og universitetsstuderende. Ved systematisk digitalisering af arkivmaterialet, inklusive de fleste citater, i en database og tilkomsten af webben blev det til gengæld muligt at gøre det tilgængeligt og åbent for en langt større brugerkreds. Dermed kunne hvem som helst konsultere det direkte med hensyn til forskellige spørgsmål og interesser angående islandske ord, deres grammatiske kendetegn og anvendelse gennem de godt 450 år som kildematerialet dækker. I den forstand kan man sige at den digitale version af ROH allerede fungerer som en vigtig historisk sprogresurse.

Brugergørsefladen og den underliggende database er efterhånden blevet renoveret og udvidet, bl.a. ved en mere detaljeret analyse af citaterne og mere fleksible søgemuligheder, for bedre at kunne tilfredsstille brugernes forskellige behov. Omkring år 2000 blev databasen til ROH gjort alment tilgængelig og søgbar på webben, og for nylig blev både databasen og brugergørsefladen renoveret, bl.a. med mere fleksible søgemuligheder (Svavarsdóttir 2024:195ff.).

En del af materialet i TOH er allerede blevet digitaliseret, dvs. lemmalisten, informantregistret, brevsamlingen og en mindre del af kommentarerne (Svavarsdóttir & Kvaran 1994), selv om det kun er tilgængeligt internt. På grund af kommentarernes diversitet og materialets mangfoldighed er det kompliceret at oprette en vel-

struktureret database som rummer hele TOH-arkivet og knytter de forskellige dele sammen på en formålstjenlig og gennemskuelig måde. Der arbejdes nu på udvikling og opbygning af en sådan database, og i fortsættelsen er det planlagt at indføre kommentarerne og andet materiale tilknyttet samlingen. Således vil hele arkivet efterhånden blive let tilgængeligt, i det mindste for leksikografer og sprogforskere, og det diskuteres også hvordan udvalgte dele af materialet kunne præsenteres på en interessant måde for almenheden.

2.2. Sprogkorporusser

På OH, og senere ÁMI, har man opbygget et antal digitale tekstsamlinger og sprogkorporusser til brug for islandsk leksikografi, sprogforskning og sprogteknologi.

2.2.1. Foreliggende korpusser

Den ældste digitale tekstsamling er *Íslenskt textasafn* (ÍT), som nu består af 65 millioner løbende ord fra gamle, men dog mest nyere tekster af forskellig slags. Den er søgbar på webben, men kun som ren tekst. Teksterne i ÍT er helt uden annotering, og eksempel materialet fra et sådant korpus bliver hurtigt uoverskueligt, især når det drejer sig om undersøgelsen af forholdsvis almindelige ord (se fx Trap-Jensen 2014:49). Den begrænsning gælder derimod ikke teksterne i nyere sprogkorporusser, hvor teksterne er grammatisk taggede og lemmatiserede, og hvor man derfor har flere og mere fleksible søgemuligheder, fx efter opslagsord, ordformer eller orddele og/eller efter forskellige grammatiske kendetegn.

Det største islandske korpus er *Risamálheildin* (RMH), et gigakorpus som består af 2,8 milliarder tokens (se Steingrímsson et al. 2018), for det meste fra forholdsvis nye tekster. Ved opbygningen af korpusset blev der fokuseret på mængden, dvs. at få fat i så mange tekster der allerede forelå digitalt, som muligt. Trap-Jensen

(2014:53) har påpeget at det netop er kendetegnende for de store korpusser “at de ikke er særlig godt balancerede.”

Et mindre korpus, men mere balanceret med hensyn til materialets mangfoldighed, er det såkaldte *Mörkuð íslensk málheild* (MÍM) som havde det britiske BNC som det vigtigste forbillede. MÍM består af 25 millioner tokens i forskellige tekster fra 2000-2010 (Helgadóttir et al. 2012) som både er grammatisk taggedede og lemmatiserede. Disse tekstsamlinger og korpusser har været en vigtig kilde ved udarbejdelsen af forskellige sprogresurser, bl.a. ordbøger over samtidssproget, den islandske bøjningsbeskrivelse *Beygingarlýsing íslensks nútímamáls* (BÍN) og *Íslenskt orðanet*, dvs. et islandsk ordnet. Derimod repræsenterer kun en brøkdel af korpusmaterialet ældre sprogbrug så disse resurser er til begrænset nytte ved en historisk beskrivelse af sproget og ordforrådets udvikling over længere tid.

2.2.2. Korpusser under udvikling

I dag er der tre nye sprogkorpusser bestående af ældre tekster på tegnebordet på ÁMI. Sammenlagt vil de dække samme tidsrum som skriftsprogsarkivet, dvs. fra ca. 1550 til 2000, og delvis også de samme kilder. Tabel 2 viser en oversigt over de korpusser der er under udvikling.

Korpus	Indehold	Omfang	Status
Historisk korpus	Håndskrifter, dokumenter og bøger fra 1550-1800	Endnu uklart	Forberedes; materialeindsamling
1800-tals korpus	Aviser og tidsskrifter	270 mio. tokens	Færdig i foreløbig udgave på webben
1900-tals korpus	Bøger	153 mio. tokens	Under udarbejdelse; begrænset adgang

Tabel 2: Nye korpusser der indeholder tekster fra 1550-2000.

Det historiske korpus der er planlagt, vil dække perioden fra midten af 1500-tallet og frem til 1800, dvs. de omtrent 260 år som svarer til den ældre del af ROH. Der foreligger kun få allerede digitaliserede tekster fra denne periode, så de må enten indtastes eller skannes og gøres læsbare med OCR. Digitalisering af gamle tekster er ikke en nem opgave, fx stammer mange tekster fra håndskrifter eller bøger trykt med fraktur, og typisk er der stor variation i skriftformer både inden for samme tekst og mellem tekster. Korpusset bliver på to niveauer, det ene med originalteksten (med latinske tegn) og det andet med en ortografisk standardiseret version for at lette automatisk analyse og søgning. Tekstbehandlingen er således kompliceret og tidskrævende, og det er endnu uklart hvor stort korpusset vil blive, og hvornår det bliver færdigt.

1800-talskorpusset består af tekster fra aviser og tidsskrifter, som på forhånd forelå digitalt (oprindeligt fra databasen *Tímarit.is*), og fejl i OCR-behandlingen var allerede til dels blevet korrigeret manuelt eller halvautomatisk. Korpusset består af 270 millioner tokens og vil indbefatte både en bogstavret version og en standardiseret version af teksterne. Det foreligger i skrivende stund i en foreløbig udgave.

Det sidstnævnte korpus, 1900-talskorpusset, består af tekster fra trykte bøger, både skøn- og faglitterære. Korpusset dækker det 20. århundrede, især før it-teknologiens indtog i 1980'erne. På grund af ophavsret kan det kun anvendes til forskningsformål.

3. Sammenligning af sprogarkiver og digitale sprogkorpusser

3.1. Resursernes forskellige natur og nytte

I afsnit 2 blev de vigtigste leksikografiske og sproglige resurser for islandsk præsenteret. Her vil vi fokusere nærmere på deres

indhold med hensyn til hvordan de kan spille sammen og opveje hinandens begrænsninger for at nå frem til en så detaljeret og pålidelig beskrivelse af ordenes grammatiske, semantiske og pragmatiske egenskaber som mulig.

3.1.1. Arkiver og korpusser som ordbogsressurser

ROH består af citater som er blevet exciperet manuelt, dvs. at mennesker har læst kilderne og udvalgt ord og udtryk som de har anset som interessante og informative. Ved excipering er der dog en fare for tilfældighed ved valg af citater, bl.a. fordi man har tendens til lettere at få øje på det usædvanlige end det almindelige, og således kan frekvente udtryksmåder blive overset. Dette kan lede til en ubalance i belægsamlingen for bestemte ord og reducere dens repræsentativitet (jf. Hovmark 2012:297-299), bl.a. ved at “ge en oproportionerligt stor tyngd åt det ovanliga och perifera språkbruket” (Svensén 2004:57). Digitale tekstsamlinger og korpusser består derimod af sammenhængende tekster, ofte hele værker, og derfor kan man i princippet finde alle eksempler af et ord som forekommer i teksterne i deres forskellige former og sammenhænge (ordforbindelser, konstruktioner osv.). Korpusserne kan således være nyttige til at supplere belæggene som findes i arkiverne (jf. Pind et al. 1993:14-15; Hovmark 2012), ikke mindst når det gælder meget almindelige og frekvente ord som gerne er underrepræsenterede i seddelsamlingerne.

Hvis et korpus skal være hovedgrundlag for en leksikografisk beskrivelse, må den dog opfylde en række krav til dets sammensætning, især teksternes variation med hensyn til alder, type, emne, stil osv. Desuden vil det være et krav at teksterne er godt analyseret, og at søgesystemet er både meget effektivt og fleksibelt. Det er også vigtigt at leksikograferne har grundigt kendskab til søgesystemets muligheder, og at de til deres rådighed har et godt redigeringsystem, helst med direkte kobling til korpusset.

3.1.2. Emneområde og tekstgenrer

Tidsmæssigt indeholder de eksisterende korpusser næsten udelukkende nyere tekster, mest fra 2000-tallet, selv om der findes enkelte ældre tekster i ÍT og i mindre grad i RMH. Men nu forberedes der nye historiske korpusser som tilsammen vil dække samme periode som ROH (se afsnit 2.2). Materialeudvalget i de fleste korpusser har hidtil beroet på at teksterne allerede fandtes digitalt, og det præger deres sammensætning med hensyn til emneområde og stil, selv hvis man har stræbt efter en så varieret sammensætning som muligt. MÍM-korpusset kommer nærmest en bred og balanceret dækning med hensyn til emne og teksttype, mens RMH hovedsagelig består af ikke-skønlitterære tekster, hvoraf omtrent halvdelen er nyheder (fra aviser, radio og tv), og andre store tekst-kategorier er sociale medier, administration (især fra *Alþingi* og domstole) og lokale eller faglige tidsskrifter. Andelen af bøger, bl.a. skønlitteratur og videnskabelige tekster, er derimod lille.

Problemer ved indsamling af materiale til skriftsprogsarkivet, ROH, og de planlagte historiske korpusser er anderledes. Her er udvalget af mulige tekster begrænset af hvad der blev skrevet, og hvad der er bevaret fra forskellige perioder. I alt blev der excerperet fra omtrent 2.000 titler til ROH, og disse kilder inddeles i mere end 30 kategorier med hensyn til teksttype og emneområde. Teksternes antal og mangfoldighed varierer imidlertid meget efter periode. Således excerperedes størstedelen af de eksisterende værker fra 1600-tallet, selv om der kun er 32 og de fleste er religiøse. Derimod er der blevet excerperet over 1.000 titler fra 1900-tallet, omfattende næsten alle kategorier (Svavarsdóttir et al. 1993:3), men som alligevel kun er et udvalg af det som publiceredes i perioden.

Det begrænsede antal af eksisterende tekster fra gammel tid betyder at tekstmaterialet i de nye historiske korpusser uundgåeligt kommer til at overlappe det som blev excerperet for ROH, især fra perioden før ca. 1800. Desuden bliver 1800- og 1900-talskorpusserne begrænset til bestemte teksttyper. Det førstnævnte vil

udelukkende bestå af aviser og tidsskrifter, mange af dem allerede excerperede for ROH, mens det sidstnævnte bygger på skanning af trykte bøger. Mange bøger i 1900-talskorpusset tilhører de samme kategorier som de excerperede tekster til ROH, mens andre repræsenterer teksttyper som ikke fylder meget i ROH, fx populærlitteratur, børnebøger og oversættelser. Til trods for overlappning i kildegrundlaget bliver de historiske korpusser et vigtigt supplement til arkivmaterialet.

3.1.3. Talesprog

Uanset resursetype er talesproget stærkt underrepræsenteret. Det meste af tiden har der ikke været metoder til at bevare talesprog, og der findes ingen direkte historiske kilder om det før 1900-tallet. Talesprogsarkivet, TOH, retter sig mod daglig tale, især dagligdags ord, ordforbindelser og ordbrug som sjældent forekommer i skriftlige kilder, og er således et vigtigt tillæg til andre resurser. Det består dog af informanternes kommentarer om ord og deres vurdering af hvad de betyder og hvordan de bruges, ikke af autentisk forekommende talesprog. De fleste informanter, i alt over 3.000 kvinder og mænd fra forskellige dele af landet, var født og opvokset i 1800- og 1900-tallet. TOH kan således anses som historisk, selv om tidsrummet er begrænset.

Hvad korpusserne angår, er det meget arbejds- og tidskrævende at indsamle og forberede store mængder talesprogsmateriale på en sådan måde at det kan lægges ind i et korpus. En lille samling af transskriberet spontant islandsk talesprog, svarende til ca. 50 timers optagelser, er dog inkluderet og søgbart både i MÍM og ÍT. Desuden findes der et korpus bestående af interview med unge mennesker, i alt omtrent 20 timer, som giver tilgang både til lyd og transskriptioner (Hilmisdóttir 2024:108-110). Selv om dette materiale ikke rækker langt i leksikografisk sammenhæng, kan det i nogle tilfælde supplere skriftsprogsilderne ved at fange udtryk som først og fremmest forekommer i tale (se Svavarsdóttir 2003).

3.2. Sociolingvistiske faktorer

TOH er i mange tilfælde en god kilde til sproglig variation som kan kobles til sociologiske faktorer. Ved hvert udsagn er der oplysninger om informanten, dvs. navn, geografisk herkomst, bopæl, køn osv., og arkivmaterialet kan således give oplysninger om sproglig variation som er knyttet til sociolingvistiske faktorer, ikke mindst dialektal udbredelse af ord, ordformer og ordbrug.

3.2.1. Dialektal ordbrug som kun findes i talesprog

Som et eksempel på at TOH er hovedkilde til et dialektalt ord, kan man tage ordet *della* i betydningen '(en type) stegt bagværk'. I arkivet er der 12 sedler hvor ordet er nævnt som et alternativ til *ástarþungur* med samme betydning. Ifølge den etymologiske ordbog svarer *della* muligvis til sidstedelen af *frikadella* 'kødbolle', et låneord fra italiensk gennem dansk (*frikadelle* < *frittadella*) (Magnússon 1989; jf. også Árnason 2011:151-152). Denne tolkning støttes af udtalen som ifølge kommentarer i TOH er langt [l:], typisk for bl.a. låneord, men ikke det hyppigere [tl] (som fx i homografen *della* 'nonsens, dille'). Alle belæg i talesprogsarkivet er knyttet til det nordlige Island, hvilket tyder på at ordet har haft en begrænset geografisk udbredelse. Desuden viser nogle af kommentarerne informanternes opfattelse af og holdninger til ordet, fx siger én at *della* blev anset som finere end det mere udbredte *ástarþungur* (antagelig med hensyn til sidstedelen *-þungur* 'þung'), og at det derfor mest blev brugt af "hispursdömmum", dvs. fine damer. En anden informant hævder at *della* var det almindelige ord, mens *ástarþungur* mest blev brugt ironisk.

Udsagnene i TOH kan således give værdifuld indsigt i sprogbruges opfattelse af ord og ordanvendelse, men de kan også være i modstrid med hinanden, så det er ofte svært at drage generelle konklusioner ud fra de få kommentarer som findes. Alligevel indeholder TOH vigtige og mangfoldige oplysninger, med hensyn

til *della* og mange andre ord, som næppe findes i skriftsproglige kilder.

3.2.2. Morfologisk variation i tale og skrift

Der er også eksempler på at ROH og TOH supplerer og støtter hinanden. Det viser fx oplysninger om homonymerne *skúr* ‘nedbør’ og *skúr* ‘(lille, enkel) bygning’. De fleste sprogbrugere anvender det første som femininum (def. *skúrin*, pl. *skúrir*) og det sidste som maskulinum (def. *skúrin*, pl. *skúrar*), men hos andre falder de to sammen morfologisk så de enten begge er i maskulinum eller begge i femininum (jf. Blöndal 1920-1924). Dette leder til variation i brugen, som ifølge Blöndal er regional: maskulinum af *skúr* ‘nedbør’ findes i nord- og østlandet, og femininum af *skúr* ‘bygning’ knytter sig til et område i Vestfjordene. Andre steder er homonymerne adskilte morfologisk så vel som semantisk.

Tabel 3 viser fordelingen mellem varianter ifølge sprogarkiverne (eksempler hvor kønnet ikke er utvetydigt, medregnes ikke).

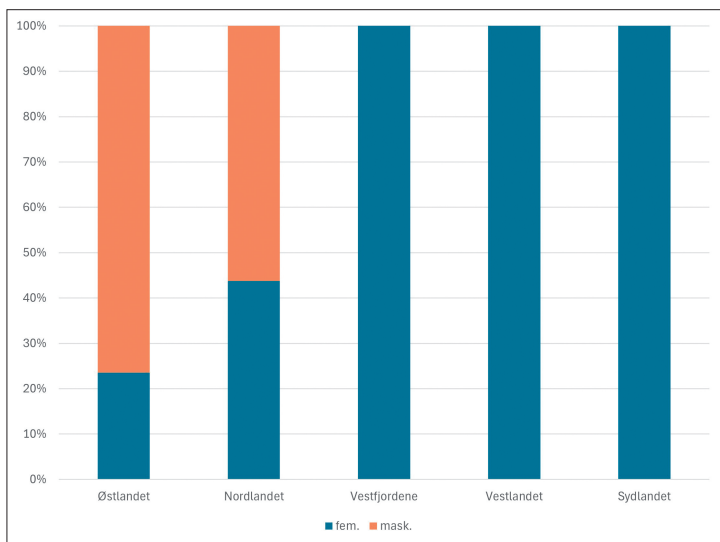
	<i>skúr</i> ‘nedbør’		<i>skúr</i> ‘bygning’	
	fem.	mask.	fem.	mask.
ROH	38 (70 %)	16 (30 %)	4 (19 %)	17 (81 %)
TOH	23 (51 %)	22 (49 %)	5 (45 %)	6 (55 %)

Tabel 3: Køn af homonymer i skrift- og talesprogarkiverne.

Det fremgår af Tabel 3 at materialet fra de to arkiver stort set stemmer overens med Blöndals beskrivelse fra 1920'erne. I begge homonymer er der variation i køn således at femininum dominerer i *skúr* ‘nedbør’, men maskulinum i *skúr* ‘bygning’. Jævnere distribution i TOH kan skyldes indsamlingsmetoden, dvs. at den sjældnere variant har været i fokus i leksikografernes forespørgsel. Det ville være tidskrævende at komme frem til sådanne oplysninger.

ger i store korpusser, fx i RMH's 13.000 og 17.000 belæg om *skúr* i femininum respektive maskulinum, fordi homonymer af samme ordklasse ikke kan adskilles i søgningen.

Oplysninger om geografisk udbredelse af varianter ville man først lede efter i TOH. Figur 1 viser hvordan brugen af maskulinum og femininum af *skúr* 'nedbør' fordeler sig efter region ifølge informanterne (hvis de har meddelt at begge køn var brugt i deres hjemegn, var begge registreret).



Figur 1: Distribution af femininum og maskulinum af ordet *skúr* 'nedbør' ifølge talesprogsarkivet (oplysninger fra 45 informanter).

Figur 1 tyder på at i syd- og vestlandet samt Vestfjordene bruges *skúr* 'nedbør' udelukkende som femininum (*skúrin*). I nord- og østlandet er maskulinum (*skúrinn*) derimod dominerende, men femininum er også kendt og brugt. Nogle informanter hævdede at maskulinum havde været hyppigere før i tiden, og det tyder på at udbredelsen af maskulinum kan være på retur.

Der er for begrænsede oplysninger om *skúr* 'bygning' i TOH til

at vise en landsdækkende distribution af varianterne. Der findes fem udsagn om brugen af femininum, og interessant nok kommer fire af dem fra samme område som Blöndal (1920-1924) knyttede varianten til, dvs. de nordlige Vestfjorde. I tillæg er der fire belæg i ROH på femininum, alle fra kilder som stammer fra Vestfjordefne. Således supplerer og støtter arkiverne hinanden og de meget begrænsede oplysninger om denne variants udbredelse i Blöndal (1920-1924).

3.3. Skriftsprogsarkivet som sprogteknologisk resurse

Den digitale version af ROH kan anses som et korpus med sine ca. 2 millioner citater indlagt i en systematisk opbygget database hvor forskellige oplysningstyper står i adskilte felter. Hvert belæg er knyttet til et opslagsord (i standardiseret nutidsform), og det kan således siges at den pågældende ordform i citatet er blevet lemmatiseret. Opslagsordene er desuden markeret med hensyn til ordklasse (plus genus for substantiver) og morfologisk struktur (simpleks, afledning med præfiks og/eller suffiks, sammensætning). I hvert belæg er ordformen afgrænset og opført bogstavret i et særfelt i databasen, og i et andet felt indføres den tilsvarende standardiserede nutidsform. I tillæg er hvert citat, og dermed den tilhørende skrift- og bøjningsform af opslagsordet, tidsfæstet. Til sidst er belægget knyttet til sin kilde.

Den analyse og kategorisering af historisk sprogmateriale som på denne måde er systematisk registreret i databasen og hovedsageligt bygger på manuel analyse, kan forventes at være nyttig ved opbygning af historiske sprogkorporer som dækker samme tidsrum som ROH. Det kan navnlig gælde ved rettelser i teksterne, deres standardisering, samt grammatisk tagging og lemmatisering af ordformer. På den måde kan det gamle skriftsprogsarkiv med de yderligere analyser som er blevet tilført i den digitale version, blive et vigtigt sprogteknologisk hjælpemiddel ved opbygning af

nye, historiske korpusser, især ved en (halv)automatisk analyse af teksterne. Samtidig vil sådan et nyt korpus blive et vigtigt supplement til ROH som kilde til en historisk, leksikografisk beskrivelse af islandsk.

4. Konklusioner

Udgangspunktet i denne artikel var at diskutere og vurdere den leksikografiske værdi af store historiske sprogarkiver, bygget op på basis af excerpering af tekster, vis-a-vis moderne digitale sprogkorpusser med sammenhængende tekster. Foruden tekstsamlinger og korpusser med nutidssprog, som længe har eksisteret, arbejdes der nu på oprettelse af historiske korpusser som dækker helt eller delvis samme periode som arkiverne. Dette giver anledning til at spørge: Vil de nye sprogresurser afløse de gamle, eller kan de fungere side om side og støtte hinanden? Det har vi drøftet i artiklen med hensyn til forskellige islandske sprogresurser og deres betydning for leksikografi og sprogforskning.

I en artikel om brugen af seddelsamlinger og korpusser ved redigering af *Ømålsordbogen* skriver Hovmark (2012:298): “det er fristende og nærliggende at anse seddelsamlinger for at være et ringe(re) kildegrundlag end tekstkorpora, bl.a. fordi korpora rummer mulighed for kvantitative undersøgelser og giver et bedre indblik i fx syntaktiske forhold og mere upåfaldende ord og betydninger der ikke altid er blevet indstreget når man har excerperet til en seddelsamling.” De to slags resurser viser imidlertid sprogets egenskaber på forskellig vis og supplerer hinanden: Korpusserne indeholder mange eksempler på almindelige ord og sprogbrug, mens seddelsamlingerne gerne fremhæver det som er sjældent forekommende (Svensén 2004:57). Hovmark (2012:305) konkluderer at begge resurser har selvstændig værdi for leksikografen, og at spørgsmålet først og fremmest drejer sig om kildekritik og

repræsentativitet. Leksikografen må huske at begge datatyper er “et udsnit af virkeligheden som skal vurderes kritisk som kildegrundlag”. Vi er kommet til lignende konklusion, dvs. at både de traditionelle arkiver og de digitale korpusser udgør værdifulde ressourcer for islandsk leksikografi og sprogforskning.

ROH holder sin værdi som historisk oversigt over ordforrådet og er en vigtig kilde til det islandske sprogs udvikling fra det 16.-20. århundrede. Den digitale version er nu i forholdsvis fin stand med nyligt renoveret database og grænseflade, tilføjelse af citater fra seddelsamlingen, rettelser og standardisering. Yderligere analyse og kategorisering af citaterne med hensyn til strukturelle og/eller semantiske egenskaber ville, med hensyn til fortsat anvendelse af arkivet, være ønskværdig. I det arbejde må der tages hensyn både til de forskellige brugeres behov, og til at materialet kan være til nytte ved opbygning af historiske sprogkorpusser, fx ved tilpasning af sprogteknologiske værktøjer til automatisk analyse af ældre tekster.

TOH indeholder forskellige oplysninger som ikke findes i andre kilder, og derfor er det en uvurderlig sprogresurse. Arbejde på digitalisering af materialet er nu i gang, og en skræddersyet database er under udvikling. Desuden diskuteres det hvordan udvalgte oplysninger fra samlingen kan præsenteres på en interessant måde, fx ved hjælp af kort som viser udbredelsen af sprogvarianter.

Således vil arkiverne efter vores vurdering fortsat udgøre vigtige kilder til vores viden om islandske ord, ordbrug og ordforråd, og samtidig er tilkomsten af nye historiske sprogkorpusser både nyttig og ønskværdig. Digitale sprogkorpusser har allerede bevist deres værdi ved undersøgelse og leksikografisk analyse af nutidsproget, og brugen af de forholdsvis få ældre tekster i eksisterende digitale tekstsamlinger har vist at de er et vigtigt supplement til arkivmaterialet.

Litteratur

Ordbøger, korpusser og digitale resurser

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Árni Magnússon-instituttet for islandske studier. <<https://bin.arnastofnun.is>> (marts 2025).

Blöndal, Sigfús (1920-1924): *Íslensk-dansk Ordbog*. Reykjavík. <<https://blondal.arnastofnun.is>> (marts 2025).

BNC = *British National Corpus*. University of Oxford. <<http://www.natcorp.ox.ac.uk>> (marts 2025).

Íslenskt orðanet. Jón Hilmar Jónsson (red.). Árni Magnússon-instituttet for islandske studier. <<https://ordanet.arnastofnun.is>> (marts 2025).

ÍT = *Íslenskt textasafn*. Árni Magnússon-instituttet for islandske studier. <<https://corpus.arnastofnun.is>> (marts 2025).

Magnússon, Ásgeir Blöndal (1989): *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.

MÍM = *Mörkuð íslensk málheild*. Árni Magnússon-instituttet for islandske studier. <<https://malheildir.arnastofnun.is>> (marts 2025).

RMH = *Risamálheildin* (2024). Árni Magnússon-instituttet for islandske studier. <<https://malheildir.arnastofnun.is>> (marts 2025).

ROH = *Ritmálssafn Orðabókar Háskólans*. Árni Magnússon-instituttet for islandske studier. Digital version: <ritmalssafn.arnastofnun.is> (marts 2025).

Tímarit.is. Landsbókasafn Íslands Háskólabókasafn. <<https://tímarit.is>> (april 2025).

TOH = *Talmálssafn Orðabókar Háskólans*. Árni Magnússon-instituttet for islandske studier.

Ømålsordbogen. *En sproglig-saglig ordbog over dialekterne på Sjælland, Lolland-Falster, Fyn og omgivende øer 1- (1992-)*. Køben-

havn: Center for Dialektforskning, Københavns Universitet, og
Universitets-Jubilæets danske Samfund.

Anden litteratur

- Árnason, Mördur (2011): Ásgeir, Orðsifjabókin og endurútgáfa Íslenskrar orðabókar 1983. I: *Orð og tunga* 13, 145-171.
- Helgadóttir, Sigrún, Ásta Svavarsdóttir, Eiríkur Rögnvaldsson, Kristín Bjarnadóttir & Hrafn Loftsson (2012): The Tagged Icelandic Corpus (MÍM). I: *Proceedings of the Workshop on Language Technology for Normalisation of Less-Resourced Languages – SaLTMiL 8 – AfLaT 2012*. Istanbul. 67-72.
- Hilmisdóttir, Helga (2024): Talspråkskorpuser som resurs för isländska ordböcker. I: *LexicoNordica* 31, 103-125.
- Hovmark, Henrik (2012): Data og repræsentativitet i ordbogsarbejdet. I: Birgit Eaker et al. (red.): *Nordiska studier i lexikografi* 11. Lund: Nordiska föreningen för lexikografi. 296-308.
- Hovmark, Henrik (2023): Seddelsamlinger – historisk arkivmateriale eller levende resurse? I: Louise Holmer et al. (red.): *Nordiska studier i lexikografi* 16. Lund och Göteborg: Nordiska föreningen för lexikografi. 165-176.
- Ingólfsson, Gunnlaugur (1988): Söfnun Orðabókar Háskólans úr mæltu máli. I: *Orð og tunga* 1, 65-72.
- Jónsson, Jón Hilmar (1988): Sagnorðagreining Orðabókar Háskólans. I: *Orð og tunga* 1, 123-174.
- Pind, Jörgen, Kristín Bjarnadóttir, Jón Hilmar Jónsson, Guðrún Kvaran, Friðrik Magnússon & Ásta Svavarsdóttir (1993): Using a Computer Corpus to Supplement a Citation Collection for a Historical Dictionary. I: *International Journal of Lexicography* 6/1, 1-18.
- Prospectus (1993) = *A Prospectus to a Dictionary of Icelandic Verbs. English Version of the Introduction*. Reykjavík: Institute of Lexicography.

- Steingrímsson, Steinþór, Sigrún Helgadóttir, Eiríkur Rögnvaldsson, Starkaður Barkarson & Jón Guðnason (2018): Risamálheild: A Very Large Icelandic Text Corpus. I: *Proceedings of LREC 2018*. Miyazaki. 4361-4366.
- Svavarsdóttir, Ásta (2003): Ordbogen og den daglige tale. Om den islandske talesprogsbank (ISTAL) og dens betydning i ord-bogsredaktion. I: Zakaris Svabo Hansen & Anfinnur Johansen (red.): *Nordiske Studier i Leksikografi* 6. Tórshavn: Nordisk for-ening for leksikografi. 43-48.
- Svavarsdóttir, Ásta (2024): Ritmálssafn Orðabókar Háskólans. Samtíð og saga. I: *Orð og tunga* 26, 185-210.
- Svavarsdóttir, Ásta & Guðrún Kvaran (1994): The computeriza-tion of the spoken language archive. I: Karl Hyldgaard-Jensen & Viggo Hjørnager Pedersen (red.): *Symposium on Lexicogra-phy* VI. Tübingen: Max Niemeyer Verlag. 127-139.
- Svavarsdóttir, Ásta, Guðrún Kvaran, Jón Hilmar Jónsson & Kristín Bjarnadóttir (1993): *Sýnihefti sagnorðabókar*. Reykjavík: Orðabók Háskólans.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ord-boksarbete i teori och praktik*. Stockholm: Nordsteds.
- Trap-Jensen, Lars (2014): Leksikografisk tradition og fornyelse: tre revolutioner på 100 år? I: Ruth Vatvedt Fjeld & Marit Hov-denak (red.): *Nordiske studier i leksikografi* 12. Oslo: Novus forlag. 42-68.

Ásta Svavarsdóttir
forskningsdocent (emerita)
Árni Magnússon-instituttet for
islandske studier
Edda, Arngrímögötu 5
IS-107 Reykjavík
asta.svavarsdottir@arnastofnun.is

Helga Hilmisdóttir
forskningsdocent, ph.d.
Árni Magnússon-instituttet for
islandske studier
Edda, Arngrímögötu 5
IS-107 Reykjavík
helga.hilmisdottir@arnastofnun.is

Fremtidens ordbøger: udfordringer og tilpasninger

Lars Trap-Jensen

The article examines some of the most pressing challenges confronting contemporary monolingual lexicography. Monolingual lexicography is probably less challenged than bilingual, but even so it is necessary to adapt the dictionary data in order to remain competitive with emerging technologies. Two key factors are investigated: the tendency towards integrated solutions over traditional stand-alone dictionaries and the impact of large language models on dictionaries and the strategies dictionary projects should adopt in response. A central question is whether artificial intelligence should be regarded as a competitor to human editors or if it can be effectively integrated into editorial workflows to enhance linguistic descriptions.

1. Indledning

Hvis ordbøger primært betragtes som sproglige værktøjer, kan man godt forestille sig mere effektive måder at formidle information om ords betydninger i konkrete tekster end at lade brugeren slå dem op i en ekstern ordbogsressource. Et godt og aktuelt eksempel er muligheden for at integrere ordbogsopslag direkte i e-bogslæsere eller på hjemmesider, ikke mindst hvis det bliver teknisk muligt at få programmet til at udvælge den betydning der er relevant for brugeren i sammenhængen, så det alene er den der vises.

En sådan udvikling vil ikke overflødiggøre den traditionelle ordbogs samlede fremstilling af ords betydninger og anvendelser. Den vil eksempelvis stadig være relevant for fremmedsprogs elever eller inden for forskning og undervisning i sprog. Men hvis inte-

grerede løsninger bliver standarden, vil det sandsynligvis gøre den traditionelle ordbog mere nichepræget. Hvis fremtidens ordbogsbaser skal forblive relevante, bliver det derfor nødvendigt at organisere data på nye måder så den integrerede brug understøttes.

I denne artikel fokuserer jeg på to områder der fremstår særlig centrale og relevante i lyset af denne udfordring. Det drejer sig for det første om de etsprogede ordbøger og modersmålsleksikografien, herunder deres fortsatte relevans og potentielle tilpasningsmuligheder, og for det andet diskussionen af de teknologiske udfordringer som kunstig intelligens og avancerede sprogmodeller medfører for leksikografien og ordbøgerne.

2. Etsprogsleksikografi

Hvis tendensen til stigende atomisering af ordbogsoplysninger forstærkes i takt med at data bruges i andre sammenhænge end den traditionelle ordbogsvisning, er det nødvendigt at ordbogsredaktionerne tilpasser data til denne nye virkelighed. Trap-Jensen & Lorentzen (2024) tog dette emne op ved at undersøge udvalgte definitionspraksisser som kunne påvises i den digitale udgave af DDO. Undersøgelsen var koncentreret om to hovedtyper benævnt henholdsvis *morfosemantisk lighed* og *nedarvning*. Disse gennemgås kort herunder.

2.1 Morfosemantisk lighed

Morfosemantisk lighed er det der kendetegner grupper af ord der er morfologisk beslægtede og har en fælles semantisk kerne. Et eksempel er tripletter som *buddhisme*, *buddhist* og *buddhistisk*. Traditionelt har sådanne ordgrupper kun haft én hovedbeskrivelse, på den måde at kun hovedbegrebet *buddhisme* får en fuldt udfoldet forklaring, mens de afledte ord ikke forklares selvstændigt.

Buddhist defineres således blot som en person der bekender sig til buddhisme, og *buddhistisk* angives som et adjektiv der vedrører hvad der har med buddhisme at gøre, men brugerne må selv slå op hvad *buddhisme* betyder, hvis de ikke kender ordet.

Denne skabelon fungerede relativt godt i papirordbogens tid, hvor artiklerne var alfabetiske naboer, og hvor det var vigtigt at spare værdifuld plads ved ikke unødigt at gentage samme lange forklaring ved hvert af ordene. Men i den digitale ordbog fungerer det mindre godt (jf. Atkins & Rundell 2008:436: "As all user research shows, it is never a good idea to require users to go to a second (or third) entry to find the information they are looking for in a first"), og skal forklaringerne kunne fungere helt uafhængigt af hinanden, er det nødvendigt at finde andre løsninger. Et eksempel på en uafhængig brug er e-bogslæsere hvor brugere med et dobbeltklik kan slå alle ord op og se ordbogens definition. Hvis brugeren slår *buddhist* op, er en forklaring som 'person der bekender sig til buddhismen' utilstrækkelig. Hvis man har slået *buddhist* op, er det usandsynligt at man ved hvad *buddhisme* er, og forklaringen på det ord er ikke umiddelbart tilgængelig.

I analysen fremgik det at DDO's definitionspraksis for denne type fordelte sig på fire forskellige skabeloner:

1. *Den traditionelle tilgang*, som undertiden har overlevet fra papirordbogens tid.
2. *Minimalforbedringen*, hvor hovedordet gøres klikbart så brugeren kan komme til den fulde forklaring.
3. *Kombinationstilgangen*, hvor klikmuligheden suppleres med en kort forklaring der i mange tilfælde vil være tilstrækkelig for brugeren.
4. *Den fulde udfoldning*, hvor hvert lemma har en selvstændig og fuldt udfoldet forklaring i alle artikler.

De fire skabeloner kan illustreres med eksempler som i tabel 1.

Skabelon (1) <i>darwinisme</i>: udviklingslære baseret på Charles Darwins teorier om plante- og dyrearternes gradvise udvikling ved naturlig favorisering af individer med bestemte egenskaber <i>darwinist</i>: person der går ind for darwinismen <i>darwinistisk</i>: vedr. eller tilhørende darwinismen
Skabelon (2) <i>dualisme</i>: filosofisk opfattelse eller teori der opfatter tilværelsen som styret af to fundamentale, modsatrettede principper <i>dualist</i>: tilhænger af <<u>dualismen</u>> <i>dualistisk</i>: vedr. eller præget af <<u>dualisme</u>>
Skabelon (3) <i>universalisme</i>: princip eller lære (fx inden for religion eller politik) der betoner almenvellets eller helhedens forrang for den enkelte eller det specifikke <i>universalist</i>: person der er tilhænger af <<u>universalismen</u>>, dvs. tror på almenvellets eller helhedens forrang for den enkelte eller det specifikke <i>universalistisk</i>: vedr. eller karakteristisk for <<u>universalisme</u>>, dvs. det princip at almenvellet eller helheden har forrang for den enkelte eller det specifikke
Skabelon (4) <i>ekstremisme</i>: yderligtgående adfærd eller holdning, især inden for politik eller religion <i>ekstremist</i>: person med yderligtgående synspunkter eller adfærd, især inden for politik eller religion <i>ekstremistisk</i>: som udviser yderligtgående synspunkter eller adfærd, især inden for politik eller religion

Tabel 1: Fire skabeloner for morfosemantisk lighed.

2.2 Nedarvning

Pladshensyn gør sig også gældende for den anden hovedtype, *nedarvning*. Med nedarvning menes at en underbetydning forudsætter, det vil sige arver, definitionen fra en overordnet betydning. Figur 1 viser to eksempler: Dansen *bossanova* er udskilt som en særbetydning fra musikken, og i tilfældet *ramie* er plantefibre der anvendes i tekstiler, udskilt som en særskilt betydning af planten.

bossanova substantiv, fælleskøn
 uofficiel, men meget almindelig form: **bossa nova**
 BØJNING -en, -er, -erne
 UDTALE [basa'no:va] 🗣️ ⓘ
 OPRINDELSE kendt fra 1962 * fra portugisisk *bossa nova* 'den nye stemme, den nye rytme'

Betydninger

1. jazzpåvirket musik der er rytmisk beslægtet med samba – også om musikstykke af denne art * opstået i Brasilien sidst i 1950'erne
 ORD I NÆRHEDEN swingmusik | bigbandmusik | bebop | bop | avantgardejazz | freejazz...vis mere
 For mig er hverken repertoireet eller grundrytmen afgørende. Jeg synes også, at bossa nova og samba er jazz JyP1992

1.a dans til denne musik
 ORD I NÆRHEDEN mambo | lambada | salsa | merengue | beguine...vis mere

ramie substantiv, fælleskøn
 BØJNING -n
 UDTALE ['ʁa:mi] 🗣️ ⓘ
 OPRINDELSE via engelsk og fransk fra malajisk *rami* 'jute, hamp, groft linnedstof'

Betydninger

1. 1-2 m høj, flerårig plante med små grønne blomster og lange, stærke fibre der bruges til tekstilfremstilling * tilhører nældefamilien
 SYNONYMER kinagræs | Boehmeria nivea
 ORD I NÆRHEDEN hampeplante | jute | agave | sisal | raffiapalme | kapoktræ...vis mere

1.a fibre fra denne plante brugt i tekstiler
 Bluse af 55% ramie og resten bomuld www2000

Figur 1: Nedarvning: *bossanova* og *ramie* i DDO.

Traditionelt har denne type været løst med brug af en anaforsk reference (“... denne musik”, “... denne plante”), en løsning der kan fungere hvis – og kun hvis – de to betydninger ses sammen. Hvis underbetydningen derimod skal vises uafhængigt, må en anden løsning tages i brug. Ligesom ved den første hovedtype kunne vi identificere fire skabeloner for definitionspraksis i tilfælde af nedarvning:

1. *Den traditionelle tilgang*, med anaforisk reference som et levn fra papirordbogens tid.
2. *Gentagelsesskabelonen*, hvor opslagsordet i den overordnede betydning gentages. Ordet gøres evt. klikbart for at kompensere for cirkulariteten.
3. *Den uafhængige skabelon*, hvor den overordnede betydning udfoldes i underbetydningen.
4. *Den integrerede skabelon*, hvor flere betydninger er integreret eller antydet i én definition.

Eksempler på de fire skabeloner er givet i tabel 2.

Skabelon	Eksempel	Definition
(1)	<i>agat</i> (DDO)	stykke af dette mineral, fx brugt som smykkesten
(2)	<i>lam</i> (DDO)	kød fra et lam
(3)	<i>lam</i> (DDNO)	kød fra et ungt får
(4)	<i>yak</i> (ODE)	a large domesticated wild ox with shaggy hair, humped shoulders, and large horns, used in Tibet as a pack animal and for its milk, meat and hide

Tabel 2: Fire skabeloner for nedarvning.

Man kan bemærke at skabelon 2 umiddelbart bryder med hvad der traditionelt har været regnet for god praksis fordi opslagsordet gentages i definitionen. Men tilgangen kan forsvares når det sker i en underbetydning, og det er indiskutabelt bedre end formuleringer som “dette dyr” eller “et sådant dyr”. Desuden er der mulighed for at gøre ordet klikbart hvis forklaringen optræder i isolerede kontekster.

Skabelon 3 repræsenterer en uafhængig definition, og den kan godt fungere hensigtsmæssigt, fx som vist i tabel 2's eksempel med *lam* fra *Den Danske Netordbog* (DDNO), men den kan virke over-

drevent omstændelig (og redundant i ordbogsvisningen) hvis der er tale om en kompliceret definition, og den stiller dermed større kognitive krav til brugeren end skabelon 2. Skabelon 4 indebærer en løsning hvor flere betydninger integreres i én definition – eller antydes indirekte. I tabel 2 er vist et eksempel fra *Oxford Dictionary of English* (ODE), hvor definitionen af *yak* ikke blot beskriver selve dyret, men også dets funktion som husdyr og kilde til kød, mælk og læder.

2.3 Konsekvenser for datastruktur

I Trap-Jensen & Lorentzen (2024) undersøgte vi for begge hovedtyper hvor udbredt de forskellige skabeloner er i to danske, en norsk, en svensk og en engelsk ordbog. Undersøgelsens resultater er beskrevet i Trap-Jensen & Lorentzen (2024). Relevansen i denne sammenhæng ligger i de konsekvenser som valg af definitionsstruktur har for organiseringen og udformningen af de leksikografiske data. Hvis det er en blivende tendens at ordbogsdefinitioner i stigende grad vises uafhængigt af den traditionelle ordbogspræsentation, må udformningen af ordbogsartikler ikke alene tilrettelægges efter målgruppen, men også efter publiceringskanalerne. Som vist indebærer det en betydningsbeskrivelse der kan stå alene.

I øjeblikket er det ikke sprogteknologisk muligt at genkende og vise den relevante betydning af et polysemt ord i løbende tekst. Derfor må integrerede ordbogsfunktioner i e-læsere og på hjemmesider nok fortsat præsentere hele artikler i et stykke tid endnu. Men vil man være forberedt på udviklingen, er det klogt at tilpasse data.

Morfosemantisk lighed og nedarvning er givet som eksempler, men de er ikke mere end netop eksempler. Der er andre områder der peger i samme retning. Et andet eksempel er faste vendinger der udviser betydelig variation.

slå en klo i
K

eller

slå/sætte klørne i
K

OVERFØRT tage eller få fat i; opnå besiddelse af; få i sin magt

ORD I NÆRHEDEN opsnappe | snuppe | tilsnige sig | gå på rov
 | gå/være på hugst | skrabe til sig...vis mere

Andy vil .. ind i vognen, men Kalle slår en klo i ham
 ThBirk85b

(varierer i form) Jalousien slår sin sorte klo i hende
 ChStra91

Figur 2: Udtrykket *slå en klo i* i DDO.

I figur 2 er vist et eksempel fra DDO hvor variationen angives med brug af “eller” og med skråstreg mellem de varierende enheder. Samtidig antydes det i kommentaren til citat nummer to at variationen er større end det der umiddelbart er fremhævet, og der vises en yderligere variant, her *slå sin sorte klo i nogen*.

Dette er løst anderledes i DDNO, som har valgt at opregne samtlige muligheder som selvstændige udtryk, jf. figur 3.

Det fremgår at selv flere betydninger af samme form er opregnet som selvstændige udtryk. Den løsning fungerer efter min mening ikke særlig hensigtsmæssigt i en ordbog. Den fører til uoverskueligt mange udtryk når der, som her, er mange faste udtryk under samme opslagsord, og det bliver derfor let uklart at der er tale om ét udtryk med to grundlæggende betydninger (‘gribe fat i’ og ‘få magt over’). Til gengæld er løsningen datastrukturelt velegnet til at matche en konkret streng i en tekst og give brugeren forklaring på netop denne streng, frem for at redegøre for udtrykkets variation. Det kan forklare hvorfor DDNO har valgt denne løsning. Metoden minder om princippet bag et morfologisk fuldformsleksikon, hvor målet er at registrere det maksimale

slå en klo i nogen
få magt over nogen

 slå en klo i nogen
gribe fat i nogen

 slå kloen i nogen
gribe fat i nogen

 slå kloen i nogen
få magt over nogen

 slå klørerne i nogen
gribe fat i nogen

 slå klørerne i nogen
få magt over nogen

 slå sine klør i nogen
gribe fat i nogen

 slå sine klør i nogen
få magt over nogen

Figur 3: Udtrykket *slå sin klo i nogen* i DDNO.

antal former for at kunne genkende specifikke bøjningsformer og knytte dem til et lemma. Det er nærliggende at pege på en tilsvarende strategi for udviklingen af et tilsvarende fuldformsleksikon for faste udtryk. På den måde ville man nemt kunne finde et bestemt udtryk og evt. nøjes med at vise den relevante variant med tilhørende forklaring ved søgning via en API-grænseflade i en integreret ordbog. Derimod er det nok ikke en god idé at vise hele inventaret for brugerne i ordbogsvisningen.

Sammenfattende ser jeg tegn på at leksikografien bevæger sig væk fra de traditionelle ordbøger som den eneste publiceringskanal og i retning af integrerede løsninger. Skal disse fungere optimalt, stiller det krav om fleksible leksikalske databaser, hvis konkrete anvendelser ikke nødvendigvis kan forudsiges. Det kræver at

data organiseres på en anden måde end i den traditionelle ordbog. Jeg har peget på nogle vigtige (men ikke nødvendigvis udtømmende) områder der trænger sig på: Definitioner må i højere grad kunne fungere uafhængigt af andre dele af en ordbogsartikel, og flerordsforbindelser bør registreres i al deres variation af hensyn til genkendelse. Men det betyder ikke at alt skal vises for brugerne. Også præsentationen kan være fleksibel så der tages hensyn til fx traditionelle visninger i ordbogsform eller forskellige overgangsløsninger hvor den traditionelle løsning udvides med link eller delvis udfoldning.

3. Kunstig intelligens

Det er vanskeligt at diskutere fremtidens udfordringer uden at inddrage den seneste revolution inden for leksikografien: Hvordan vil chatbotter og AI-assistenter påvirke både arbejdsmetoder og leksikografiske produkter? Kunstig intelligens som disciplin er i sig selv ikke ny; feltet har eksisteret siden 1950'erne, men indtil for nylig har det primært været et felt forbeholdt en snæver kreds af dataloger, sprogteknologer og computeringeniører. Først med lanceringen af ChatGPT i november 2022 blev teknologien for alvor bredt ud til offentligheden, og siden er udviklingen accelereret markant. Det er formentlig også den udvikling der mere end noget andet er bestemmende for hvilke udfordringer, tendenser og perspektiver der aktuelt tegner sig for leksikografien.

Siden lanceringen af ChatGPT er der fremkommet en betydelig mængde forskning og debat om emnet. Konferencen eLex 2023 havde fx dette som tema, og derfra stammer en større samling undersøgelser og overvejelser. I forbindelse med konferencen Lexicography in the XXI Century i 2023 foretog jeg selv som en del af min forberedelse en systematisk afprøvning af ChatGPT i mit daglige ordbogsarbejde gennem det meste af et år. På baggrund af

disse erfaringer samt en gennemgang af den eksisterende litteratur har jeg samlet hvad jeg mener at kunne konkludere på nuværende tidspunkt om ChatGPT's præstationer som leksikografisk værktøj (Trap-Jensen 2024). Den der selv har benyttet teknologien flittigt, har formentlig også dannet sig en opfattelse af dens styrker og begrænsninger. Det er imidlertid tankevækkende at erfaringerne på området er relativt divergerende.

I det følgende vil jeg fremlægge hvad forskellige forfattere har rapporteret om de enkelte elementer i en ordbogsartikel. Man skal være opmærksom på at de undersøgte aspekter varierer fra studie til studie, og det medfører at antallet af kilder bag hver vurdering også er forskelligt.

3.1 Styrker og potentialer

En interessant iagttagelse er at flere undersøgelser har evalueret kvaliteten af ChatGPT's definitioner, og der synes at være bred enighed om at de er af høj kvalitet. Ligeledes har tre studier beskæftiget sig med stilmarkører og vurderet dem som velfungerende (McKean & Fitzgerald 2023, Jakubíček & Rundell 2023, de Schryver 2023). Når det gælder eksempler og citater, er erfaringerne til gengæld mere blandede.

På baggrund af både egne og andres erfaringer vil jeg mene at der er generel enighed om at ChatGPT klarer sig godt på følgende områder:

1. *Hurtighed*: Dette dækker ikke blot responstiden, men også den tid brugerne behøver for at vænne sig til teknologien. Grænsefladen er intuitiv og kræver minimal tilvænning, hvilket gør det nemt at formulere effektive forespørgsler.
2. *Omkostningseffektivitet*: De fleste AI-tjenester tilbydes i gratisversioner der dækker behovene for de fleste almindelige brugere. I mine egne test brugte jeg gratisversionerne

af henholdsvis GPT-3.5 og senere GPT-4.0. Sidstnævnte giver mulighed for forespørgsler der kræver onlineadgang, modsat version 3.5, der kun har adgang til viden op til et bestemt tidspunkt.

3. *Velformulerede svar*: ChatGPT producerer generelt flydende, mundrette og venlige svar. Desuden har den en tendens til at undskylde for eventuelle fejl, hvilket bidrager til en positiv brugeroplevelse.
4. *Naturlig kommunikation*: Flere har peget på at interaktionen med en chatbot kan opleves som en mere "naturlig" måde at søge information på sammenlignet med traditionelle opslagsværker. Dette aspekt er vigtigt da det potentielt kan true ordbogen i sin nuværende form.
5. *Styrker inden for leksikografi*: Der synes at være enighed om at ChatGPT er særdeles kompetent til at formulere præcise definitioner, identificere stilmarkører samt foreslå synonymer og faste ordforbindelser (det sidste baseret på egne erfaringer).

Disse resultater peger på både muligheder og udfordringer for leksikografien i en tid hvor AI-teknologier spiller en stadig større rolle. ChatGPT forbedres løbende og i hastigt tempo. Meget har allerede ændret sig siden lanceringen i november 2022, og modellen lærer hele tiden af brugernes feedback. Dette afspejles i den eksplosive vækst i brugertrafikken:

november 2022:	1 million	brugere (efter fem dage)
januar 2023:	30 millioner	brugere
november 2023:	100 millioner	brugere
december 2024:	300 millioner	brugere
februar 2025:	400 millioner	brugere ¹

1 Der findes forskellige opgørelser over brugertal. Dem jeg præsenterer her, stammer fra <https://explodingtopics.com/blog/chatgpt-users> (marts 2025).

Disse tal dækker alene OpenAI's ChatGPT-plattform, mens konkurrerende modeller som Googles Bard, X's Grok, Microsofts Copilot, Apples Siri og Amazons Alexa naturligvis også bidrager til det samlede billede.

En af ChatGPT's styrker er at den nemt kan integreres i andre programmer. Eksempelvis kunne Joffe og de Schryver blot tre måneder efter lanceringen af ChatGPT præsentere en integreret løsning til ordbogsredigeringsystemet TLex (de Schryver & Joffe 2023). Tilsvarende eksperimenter finder sted på min egen arbejdsplads, Det Danske Sprog- og Litteraturselskab, og formentlig også ved andre ordbogsredaktioner i Norden.

3.2 Begrænsninger og udfordringer

ChatGPT og lignende chatbotter bygger på såkaldt store sprogmodeller (Large Language Models, LLM) der er trænet på enorme mængder tekstdata ved hjælp af maskinlæring. Disse modeller efterligner hjernens neurale netværk og anvender en arkitektur kaldet transformernetværk. Navnet GPT står for *Generative Pre-trained Transformer* som udtryk for at modellen genererer tekst baseret på forudgående træning.

En væsentlig egenskab ved ChatGPT er dens probabilistiske natur. I modsætning til deterministiske systemer er svarene ikke faste, men baserer sig på sandsynligheder beregnet af en algoritme. Det betyder at:

- modellen ikke besidder viden eller ræsonnerer deduktivt. Den genererer i stedet tekst ud fra beregnede sandsynligheder.
- den samme forespørgsel kan resultere i forskellige svar, afhængigt af de sandsynlighedsbetingede valg modellen foretager.

- senere versioner af ChatGPT (efter 3.5) er forsynet med en “prøv igen”-funktion der giver mulighed for at få alternative svar på samme spørgsmål.

ChatGPT kan tilpasse sig efter den feedback den får, og brugerne kan aktivt evaluere de modtagne svar ved at give en tommelfinger op eller ned. Denne mekanisme bidrager til den løbende udvikling af modellen.

Selvom det kan opleves positivt af brugeren, skaber det udfordringer i en forskningsmæssig kontekst. Når svarene varierer afhængigt af hvornår og hvordan et spørgsmål stilles, bliver det svært at sammenligne og vurdere dem konsistent. Teknologien udvikler sig desuden løbende, hvilket betyder at et svar givet i april 2024 kan afvige ganske meget fra et svar givet i februar 2025.

Det står i kontrast til et grundlæggende princip i videnskabelig forskning: Resultater skal være intersubjektivt overførbare, dvs. det skal være muligt at nå til samme konklusion hvis et eksperiment gentages af andre. Det er ikke tilfældet med store sprogmodeller, hvilket får nogle til helt at afvise dem som brugbare leksikografiske redskaber. Fx gav Michael Rundell udtryk for sådan en bekymring under paneldiskussionen ved eLex 2023: De algoritmer der driver chatbotten, er en lukket “sort boks”, hvilket betyder at man hverken kan teste hvordan svarene bliver til, eller systematisk sammenligne output.

Et velkendt problem er desuden de såkaldte “hallucinationer” hvor modellen leverer overbevisende, men faktisk forkerte svar. Det kan til tider være underholdende, men naturligvis også problematisk. Da jeg fx bad ChatGPT om en ordbogsartikel for ordet *spoon*, returnerede den et svar med en opdigtet betydning 2: “a term of endearment used towards a romantic partner”. Da jeg opfølgende spurgte til autentiske eksempler, leverede den flere falske citater (Trap-Jensen 2024:34-35). En anden udfordring er at ChatGPT kun er så god som de tekster den er trænet på.

Udsagn der forekommer hyppigt i træningsdata, får større vægt i modellens svar – uanset om de er korrekte eller passende. Det kan resultere i en forstærkning af stereotype forestillinger og spredning af misinformation. I tillæg til denne skævhed kommer en tilsvarende sproglig og kulturel skævhed. ChatGPT er udviklet i en amerikansk kontekst, og engelsk udgør typisk 90 % eller mere af træningsmaterialet. Dansk, norsk og svensk tilsammen udgør blot 2-4 %, hvilket forklarer dels hvorfor modellen oftere hallucinerer på disse sprog, dels at de modtagne svar er påvirket af engelsk. Fordi størstedelen af træningsmaterialet er på engelsk, betyder det at de mønstre i sprog, logik og viden modellen har lært, i høj grad er baseret på engelsksprogede kilder. Når chatbotten efterfølgende svarer på dansk, bruger den generelle sproglige “regler” og koncepter den har lært på engelsk, og tilpasser dem til dansk på grundlag af det danske materiale den *også* er blevet trænet på. Det giver den genererede danske tekst et særligt præg med fx indslag af anglicismer og afsmitning fra engelsk syntaks, og det giver teksten et glat, homogeniseret præg. Det er desuden blevet nævnt at nuancer går tabt fordi chatbotterne har særlig svært ved at forstå eksempelvis metaforiske virkemidler (jf. Pedersen et al. 2024:9), og problemerne vokser når det drejer sig om specialiserede domæner hvor der kun er begrænsede mængder materiale i det danske træningsmateriale.

Det påvirker også svarenes kulturelle præcision – for eksempel kunne ChatGPT (især i den tidlige version) have svært ved at gengive fx fødselsdagssange eller årstidens traditioner korrekt for en dansk eller nordisk kontekst i mine undersøgelser (Trap-Jensen 2023:35-36). Det harmonerer godt med de fejloverbørsler der rapporteres for eksempelvis *flødebolle* og *æbleskive* i Pedersen et al. (2024). Se også Zhang et al. (2023) for andre sprog end dansk.

Disse betænkeligheder bør også indgå når man bedømmer de erfaringer der beskrives, fx de ovenfor nævnte fra eLex 2023. En positiv vurdering af ChatGPT's præstation på engelsk kan ikke

bare overføres til en tilsvarende vurdering når det gælder andre sprog.

Endelig bør de etiske og politiske betænkeligheder ved brugen af chatbotter også nævnes. Et væsentligt kritikpunkt er det betydelige energiforbrug der er forbundet med både driften og træningen af de store sprogmodeller. En enkelt forespørgsel i ChatGPT estimeres at bruge langt mere energi end fx søgninger i Google (ca. 0,19 kWh per prompt mod 0,0003 kWh for en enkelt Google-søgning²). Når dette skales op til de omkring 400 millioner brugere (per februar 2025) der samlet foretager omtrent en milliard søgninger dagligt, bliver den samlede miljøpåvirkning betydelig. Hertil kommer at selve træningen af modellerne er ekstremt energikrævende. Disse forhold udgør en væsentlig udfordring i en tid præget af klimaforandringer – især hvis løsningen på energibehovet udelukkende baseres på øget fossil udvinding.

Derudover rejser udviklingen og udbredelsen af chatbotter en række politiske spørgsmål. Store amerikanske teknologivirksomheder dominerer markedet med en efterfølgende koncentration af økonomisk magt hos aktører som Elon Musk, Mark Zuckerberg og Jeff Bezos – for ikke at nævne mere kontroversielle platforme som Donald Trumps Truth Social. Den potentielle anvendelse af AI-baserede systemer til at understøtte specifikke ideologiske dagsordener, sådan som det allerede ses på sociale medier, giver anledning til bekymring. Spørgsmålet er derfor om eller i hvilket omfang det er etisk forsvarligt at bruge og dermed indirekte støtte disse teknologier.

² Ifølge Ingeniørforeningen IDA på <https://www.dr.dk/nyheder/viden/teknologi/chatbots-sviner-klimaet-kaempe-datacentre-kan-om-faa-aar-sluge-hele-japans>.

3.3 Vurdering af ChatGPT i leksikografi

Sammenfattende tegner der sig et komplekst billede af ChatGPT's anvendelse i leksikografien. Udviklingen sker i et meget højt tempo, og det indebærer at konklusioner truffet i dag ikke længere er gyldige om blot få måneder. Der er en række åbenlyst problematiske aspekter som man må forholde sig til, fx at de svar der genereres, ikke er verificerbare eller konsekvent reproducerbare fordi de underliggende algoritmer fungerer som en lukket, uigennemsigtig "sort boks".

Samtidig er min egen erfaring at ChatGPT kan være en værdifuld sparringspartner for den menneskelige leksikograf. Den kan være et nyttigt værktøj, særligt når den bruges bevidst og kritisk. Det er derfor afgørende at udvikle en metodisk tilgang til brugen af AI i leksikografisk arbejde – herunder at formulere præcise og velovervejede spørgsmål samt at være opmærksom på den engelsksprogede bias der præger mange af modellens svar. Men vi er endnu langt fra at kunne bruge ChatGPT automatisk, uden menneskelig redaktionel mellemkomst, med den kvalitet AI har i øjeblikket.

Udover de områder jeg har berørt her, og som vedrører direkte anvendelse til ordbogsredigering, tegner der sig også muligheder i beslægtede områder. Fx kan jeg sagtens forestille mig at AI vil kunne spille en rolle i leksikografisk præprocessering, eksempelvis til udvælgelse af gode citater i et korpus eller til fremfindning af kandidater til flerordsudtryk.

Det er altså et uklart billede der tegner sig, med både faglige fordele og teknisk-metodiske betænkeligheder. Hertil kommer de etiske og politiske udfordringer, herunder risikoen for øget økonomisk ulighed og fremvæksten af et overvågningssamfund. Det er overvejelser vi hver især må forholde os til.

Litteratur

Ordbøger

DDNO = *Den Danske Netordbog*. Tilgængelig på <<https://ordbogen.com>> (april 2025).

DDO = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. <<https://ordnet.dk/ddo>> (april 2025).

ODE = *Oxford Dictionary of English*, 3rd ed. Angus Stevenson (ed.). <<https://www.oxfordreference.com/display/10.1093/acref/9780199571123.001.0001/acref-9780199571123>> (april 2025).

Anden litteratur

Atkins, B. T. Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.

de Schryver, Gilles-Maurice (2023): Generative AI and Lexicography: The Current State of the Art Using ChatGPT. I: *International Journal of Lexicography* 36(4), 355-387. <<https://doi.org/10.1093/ijl/ecado21>> (april 2025).

de Schryver, Gilles-Maurice & David Joffe (2023): The End of Lexicography, Welcome to the Machine: On How ChatGPT Can Already Take over All of the Dictionary Maker's Tasks. Paper presented at the 20th CODH Seminar. Center for Open Data in the Humanities, Research Organization of Information and Systems, National Institute of Informatics, Tokyo, Japan, 27 February 2023. <<https://www.youtube.com/watch?v=mEorwoyefAs>> (april 2025).

de Schryver, Gilles-Maurice, Michael Rundell, Arvi Tavast, Pavel Rychlý, Marko Kokol (panellists) & Simon Krek (moderator) (2023): Round Table on "Large Language Models and AI in Lexicography". Panel at the 8th Electronic Lexicography in the

- 21st Century Conference, Brno, Czech Republic, 27-29 June 2023.
- Jakubíček, Miloš & Michael Rundell (2023): The end of lexicography? Can ChatGPT outperform current tools for post-editing lexicography? I: Marek Medveď, Michal Měchura, Carole Tiberius, Iztok Kosem, Jelena Kallas, Miloš Jakubíček & Simon Krek (eds.): *Electronic lexicography in the 21st century (eLex 2023): Invisible Lexicography. Proceedings of the eLex 2023 conference*, Brno, 27-29 June 2023. Brno: Lexical Computing CZ s.r.o. 508-523.
- McKean, Erin & Will Fitzgerald (2023). The ROI of AI in Lexicography. I: *Proceedings of the 16th International Conference of the Asian Association for Lexicography: "Lexicography, Artificial Intelligence, and Dictionary Users"*. Seoul: Yonsei University. 10-20.
- Pedersen, Bolette Sandford, Nathalie Carmen Hau Sørensen, Sussi Olsen & Sanni Nimb (2024): Evaluering af sprogforståelsen i danske sprogmodeller – med udgangspunkt i semantiske ord-bøger. I: *NyS – Nydanske Sprogstudier* 65, 8-40. <<https://doi.org/10.7146/nys.v1i65.143072>> (august 2025).
- Trap-Jensen, Lars (2024): The Best of Two Worlds: Exploring the Synergy between Human Expertise and AI in Lexicography. I: Tinatin Margalitadze (ed.): *Proceedings of the I International Conference. Lexicography in the XXI Century*. Centre for Lexicography and Language Technologies, Ilia State University. 26-39. <https://lexicography21.iliauni.edu.ge/wp-content/uploads/2024/10/03_Lars-Trap-Jensen.pdf> (april 2025).
- Trap-Jensen, Lars & Henrik Lorentzen (2024): Back to Basics: Meaning Description for Human Users and for Computers. I: K. Š. Despot, A. Ostroški Anić & I. Brač (eds.): *Lexicography and Semantics. Proceedings of the 21st EURALEX International Congress*, 8-12 October 2024, Cavtat/Dubrovnik, Croatia. 191-209. <<https://euralex.org/publications/back-to-basics-mean>

ing-description-for-human-users-and-for-computers> (april 2025).

Zhang, Xiang, Senyu Li, Bradley Hauer, Ning Shi & Grzegorz Kondrak (2023): Don't trust ChatGPT when your question is not in English: A study of multilingual abilities and types of LLMs. I: *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. Singapore: Association for Computational Linguistics. 7915-7927. <<https://aclanthology.org/2023.emnlp-main.491>> (august 2025).

Lars Trap-Jensen
seniorredaktør, cand.mag. & MPhil
Det Danske Sprog- og Litteraturselskab
Christians Brygge 1
DK-1219 København K
ltj@dsl.dk

Kjønnsrepresentasjon og -fremstilling i *Bokmålsordboka*s brukseksempel

Emma Josefin Ölander Aadland

The article presents a study investigating how gender is represented and portrayed in the Norwegian dictionary *Bokmålsordboka* before and after revision. Focusing on usage examples, the analysis examines typical gender role patterns, descriptions of appearance and character traits, and the allocation of agent versus patient roles. Results indicate a trend toward less stereotypical portrayals, although traces of outdated norms persist. The study underscores the challenge of balancing accurate language reflection with the avoidance of reinforcing gender stereotypes in dictionaries.

1. Innledning

I årene 2018–2024 gjennomgikk de to standardordbøkene *Bokmålsordboka* (BOB) og *Nynorskordboka* (NOB) den første fullstendige revisjonen siden de ble trykt i 1986 (*Revisjonsprosjektet*, u.å.). Revideringsarbeidet omfattet dels inkludering av nye lemmar og dels å gjøre både lemmautvalg og struktur i begge ordbøkene bedre samkjørt. Videre ble definisjoner, brukseksempel og faste uttrykk gjennomgått med formål å sikre at disse er i tråd med dagens språkbruk (jf. Selback et al. 2025).

BOB og NOB er deskriptive samtidsordbøker med et uttalt formål om å beskrive og dokumentere språket i bruk i samtiden, og de har en sammensatt brukergruppe som omfatter alt fra oftestlig forvaltning til skoleelever og andrespråksinnlærere (Rauset 2022:98). Utover å være deskriptive kan ordbøkene også sies å ha en normerende funksjon (Lyse & Svoldal 2020:80–81), dels ved at de fungerer veiledende når det gjelder ords betydning og bruk, og dels ved at språkbrukerne tillegger dem en viss normativ autoritet.

I lys av den posisjon ordbøkene har, er det interessant å undersøke i hvilken grad tradisjonelle kjønnsroller og kjønnsstereotypiske fremstillinger av kjønnene forekom i ordbøkene før det omfattende revideringsarbeidet, samt i hvilken grad det fremdeles eksisterer spor av slikt i de reviderte utgavene. I studien som presenteres i denne artikkelen, undersøker jeg dette ved å sammenligne brukseksempel i BOB før og etter revisjonen. Konkret undersøker jeg forekomsten av typiske kjønnsrollemønstre, hvordan kjønnene beskrives utseendemessig og når det gjelder karakterdrag, samt i hvilken grad respektive kjønn innehar en agens- eller patiensrolle. Studien kan bidra til å gi ny kunnskap om språkbruken tilknyttet kjønn i ordbøker, som kan være verdifull og bevisstgjørende for det videre leksikografiske arbeidet.

Artikkelen har følgende struktur: Innledningsvis (avsnitt 2) gis et teoretisk rammeverk for studien, og videre i avsnitt 3 presenteres datamaterialet og den metodiske fremgangsmåten. Avsnitt 4 inneholder en redegjørelse for resultatene og en komparativ analyse av funn i de to utgavene av ordboken. Artikkelen oppsummeres i den avsluttende del 5.

2. Teori og tidligere forskning

2.1. Et normkritisk perspektiv på ordbøker

Ordbøker bør etterstrebe å gi en objektiv beskrivelse og dokumentasjon av ordforrådet i et språksamfunn, ettersom de gjerne betraktes som troverdige kilder til kunnskap om hva ord betyr og hvordan de skal brukes (Milles 2009:125). Ordbøkers beskrivelse av språket kan dertil også ha en normgivende rolle (se f.eks. Moon 2014). At ordbøker kan inneha både en deskriptiv og normativ funksjon, har i senere år medført en stadig større forventning om at de oppfyller ideelle forestillinger om eksempelvis likestilling og

språklig inkludering (Trap-Jensen 2020a:33). Fra et normkritisk perspektiv handler dette om å utjevne eventuelle uønskede skeivheter, som innbefatter ordbøkernes fremstilling og representasjon av eksempelvis kjønn, kjønnsoverskridende identitet eller uttrykk, etnisk bakgrunn, religion, funksjonsnedsetting, seksuell legning og alder. En ordbok kan likevel av natur ikke være upolitisk, ubunden eller usosial (Chen 2019:362). På den ene siden kan ordbøker altså påvirke både språkoppfatninger og generelle språkholdninger i samfunnet, og på den andre siden er de i stor grad preget av samtiden og kulturen de er bundne til.

At en ordbok kan betraktes som «et vindu til den kulturen den er laget i» (Fjeld 2015:40), vil være reflektert så vel i ordbokens lemmaseleksjon som i utformingen av definisjoner samt valg av bruksmarkører og brukseksempel i ordbokartiklene (jf. H. Nikula 1995; Grønvik 2002). Når det gjelder brukseksempel, har disse gått fra i hovedsak å være formulert av ordbokredaktørene i form av såkalte introspektive data, altså basert på egen språklig intuisjon, fantasi og erfaring, til i større grad å baseres på eksempel hentet fra korpustekster (Fjeld 2015:40). Brukseksempelene kan sies å ha en sammensatt funksjon ved å skulle gjenspeile både sosiale, historiske og kulturelle aspekt ved språksamfunnet (H. Nikula 1995:312). Det forekommer dermed noen åpenbare utfordringer ved bruk av korpusbelegg; disse kan være skeivt sammensatt for eksempel når det gjelder sjanger, tidsspenn eller representasjon av kjønn eller ulike minoritetsgrupper (Nimb 2024). Bruk av korpuslingvistiske metoder i ordbokarbeid krever dermed en bevissthet rundt dette, eksempelvis for å unngå at kjønnsstereotyper overføres fra korpus til ordbok.

At ordbøker viderefører kjønnsstereotyper, forekommer blant annet ved at kvinner omtales nedlatende, i kjønnsstereotypiske roller og som avhengige av menn (Fjeld 2001:13). Et typisk normativt kjønnsrollemønster fremtrer dessuten i fordelingen mellom kvinner og menn i sosiale roller, yrkes- eller samfunnsmessig.

Videre kan det anses som kjønnsstereotypisk å fremstille kvinner som objekter med referanse til deres utseende samt i foreldede roller som eksempelvis er tilknyttet hjemmet. En kjønnsstereotypisk fremstilling av menn vil være som aktive og agerende subjekt, gjerne tilknyttet viktige yrkes- eller samfunnsposisjoner og med stor aksjonsradius (Fjeld 2015:38). I senere år har den vitenskapelige oppmerksomheten rundt kjønnsstereotyper i ordbøker vokst (Vacalopoulou 2022:168), men blant andre Sköldberg (2020:21) peker på behovet for flere normkritiske studier av ordbøker.

2.2. Tidligere forskning

Den leksikografiske ordbokforskningen med et kjønnskritisk perspektiv fikk et tydelig oppsving i Norden på 1990-tallet. Hageberg (1990) var tidlig ute med å undersøke hvordan kvinner og menn blir representerte i norske ordbøker, og hennes funn var kortfattet at kvinner er underrepresenterte i forhold til menn, samt at kvinnene ofte blir fremstilt unyansert. Fjeld (2001) viser til en leksikalsk subtil kjønnsdiskriminering som gjenspeiles i at ordbokartikler for mer tradisjonelle kvinneedominerte interesser er mangelfulle i flere norske ordbøker, og videre at eksempel ofte gir en stereotypisk fremstilling av kvinner med tanke på yrkesvalg og avhengighet av menn (2001:11–13). I forbindelse med revideringsarbeid i senere år har lemmalakuner blitt færre, men Fjelds komparative studie av *Bokmålsordboka* (BOB), *Svensk Ordbok* (SO) og *Den Danske Ordbog* (DDO) viser at både kjønnsскеivheter og stereotyper fremdeles lever i beste velgående (2015:35–36). Blant annet gir nevnte ordbøker det mannlige kjønn større plass.

Flere andre svenske ordbøker har også vært under lupen de siste tiårene. En studie av svenske SOB (1986) fra 1990-tallet viser at diskrimineringen av kvinner som forekommer i ordboken, til en viss grad er skjult, objektiviserer og passiviserer kvinnen samt underbygger foreldede kjønnsrollemønstre som til en viss grad

kan sies å nedverdige kvinnen (K. Nikula 1997:196). Milles (2009) analyserer ordbokartikler i ni ulike svenske ordbøker utgitt mellom 1937 og 2006. Hun finner at ordbøkene delvis er normbrytende i den forstand at flere kan sies blant annet å vise til et åpnere syn på seksualitet og på kvinnelig seksualitet spesifikt (2009:134). Derimot viser en undersøkelse av den tospråklige svensk-finske ordboken *Stora finsk-svenska cd-romordboken* fra 1998 en tydelig skeivhet både i representasjon og fremstilling av kjønnene (Pilke 2009). Undersøkelsen tar for seg hvordan pronomenene *han* og *hun* fremstilles yrkesmessig i brukseksempel i ordboken. Sentrale funn er at *han* brukes mer frekvent, og at ordboken gir en nokså stereotypisk fremstilling av begge kjønn når det gjelder yrkesmessig tilhørighet.

Videre har Larsson (2017) i en studie av den syvende utgaven av den svenske ordlisten SAOL, som utkom i 1900, undersøkt hvordan innføringen av både grammatisk og seksuelt kjønn medførte en nokså snever fremstilling af og plass tilegnet kvinner i ordlisten. Sköldbbergs studie av *Svensk ordbok* (SO 2009) fra 2020 viser at jenter i større grad portretteres fra et mannlig perspektiv, og at jenter oftere er fremstilt objektiverende og med beskrivelser som berører utseende i eksempel, mens menn i større grad kobles til karakterdrag (Sköldbberg 2020:14–15). I begynnelsen av 2000-tallet gjennomgikk SO en større revisjon, der et overordnet mål var å gjøre ordboken mer inkluderende og likestilt samt å unngå stereotypiske eller foreldede fremstillinger av kjønnsroller (Hult 2010:216).

Det finnes også nordiske studier som undersøker sammenheng mellom kjønn og språk i korpustekster (se Mattisson 2006; Fjeld & Knudsen 2013; Tonne, Uri & Johnsen 2022). Et eksempel er den komparative studien Trap-Jensen (2020b) har gjennomført, som undersøker forholdet mellom korpus og ordbøker i danske og norske ordbøker. Han påpeker at det må skilles mellom skeivheter i språket, skeivheter i samfunnet, skeivheter i korpus og skeivheter

i ordbøkene (2020b:33), samtidig som det sannsynligvis forekommer sammenhenger mellom disse fire som kan være relevante å undersøke nærmere.

For å sammenfatte viser den nordiske forskningen til en historisk skeivhet i kjønnes representasjon og en stadfesting av stereotypiske gestalter i flere nordiske ordbøker. Den nordiske forskningen kan også sies å stå i samsvar med internasjonal leksikografisk forskning (se f.eks. Sköldberg 2020:8). Studien som presenteres i denne artikkelen, bidrar med ny kunnskap om hvordan kjønn fremstilles og er representert i brukseksempel i BOB, samt hvordan den gjennomgående revisjonen har påvirket språkbruken om kjønn.

3. Datamateriale og metode

3.1. Datamateriale og avgrensninger

Datamaterialet for studien er hentet ut via språkdatainfrastrukturen INESS (Rosén et al. 2012), som ligger under CLARINO Bergen Centre.¹ INESS inneholder en omfattende samling trebanker, som alle er laget av ett eller flere korpus. Herunder finnes trebanker for begge standardordbøkene både før Revisjonsprosjektet startet i 2018 og med status i desember 2024.² Trebankene for BOB er brukt for å hente ut relevant datamateriale for denne studien.

Studien er avgrenset til å inkludere substantivene *kvinne* og *jente* (heretter FEM) samt *mann* og *gutt* (heretter MASK) og undersøker alle brukseksempel som inneholder disse både før

¹ <https://clarino.uib.no>.

² Trebankene ble laget 9. desember 2024 basert på alle artikler som ikke var markert som utgått eller slettet. Ettersom revideringsarbeidet fremdeles pågikk, og enkelte artikler fremdeles var under redigering, vil det dermed kunne være enkelte avvik fra den tilgjengelige ordboken på nett.

og etter revisjonen. Avgrensningen medfører at studien ikke kan påstå å gi et fullstendig bilde av kjønnsfremstillingen i ordboken, men den kan peke på noen relevante tendenser i de to utgavene både når det gjelder representasjon og fremstilling.

Datasettet er hentet ut fra INESS med et søkeuttrykk som gir både ubøyde og bøyde former av ordene: /kvinne|jente/ og /mann|gutt/. Dataene, som vist med frekvensoversikt i tabell 1, er eksportert til Excel med tilhørende metadata (artikkel-id, lemma, brukseksempel) for videre klassifisering og analyse.

Ordbokutgave	kvinne jente	mann gutt
BOB2018	143	214
BOB2024	260	369

Tabell 1: Antall brukseksempel som inneholder *kvinne* eller *jente* og *mann* eller *gutt* i 2018- og 2024-utgaven av BOB.

3.2. Metode

Studiens metodiske fremgangsmåte kan betegnes som en kvalitativ konkordansanalyse (Stefanowitsch 2020:54), med innslag av kvantitative aspekt. Den kvalitative analysen bygger på kategorier som Fjeld benytter i en studie av brukseksempel som inneholder *han/ham/hans* og *hun/henne/hennes* med avgrensning til lemmautvalget for bokstaven *a* og *f* (Fjeld 2015:45). Noen tilpasninger er likevel gjort for min studie, og analysen tar i hovedsak for seg å beskrive forskjeller i ren frekvens for de undersøkte kategoriene. En fullstendig oversikt over kategoriene med tilhørende eksempel presenteres i tabell 2.

Kategori	Eksempel
Kjønn nødvendig	
Ja	en <u>svanger</u> kvinne
Nei	mennene i slekta var kjent for sin <u>hardhet</u>

Kategori	Eksempel
Utseende fysisk/attraktiv	
Fysisk	<i>en <u>enarmet</u> mann</i>
Attraktiv	<i>en <u>sexy</u> kvinne</i>
Fysisk + negativ	<i>gutten var en blek liten <u>krok</u></i>
Attraktiv + negativ	<i>jenta var <u>småpen</u>³</i>
Karakterdrag	
Positiv	<i>en <u>modig</u> jente</i>
Negativ	<i>det er ingen <u>framferd</u> i den gutten</i>
Kjønnsrollemarkert tema	<i>tradisjonelt har bilen vært mannens <u>domene</u></i>
Agens/patiens	
Agens	<i>jenta klatrer opp i et <u>lekestativ</u></i>
Patiens	<i>jeg møtte en <u>deilig</u> gutt på byen i går</i>
Nøytral	<i>både kvinner og menn er <u>representert</u> i styret</i>

Tabell 2: Oversikt over kategorier med tilhørende eksempel. Understreket ord viser til lemmaet eksempelet tilhører.

For kategorien «Kjønn nødvendig» har jeg gjennomgått alle brukseksempel for å avklare om bruken av MASK- eller FEM-ordene kan anses nødvendig. Eksempelvis vurderer jeg det som nødvendig å innlemme *kvinne* i brukseksempelet når lemmaet tilsier at det er kjønnsbetinget, som i tilfellet *svanger*. Disse har blitt ekskludert fra videre klassifisering. For kategorien «Utseende» er alle brukseksempel som beskriver utseende, klassifisert som enten «Fysisk» eller «Attraktivitet». Inndelingen kan avdekke eventuelle kjønnsforskjeller tilknyttet ulike typer utseendebeskrivelser. Disse har dertil blitt gjennomgått med hensikt å identifisere de som kan tolkes som negative utseendebeskrivelser. En lignende kategorisering er gjort for «Karakterdrag», med en binær inndeling mellom negative og positive beskrivelser av karakterdrag. For kategorien «Kjønnsrollemarkert tema» markeres de brukseksempel som kan

3 Dette eksempelet er ikke entydig negativt, og jeg har derfor markert det som semantisk ambivalent i analysen.

tolkes å fremstille respektive kjønn i stereotypiske kjønnsroller (se avsnitt 2.1). Den siste kategorien «Agens/patiens» skiller mellom den som utfører handlingen i setningen (agens), og den som handlingen påvirker (patiens). I datamaterialet forekommer flere brukseksempel som mangler en konkret handling, og bare de eksempler som har en verbhandling, er inkludert i denne klassifiseringen. Videre gis tilfeller der begge kjønn har samme rolle i et brukseksempel, markeringen «Nøytral».

4. Resultat og funn

I denne delen tar jeg innledningsvis for meg resultatene fra analysen av BOB2018-utgaven, for deretter å gå videre til BOB2024-utgaven. Avslutningsvis sammenligner jeg de to utgavene av ordboken og diskuterer studiens funn.

4.1. 2018-utgaven av *Bokmålsordboka*

I BOB2018 er det 2 tilfeller blant de 214 brukseksempelene der MASK vurderes som nødvendig.⁴ Blant de 143 brukseksempel som inneholder FEM, er det 4 der bruken av enten *kvinne* eller *jente* vurderes som nødvendig.⁵ Disse ble ekskludert for begge kjønnskategoriene. Videre forekommer det totalt 18 brukseksempel der begge kjønn er inkludert i det samme eksempelet. Jeg har vurdert det som relevant å ha med disse. Dermed er det 212 brukseksempel for MASK og 139 brukseksempel for FEM som er inkludert i den videre klassifiseringen.

4 Brukseksempelene er for lemmaene *X-kromosom* og *bakkenbart* ('kinnskjegg').

5 Brukseksempelene er for lemmaene *X-kromosom*, *nedkomme*, *svanger* og *syklus*.

4.1.1. Utseende

For MASK forekommer det 24 brukseksempler (11 % av det totale antallet MASK-eksempler) som innehar utseendebeskrivelser i BOB2018. 8 av disse er semantisk ambivalente tilfeller, med bakgrunn i at de kan beskrive enten fysikk, karakterdrag eller alder.⁶ Totalt 35 brukseksempler (25 % av det totale antallet FEM-eksempler) beskriver utseende for FEM, hvorav 14 er kategorisert som semantisk ambivalente.

I 2018-utgaven er det mer vanlig at FEM beskrives som attraktive enn MASK (12 tilfeller mot 2). Det er derimot en nokså jevn forekomst av beskrivelser av fysisk utseende for begge kjønn (23 tilfeller for FEM mot 22 for MASK), som vist i tabell 3.

Kjønn	Fysisk	Attraktivitet	Negativ
MASK (n = 24)	14 + 8	2 + 0	9
FEM (n = 35)	13 + 10	8 + 4	1

Tabell 3: De ulike underkategoriene for utseende fordelt mellom kjønnene i BOB2018. Tall etter plusstegn viser de semantisk ambivalente tilfellene.

En nærmere undersøkelse av brukseksempelene som beskriver attraktivitet, viser at enkelte i tillegg er seksualiserende. Om FEM finner vi eksempelvis «en het kvinne» og «en lekker jente». Et slikt bilde tegnes derimot ikke av MASK, som under kategorien «Attraktivitet» kun har de to eksemplene «være pen gutt» og «en attraktiv mann».

Det forekommer derimot flere negative beskrivelser av utseende i kombinasjon med MASK (n = 9) enn FEM (n = 1). Flere av de negative utseendebeskrivelsene om MASK innehar adjektiv som beskriver kroppslig størrelse eller et eldre utseende. Det er kun ett eksempel på en mulig negativ utseendebeskrivelse om jenter, nem-

⁶ Eksempel på en semantisk ambivalent utseendebeskrivelse er «han er blitt en stor mann».

lig «jenta var småpen». Dette eksempelet er ikke semantisk entydig negativt, men kan tolkes i retning å være mer negativt enn positivt. Om kvinner forekommer det ingen eksempel som tolkes negativt.

4.1.2. Karakterdrag

Totalt 66 brukseksempel for MASK innehar en beskrivelse av karakterdrag, og av disse er 40 klassifisert som positive mot 26 negative. Blant brukseksempelene for FEM er det totalt 38 brukseksempel som beskriver karakterdrag, og her er fordelingen 31 positive mot 7 negative.

MASK kommer dermed dårligere ut enn FEM når det gjelder hvor ofte kjønnet fremstilles med negative karakterdrag i brukseksempel; MASK har 39 % av det totale antallet beskrivelser av karakterdrag som er klassifisert som negative, mot 18 % for FEM. Eksempelvis beskrives jenter og kvinners karakterdrag negativt med adjektiv som *fjasete*, *forstyrret* og *lidende*. Gutter og menns karakterdrag beskrives negativt med adjektiv som *besatt*, *lidderlig* og *uvillig*.

4.1.3. Kjønnnsrollemarkert tema

Det er et høyere antall brukseksempel som innehar kjønnnsrollemarkert tema for MASK (n = 42) enn FEM (n = 35). Brukseksempelene som omhandler kvinner, viser til deres stilling i samfunnet både yrkesmessig og kulturelt, som eksempelvis i «kvinnene er underrepresentert i offentlig styre og stell», «det er overtall av kvinner i omsorgsykker», «dagens kvinne vil ikke bare bli betraktet som kjønnnsobjekt» og «behandle kvinner som leketøy». Kjønnnsrollemerkerte brukseksempel om jenter er blant andre «kjøpe li-kedanne kjoler til jentene». Om menn forekommer eksempelvis «mennene i slekten var kjent for sin hardhet», «styret består av bare menn» og «vis deg nå som en mann!». Om gutter finner vi «guttene smugleste pornoblader» og «det er bare sport som står i hodet på gutten».

4.1.4. Agens- og patiensrolle

I 2018-utgaven forekommer MASK som agens eller patiens i totalt 116 brukseksempel (55 % av det totale antallet MASK-eksempel). Av disse er 77 i agensrolle (hvorav 5 tilfeller har tilleggsmarkering nøytral), mens 39 brukseksempel gir MASK en patiensrolle (hvorav 5 tilfeller har tilleggsmarkering nøytral). Fordelingen mellom agens- og patiensrolle er 66 % mot 34 %. Menn forekommer som agens i eksempel som «mennene i slekten var kjent for sin hardhet» og «alle mann var pirum» ('beruset'), mens guttene figurerer i agensrolle blant annet i «guttene smugleste pornoblader». I patiensrolle finner vi MASK eksempelvis i «hun har dressert mannen sin godt», «bølgene skylte en mann over bord», «jeg leter etter en liten gutt i blå skjorte» og «jenta vraket gutten».

For FEM er det 69 brukseksempel (50 % av det totale antallet FEM-eksempel) som inngår i denne klassifiseringen, og av disse er fordelingen 43 i agensrolle (hvorav 4 tilfeller har tilleggsmarkering nøytral) mot 26 i patiensrolle (hvorav 4 tilfeller har tilleggsmarkering nøytral). FEM er dermed plassert i en agensrolle i 62 % av tilfellene og figurerer i en patiensrolle i de resterende 38 %. Eksempel på jenter og kvinner i agensrolle er «kvinnene krever full jamstilling» og «det er tæl i den jenta». FEM har en patiensrolle blant annet i «partiene bør nominere flere kvinner», «voldta en kvinne» og «han har forgått seg mot en liten jente».

4.2. 2024-utgaven av *Bokmålsordboka*

I datamaterialet som er hentet ut fra den reviderte utgaven av BOB, har jeg for kategorien «Kjønn nødvendig» vurdert det som nødvendig å ha med MASK i 2 av totalt 369 brukseksempel.⁷ Totalt er det 18 dubletter eller triplerter av samme brukseksempel i datamaterialet som ekskluderes, og uten disse gjenstår 349 bruk-

7 Brukseksempelene er for lemmaene *ejakulasjon* og *X-kromosom*.

seksempel for videre klassifisering.⁸ For FEM er det 5 av totalt 260 brukseksempel der kjønn vurderes som nødvendig.⁹ Når disse samt 8 tilfeller av dubletter og tripler ekskluderes, gjenstår 247 brukseksempel. Det forekommer totalt 22 brukseksempel som inneholder begge kjønn i 2024-utgaven, og disse er inkludert for videre klassifisering.

4.2.1. Utseende

I BOB2024 forekommer det 51 utseendebeskrivelser for MASK (15 % av det totale antallet MASK-eksempel). Av disse er 13 klassifisert som semantisk ambivalente. For FEM forekommer det 57 utseendebeskrivelser (23 % av det totale antallet FEM-eksempel), og av disse er 17 semantisk ambivalente, som vist i tabell 4.

Kjønn	Fysisk	Attraktivitet	Negativ
MASK (n = 51)	31 + 10	7 + 3	12 + 1
FEM (n = 57)	29 + 14	11 + 3	2 + 2

Tabell 4: De ulike underkategoriene for utseende fordelt mellom kjønn i BOB2024. Tall etter plusstegn viser de semantisk ambivalente tilfellene.

En nærlesing av brukseksempelene viser noe differanse mellom kjønnene. Eksempelvis er det flere tilfeller med utseendebeskrivelser som kan tolkes som negative for MASK (n = 13) enn FEM (n = 4). Derimot beskrives FEM (n = 14) mer frekvent som attraktive enn MASK (n = 10). Blant brukseksempelene forekommer det om begge kjønn enkelte som kan tolkes seksualiserende, blant andre «jeg møtte en deilig gutt på byen i går» og «en kvinne med yppige former».

8 Dubletter og tripler av samme brukseksempel i datamaterialet opptrer når et oppslagsord har flere rettskrivningsvarianter (eksempel: *fjaset* og *fjasete*).

9 Brukseksempelene er for lemmaene *svanger*, *syklus*, *nedkomme*, *utflod* og *X-kromosom*.

Videre er det omtrent like mange fysiske utseendebeskrivelser for begge kjønn, med 41 brukseksempel for MASK og 43 for FEM. Innholdsmessig tegner det seg derimot et noe ulikt bilde av kjønnene. Jenter er oftere beskrevet som fysisk små ($n = 9$) enn gutter ($n = 4$). Videre er jenter kledd i klesplagg som «mini», «bunad» eller bare «lettkledd», mens gutter ikke kles i noen plagg i hele tatt. Om kvinner finner vi fysiske utseendebeskrivelser med adjektiv som *høyreist*, *lyseblond*, *rund* og *ferm*, mens menn fysisk er *sterke*, *blekfete*, *krumryggede* og *dresskledd*.

4.2.2. Karakterdrag

Totalt 109 av 349 brukseksempel (31 %) med MASK inneholder en beskrivelse av karakterdrag, og av disse er fordelingen 51 % positive mot 49 % negative. For FEM er det totalt 68 av 247 brukseksempel (28 %) som beskriver karakterdrag, med fordelingen 72 % positive mot 28 % negative. For begge kjønn er det flere positive enn negative beskrivelser av karakterdrag. Derimot har MASK en jevnere fordeling mellom positive og negative karakterdrag sammenlignet med FEM, som har en markant overvekt av positive eksempler. De positivt klassifiserte brukseksempelene om jenter og kvinner har adjektiv som *from*, *monden*, *ærbar*, *flink*, *arbeidsom* og *drivende*. Om gutter og menn brukes adjektiv som *varmhjertet*, *verdensklok*, *holden*, *rettskaffen*, *sjenerøs* og *sjarmerende*.

MASK har dessuten flere negative beskrivelser av karakterdrag enn FEM, med 49 % mot 28 %. Gutter og menn beskrives blant annet som *undersetsig*, *besatt*, *glemsom*, *liddelig*, *rasende* og *nedbrutt*. De negative beskrivelsene av jenter og kvinners karakterdrag inneholder på sin side adjektiv som *naiv*, *usikker*, *nervøs*, *ømfintlig*, *farlig* og *redd*.

4.2.3. Kjønnssrollemarkert tema

For begge kjønn har det tilkommet 22 brukseksempel med kjønnssrollemarkert tema etter revisjonen, og totalt har MASK 64 slike

eksempel og FEM 57 eksempel i BOB2024. Av disse er 2 for MASK og 3 for FEM omformulerte eksempel. Blant de nye brukseksempelene med kjønnsrollemarkert tema finnes om jenter og kvinner «hestegale jenter» og «kvinnen holdt mannen helt utenfor planleggingen», mens vi om gutter og menn finner «en nerdete gutt med stor interesse for data» og «menn i førtiårskrisen som kjøper seg motorsykkel».

4.2.4. Agens- og patiensrolle

MASK innehar en agens- eller patiensrolle i 190 brukseksempel (51 % av det totale antallet MASK-eksempel). Av disse er 144 (76 %) klassifisert som tilfeller med agens (hvorav 9 er tilleggsmarkerte som nøytrale) og 46 (24 %) som tilfeller med patiens (hvorav 5 har fått tilleggsmarkering som nøytrale). FEM har en agens- eller patiensrolle i 125 brukseksempel (51 % av det totale antallet FEM-eksempel). Her er 89 (71 %) klassifisert som tilfeller med agens (hvorav 8 er tilleggsmarkerte som nøytrale) og 36 (29 %) som tilfeller med patiens (hvorav 4 har fått tilleggsmarkering som nøytrale). Begge kjønn har altså en nokså lik fordeling mellom agens- og patiensroller i forhold til det totale antallet brukseksempel for respektive kjønnskategori.

Blant de nye brukseksempelene i BOB2024 forekommer kvinner i en patiensrolle i eksemplene «porno er preget av objektivisering av kvinnen» og «han voldtok flere kvinner». Om jenter finner vi blant annet «kjøpe likedanne kjoler til jentene» og «være omgitt av et harem av lettkledder jenter». Menn på sin side er plassert i en patiensrolle i brukseksempel som «kvinnen holdt mannen helt utenfor planleggingen» og «hun har dressert mannen sin godt». Eksempel med gutter er «han tok et knipetak rundt halsen på gutten» og «hun var blitt kjær i en gutt».

4.3. En komparativ analyse av de to utgavene

4.3.1. Kjønnrepresentasjon

Revisjonen av BOB har medført en økning på 70 % flere brukseksempel som inneholder *jente*, *kvinne*, *gutt* og *mann*. Det er en overrepresentasjon av MASK i begge utgavene av ordboken. I BOB2018 forekommer gutter og menn i 60 % av alle undersøkte brukseksempel, mot 59 % i BOB2024. Denne kjønnskeivheten kan også beskrives med kvotienter, som vist i tabell 5. Kvotienten regnes ut ved å dividere forekomsten av den ene kjønnskategorien med den andre (Trap-Jensen 2020b:152). Når kvotienten er over 1, slik den er i dette tilfellet, viser det en dominans av MASK over FEM.

Kjønnskategori	BOB2018	BOB2024
MASK	212	349
FEM	139	247
Kvotient MASK/FEM	1,5	1,4

Tabell 5: Kvotienter for fordeling av kjønnskategoriene MASK og FEM i de to utgavene av BOB.

At menn gis mer plass enn kvinner i ordbøker, påvises også i tidligere studier (se f.eks. Hageberg 1990; Pilke 2009; Fjeld 2015; Larsson 2017). Et interessant unntak her er Fjelds undersøkelse fra 2015 som påviser at brukseksempel med *hun* og *han* har en nokså jevn kjønnsfordeling i BOB (2015:47). At dette funnet avviker fra funnene i min studie, viser at ingen av studiene gir et fullstendig bilde av kjønnsrepresentasjonen i ordboken, og at funn kan være forskjellig avhengig av hvilke kjønnsbeskrivende ord eller pronomen som undersøkes. Det fremhever dermed behovet for en mer omfattende studie som kan gi et mer representativt bilde av kjønnsfremstillingen og representasjonen i BOB.

4.3.2. Utseende

Det er flere brukseksempel med utseendebeskrivelser tilknyttet FEM enn MASK i begge utgavene av ordboken. Forskjellen har likevel blitt merkbart mindre etter revisjonen ved at differansen er justert ned med 6 prosentpoeng fra 14 % til 8 %. At det kvinnelige kjønn oftere beskrives utseendemessig, samsvarer med funn i en studie av brukseksempelene i *Svensk ordbok* (SO), som påviser flere utseendebeskrivelser tilknyttet *tjej* enn *kille* (Sköldberg 2020:15).

I BOB2018 forekom bare to brukseksempel om MASK som beskriver attraktivitet, mens antallet for FEM var 12. I BOB2024 har MASK 10 slike eksempler mot 14 for FEM. Balansen mellom kjønnene når det gjelder denne type utseendebeskrivelser, har dermed blitt jevnere etter revisjonen. Derimot innehar MASK flere negative utseendebeskrivelser enn FEM i begge utgavene. Dette er dermed en vedvarende skeivhet mellom kjønnene som har et uheldig utfall for MASK.

4.3.3. Karakterdrag

Totalt er det 31 % av alle brukseksempel med MASK som beskriver karakterdrag i begge utgavene av BOB. For FEM er andelen i BOB2018 27 % mot 28 % i BOB2024. Ettersom det forholdsvis er flere MASK- enn FEM-eksempel i begge utgavene, samsvarer dette med Sköldbergs funn i undersøkelsen av SO, som påviser flere brukseksempel med beskrivelser av karakterdrag tilknyttet det mannlige kjønn (2020:15).

Videre har revisjonen av ordboken medført en økning i antall negative beskrivelser av karakterdrag for begge kjønn, da både MASK og FEM har hatt en økning på 10 prosentpoeng flere eksempler av dette slag. Dertil er det en gjennomgående tendens til at MASK mer frekvent fremstilles med negative karakterdrag.

4.3.4. Kjønnnsrollemarkert tema

Andelen brukseksempel med kjønnnsrollemarkert tema har også

økt med revisjonen, da begge kjønn har fått 22 flere brukseksempel med denne klassifiseringen. I tillegg er enkelte eksempel omformulert. Blant annet finner vi «kvinner blir ofte referert til som datter eller kona til den og den» i BOB2018 som er endret til «før ble kvinner ofte referert til som datter eller kona til den og den» i BOB2024. Med denne omformuleringen signaliseres det at en slik referering er utdatert, og her antydes en spesifikk bevissthet hos leksikografene rundt språkbruk om kjønn.

4.3.5. Agens- og patiensrolle

Andelen brukseksempel der FEM er i agensrolle, har økt med 9 prosentpoeng etter revisjonen, fra 62 % til 71 %. Det samme har skjedd for MASK, som etter revisjonen har 76 % der det mannlige kjønn er i agensrolle, mot 66 % før revisjonen. Her antydes dermed en redaksjonell praksis som har medført at begge kjønn i større grad figurerer i agensrolle i brukseksempelene i den reviderte utgaven av ordboken.

Derimot er det en tendens i begge utgavene at FEM plasseres oftere i patiensrolle enn MASK. En lignende kjønnsforskjell mellom aktive og passive subjekt påvises i tidligere studier (se f.eks. Fjeld 2015:48; Sköldberg 2020:14–15), og dette antyder at forholdet mellom kvinnen som det passive objekt mot mannen som det aktive subjekt er vedvarende kjønnsstereotypisk i flere ordbøker.

5. Oppsummering

Studien presentert i denne artikkelen påviser at det er en tendens til at BOB gir en mindre kjønnsstereotypisk fremstilling av kjønn i brukseksempel etter den gjennomgående revisjonen, men også at gutter og menn er mer synlige i brukseksempelene i begge utgavene av ordboken. Dertil viser de kvalitative funnene at ordboken tegner et noe forskjellig bilde av kjønnene. Generelt forekommer

utseendebeskrivelser oftere om FEM, mens negative utseendebeskrivelser er vanligere om MASK. Ordboken har også flere negative beskrivelser av karakterdrag om MASK. Revisjonen har dertil medført flere brukseksempel med kjønnsrollemarkert tema for begge kjønn, men også plassert både FEM og MASK i agensrolle oftere.

Her bør det likevel gjentas at studiens avgrensninger medfører at dette ikke er et fullstendig bilde av kjønnene i ordbokens brukseksempel. Den metodiske fremgangsmåten kan også sies å ha svakheter, dels ved at kategoriene som er benyttet, ikke har absolutte definisjoner og eventuelt kunne vært supplert med flere og andre nyanseringer, og dels ved at det kvantitativt først og fremst er frekvenser som undersøkes. En fremtidig studie som inkluderer begge standardordbøkene samt flere kjønnsbeskrivende ord og pronomen er dermed formålstjenlig.

Som deskriptiv ordbok skal BOB beskrive den reelle språkbruken i vår samtid. Å fjerne uheldige brukseksempel for å unngå reproduksjon av kjønnsstereotyper kan medføre at man i for stor grad fraviker dette. Samtidig bør et nyansert og variert bilde av kjønnene etterstrebes (Sköldberg 2020:23). Et enkelt brukseksempel er ikke nødvendigvis problematisk i seg selv, men gjennomgripende mønster kan medføre at ordboken som helhet fremstiller kjønnene eksempelvis i foreldede kjønnsroller. Slik det redaksjonelle revideringsarbeidet foregår, med flere leksikografer som jobber parallelt, er det vanskelig for hver enkelt å vite hvordan helhetsbildet ser ut. Det er dermed en reell utfordring, og kanskje til og med en umulighet, å unngå reproduksjon av kjønnsstereotyper eller skeivheter i kjønnsrepresentasjonen. Til tross for noe redaksjonell bevissthet rundt forekomsten av skeivheter i kjønnsbalansen i forbindelse med revisjonen av standardordbøkene (se Selback et al. 2025) kan en større bevissthet rundt dette være en fordel i det videre leksikografiske arbeidet.

Litteratur

Ordbøker, korpus og digitale resurser

- BOB = *Bokmålsordboka*. Nettutgave (1994–): <<https://ordbokene.no/>> (mars 2025). Språkrådet og Universitetet i Bergen.
- NAOB = *Det Norske Akademis ordbok*. Oslo: Det Norske Akademi for Språk og Litteratur. <<https://naob.no/>> (juni 2025).
- NOB = *Nynorskordboka*. Nettutgave (1994–): <<https://ordbokene.no/>> (mars 2025). Språkrådet og Universitetet i Bergen.
- INESS = Rosén, Victoria, Koenraad De Smedt, Paul Meurer & Helge Dyvik (2012): «An Open Infrastructure for Advanced Treebanking». I: Jan Hajič, Koenraad De Smedt, Marko Tadić & António Branco (red.): *META-RESEARCH Workshop on Advanced Treebanking at LREC2012*. Istanbul, Turkey. 22–29.
- Revisjonsprosjektet. Universitetet i Bergen. <<https://www.uib.no/lle/revisjonsprosjektet>> (mars 2025).

Annen litteratur

- Chen, Wenge (2019): Towards a Discourse Approach to Critical Lexicography. I: *International Journal of Lexicography* 32 (3), 362–388. <<https://doi.org/10.1093/ijl/ecz003>>.
- Fjeld, Ruth Vatvedt (2001): Ingenting å snakke om – kvinneord og ordbøker. I: *Språknytt* 29 (4), 10–13.
- 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*. Institutet för de inhemska språken 39. 35–65.
- 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*. Surabaya: Airlangga University Press. 360–361.
- Grønvik, Oddrun (2002): Er det stygt å seie neger? Om konnotasjonar, bruksmarkering og nemne for folkegrupper. I: *LexicoNordica* 9, 149–172. <<https://tidsskrift.dk/lexn/article/view/18721>>.
- Hageberg, Arnbjörg (1990): Kvinner og menn og ordbøker. I: *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: Harry Lönnroth & Kristina Nikula (red.): *Nordiska studier i lexikografi 10. Rapport från Konferensen om lexikografi i Norden. Tammerfors 3-5 juni 2009*. 209–222.
- Larsson, Lennart (2017): Vilken plats hade kvinnan?: Om grammatiskt och sexuellt genus i SAOL:s sjunde upplaga 1900. I: *Språk och stil* 27, 49–66.
- Lyse, Gunn Inger & Terje Svardal (2020): «Den raude rase» og «negerfolk»: kontroversielle ord i *Bokmålsordboka* og *Nynorsk-ordboka*. I: *LexicoNordica* 27, 77–96. <<https://tidsskrift.dk/lexn/article/view/123217/175471>>.
- Mattisson, Anki (2006): Genusperspektiv på SAOB:s källor. I: *LexicoNordica* 13, 55–68.
- Milles, Karin (2009): ”Klitoris orsakar orgasm”. Kön och sexualitet i svenska ordböcker. I: *Tidsskrift för genusvetenskap* 30 (2–3), 119–137. <<https://doi.org/10.55870/tgv.v30i2-3.3748>>.
- 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. 15-19 July 2014*. Bolzano/Bozen: Institute for Specialised Communication and Multilingualism. 85–105.

- Nikula, Henrik (1995): Exemplens funktion i ordböcker. I: *Nordiske Studier i Leksikografi* 3, 311–319. <<https://tidsskrift.dk/nsil/article/view/19652>>.
- Nikula, Kristina (1997): Språkprov i könsperspektiv. I: Rune Ingo et al. (red.): *Fackspråk och översättningsteori. VAKKI-symposium XVII*, 194–206.
- Nimb, Sanni (2024): Fra ‘sandheden om sproget’ til et opgør med stereotyper: ikkekorpusbaserede metoder til leksikografiske beskrivelser af kontroversielle ord. I: *LexicoNordica* 31, 151–174. <<https://tidsskrift.dk/lexn/article/view/151289>>.
- 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: *Language and Power. VAKKI Symposium XXIX*, 332–342.
- Rauset, Margunn (2022): Brukarmedverknad i utvikling av nettsida ordbøkene.no. I: *LexicoNordica* 29, 97–118. <<https://tidsskrift.dk/lexn/article/view/134777>>.
- Selback, Bente, Klara Sjø, Knut E. Karlsen & Sturla Berg-Olsen (2025): Revisjonen av *Bokmålsordboka* og *Nynorskordboka*: oppdatering og tilpasning av ordbøkene til en ny tid. I: *LexicoNordica* 32 (dette bind).
- Sköldberg, Emma (2020): En mer inkluderande lexikografi – bortom frågan om kvinnor och män. I: *Oslo Studies in Language* 11 (1), 7–30. <<https://doi.org/10.5617/osla.8165>>.
- Stefanowitsch, Anatol (2020): *Corpus Linguistics: A Guide to the Methodology*. Language Science Press.
- Tonne, Ingebjørg, Helene Uri & Lars Johnsen (2022): Om kjønn og adjektiv: ‘Trond er så eksepsjonell/nydelig/skjønn/grønn’. I: *Nordlyd* 46 (1), 241–257. <<https://doi.org/10.7557/12.6386>>.
- Trap-Jensen, Lars (2020a): Inklusion eller mindretalsdiktatur? Om politisk korrekthet, minoritetshensyn og leksikografisk deskriptivisme i *Den Danske Ordbog*. I: *LexicoNordica* 27, 137–156. <<https://tidsskrift.dk/lexn/article/view/123220>>.

- Trap-Jensen, Lars (2020b): Skævheder og ligestilling: i ordbogen, i korpus og i samfundet. I: *Oslo Studies in Language* 11 (1), 31–46. <<https://journals.uio.no/osla/article/view/8166/7258>>.
- Vacalopoulou, Anna (2022): Gender Stereotypes in Greek Children's Dictionaries. I: *Dictionaries: Journal of the Dictionary Society of North America* 43 (1), 167–192.

Emma Josefin Ölander Aadland
Universitetsbibliotekar
Universitetet i Bergen
Universitetsbiblioteket
Postboks 7808
NO-5020 Bergen
emma.aadland@uib.no

IKKJE-TEMATISKE BIDRAG

Har de kommit för att stanna? Om nyordslistornas ord i SO och SAOL

Andreas Brimberg & Linnea Hanell

The present study investigates the extent to which neologisms from the yearly lists published by the Language Council of Sweden, and its predecessors, later occur in contemporary dictionaries published by the Swedish Academy. The results show that among the 1,504 words published in the neologisms lists from 1986 to 2020, 46% have later occurred in an edition of *Svensk ordbok* or *Svenska Akademiens ordlista*. These results are discussed in terms of their ability to indicate whether the words in the yearly neologisms lists have come to stay in the Swedish language, which is a recurrent question from the public when the yearly list of new words is presented.

1. Inledning

Språkrådet publicerar årligen en nyordslista med en samling nytillskott i det svenska lexikonet. Publiceringen av nyordslistan, som sedan år 2013 sker i samarbete med Språktidningen, får ofta medial uppmärksamhet och företrädare för nyordsredaktionen blir intervjuade i tidningar, teve och radio. En fråga som ofta återkommer är hur många av nyorden som har kommit för att stanna i det svenska språket, och hur många som kommer att försvinna. Frågan kan förstås besvaras på olika sätt. En antydan till svar ges i en undersökning av Språktidningens nyordsredaktör Anders Svensson (2017), som visar att ungefär en tredjedel av nyorden från listorna 2007–2014 kom med i upplagan av *Svenska Akademiens ordlista* (SAOL) som publicerades 2015.

Svanlund (2009) analyserar ords tidiga etablering i språket utifrån nio ord, av vilka fyra varit med på Språkrådets nyordslista. Svanlund visar att ordens etableringsgrad tycks hänga sam-

man med huruvida orden upplevs som intuitivt transparenta, men också med deras förmåga att benämna fenomen som är aktuella i samhällsdebatten. Hanell (2023) undersöker 132 nyord från miljö- och klimatområdet, bland annat med hänsyn till spridning i tidningsspråk, och visar att 109 av de 132 undersökta orden (= 83 %) har använts minst 100 gånger i tidningar över tid, och att 16 av orden (= 12 %) har använts minst 10 000 gånger. Utöver detta har några studentuppsatser behandlat etableringen av och förståelsen för några nyord (t.ex. Palmblad 2008, Abellsson 2016, Bidemo 2017, Wikström 2017, Lönnfors 2021). Svensson (2020) går igenom nyorden mellan 1986 och 2019, främst med fokus på allmänna karaktärsdrag för orden, till exempel ordklass, inlåningsspråk, ordlängd och så vidare. Mer heltäckande underlag för att besvara frågan om nyordens etableringsgrad har dock saknats, vilket är motiveringen till att föreliggande studie görs.

Vi följer Svensson (2017) i att analysera etableringsgraden genom att undersöka hur många av orden från nyordlistorna som tas med i ordböcker. I föreliggande studie undersöks hur stor del av samtliga nyord från listorna 1986–2020 som också finns med i någon av Svenska Akademiens samtidsordböcker, alltså *Svensk ordbok* (SO) från 2009 och 2021 respektive *Svenska Akademiens ordlista* (SAOL) från 1986, 1998, 2006 och 2015. Studien undersöker dessutom hur lång tid det är mellan att ett ord tas med i nyordlistan och att det tas med i en ordbok.

Artikeln är strukturerad som följer. Först beskrivs materialet, som utgörs av nyordlistorna samt ordböckerna SO och SAOL. Därefter redogörs för metoden, med fokus på vad som i undersökningen räknas som nyord och när dessa anses vara med i en ordbok. Därefter redovisas resultaten, följt av en diskussion av gränfall och några avslutande slutsatser.

2. Material

I detta avsnitt beskrivs dels nyordslistorna, dels Svenska Akademiens samtidsordböcker.

2.1. Nyordslistorna

Den organisation som idag heter Språkrådet har sedan starten arbetat med att beskriva och analysera nyheter i språket. När organisationen ursprungligen grundades år 1944, under namnet Nämnden för svensk språkvård, hette det i stadgarna att nämnden hade till uppgift att ”följa det svenska språkets utveckling” (Hänninger 1969:7), och redan i början registrerades just nya ord som ett led i detta arbete. Organisationen fick ny form 1974 och blev då en delvis statlig stiftelse, som hade namnet Svenska språknämnden. Denna organisation började år 1986 publicera en årlig lista med rubriken ”Några nyare ord i svenskan” i sin tidskrift *Språkvård*. Svenska språknämnden uppgick år 2006 i den nya avdelningen Språkrådet inom myndigheten Institutet för språk och folkminnen, och det är sedan dess Språkrådet som årligen publicerar nyordslistan. Sedan 2013 tas listan fram i samarbete med den populärvetenskapliga tidskriften *Språktidningen*. Utöver listorna har det dessutom getts ut tre böcker baserat på nyordsmaterial från Språkrådet och dess föregångare (Svenska språknämnden 1986, Moberg 2000, Agazzi 2015).

I nyordslistan publiceras ett urval av ord eller fraser som tolkas som nya i språket. Antalet har varierat mellan åren, men har på sistone varit omkring 30. Ett nyord, eller med lexikografisk terminologi en *neologism*, kan definieras som en ”lexikalisk enhet som vid en viss tidpunkt anses vara nyskapad, eller en ny betydelse hos en existerande lexikalisk enhet” (NLO 1997, vår översättning). För nyordslistan har detta operationaliserats som ord eller fraser som har bedömts som helt nya i språket, eller som har bedömts som

relativt nya på grund av att de gått från fackspråk till allmänspråk eller från att användas mycket sällan till att bli betydligt mer frekventa (Karlsson 2021:106). Enligt Karlsson (2021) är avsikten att nyordslistan ska spegla det gångna året och visa på en variation av både ämnesområden och ordbildningsmönster. Även om nyordslistans huvudsakliga syfte är att beskriva det faktiska språkbruket är det inte någon uttalad avsikt att endast ta med sådana ord som kan komma att tas med i en ordbok.

Metoderna för att identifiera kandidater och sammanställa listan har utvecklats över åren, men övergripande kan sägas att arbetet är av manuell karaktär: nyordsredaktörerna (historiskt en, två eller tre till antalet) bevakar språkbruket i sin vardag och i media, noterar potentiella nyord och kontrollerar sedan dessa mot databaser över samtida språkbruk – i första hand skriftliga och delvis muntliga massmediala källor – för att säkerställa att orden verkligen används och att de verkligen är nya (Karlsson 2021). I dagsläget används främst Retrievers databas Mediearkivet, medan äldre listor sammanställdes utan tillgång till denna typ av digitala verktyg. Förutom redaktörernas egna observationer och förslag från andra språkvårdare och språkvetare, liksom från allmänheten, nyttjas även sökningar i databaser efter metalingvistiska uttryck (till exempel ”så kallad” eller ”det som numera heter”) som en ytterligare metod för att hitta nyordskandidater. Några automatiserade excerperingsmetoder används för närvarande inte, huvudsakligen på grund av begränsade tekniska resurser. Samtliga nyordslistor finns på Språkrådets webbplats (Institutet för språk och folkminnen 2024).

Materialet för den här studien sträcker sig från den första nyordslistan för år 1986 till nyordslistan för år 2020, alltså till och med året före utgivningen av den senast publicerade ordboken i materialet (se avsnitt 2.2). Det innebär att 1 504 nyord ingår i studien.

2.2. Ordböckerna

Ordböckerna som nyorden jämförs mot är de utgåvor av Svenska Akademiens samtidsordböcker som har getts ut sedan nyordslistorna började publiceras och fram till genomförandet av denna studie hösten 2024. Detta innefattar *Svenska Akademiens ordlista*, SAOL, utgåvorna från 1986, 1998, 2006 och 2015 samt *Svensk ordbok utgiven av Svenska Akademien*, SO, utgåvorna från 2009 och 2021. Dessa ordböcker har valts på grund av deras centrala roll som normauktoriteter för det samtida svenska språkbruket.

Den allra första upplagan av SAOL kom 1874 och sedan dess har ytterligare tretton utgivits. Ordlistan samlar många uppslagsord med fokus på stavning, böjning och ordbildning (SAOL 2015:XI) och har kommit att bli den inofficiella normen för detta (Gellerstam 2009a:11). Genom åren har redaktionen haft olika inställning till exempelvis lånord, vardagliga ord och sammansättningar, vilket har gjort att ordlistans omfång har varierat (jfr Gellerstam 2009b). De senaste årtiondena har tekniska verktyg förändrat arbetet med att undersöka språkbruket, men de normativa inslagen är fortfarande viktiga (SAOL 2015:X–XII). Den senaste utgåvan av SAOL innehåller omkring 126 000 uppslagsord.

SO innehåller färre uppslagsord, omkring 65 000, och är inriktad på ords betydelse, användning och historia (SO 2009:VII). Det är en definitionsordbok (Sköldberg & Hannesdóttir 2017:330). Ordboken beskrivs som deskriptiv till skillnad från den normativa SAOL (Sköldberg 2022:141–142). Till den andra upplagan år 2021 arbetades med att ”renodla SO:s deskriptiva karaktär” (Sköldberg 2022:143) genom att exempelvis ta bort de texttrutor med stilrekommendationer som tidigare fanns och genom att försöka spegla flera sorter av bruket som det ser ut (Sköldberg 2022:144–145). Även i detta arbete är tillgången till stora och breda korpusar viktig (Sköldberg 2022:142).

De sex aktuella ordböckerna samlar minst 136 000 unika

uppslagsord plus 5 000 idiom och 18 000 morfologiska exempel. Exakta siffror är svåra att få fram, men totalt kan det röra sig om uppemot 150 000 unika uppslagsord. De fyra upplagorna av SAOL innehåller mellan cirka 115 000 (1986) och 126 000 uppslagsord (2015) (Gellerstam 2009b:56; Svenska Akademien 2024). Till de nya upplagorna (från 1998, 2006 och 2015) har 5 000, 10 500 respektive 13 000 ord lagts till (Gellerstam 2009b:75, 78; SAOL 2015:X), vilket gör att de fyra upplagorna av SAOL tillsammans borde bestå av runt 140 000 unika ord.

SO innehåller ungefär 65 000 uppslagsord i båda utgåvorna (SO 2009:VII; Svenska Akademien 2024). Även här har ord både tillkommit och utmönstrats (jfr Sköldberg 2022:145) men exakta antalsuppgifter om detta finns inte. Utöver detta finns i (den första utgåvan av) SO 5 000 idiom och 18 000 morfologiska exempel (främst sammansättningar) (SO 2009:VII), vilka utöver uppslagsorden är sökbara genom webbgränssnittet SAOLhist (se vidare avsnitt 3.3).

Ungefär 55 200 uppslagsord är gemensamma mellan SAOL (14) och SO (2021). Ordböckerna har något parallella utvecklingslinjer: SAOL:s utgåvor har sin egen historia (jfr Gellerstam 2009b) medan SO följer upp andra ordböcker (Sköldberg & Hannesdóttir 2017:330). Ordböckerna har numera likheter i vilka korpusar som orden hämtas från (Sköldberg & Hannesdóttir 2017:330; jfr Josephson 2022:255). I arbetet med att ta fram respektive ordbok jämförs de också med varandra, för att till exempel hitta nya ord (SAOL 2015:XIII; Sköldberg & Hannesdóttir 2017:334). En skillnad som är särskilt framträdande i denna undersökning är att SO inkluderar konstruktioner, uttryck och sammansättningar som inte tas upp i SAOL. Vidare är en skillnad att SAOL har status som en mer normativ ordlista som ger information om stavning, böjning, ordklass och i viss mån uttal och stilnivå. SO har snarare status som en beskrivande betydelseordbok och ger, till skillnad från SAOL, utförliga betydelsebeskrivningar, användningsexempel och

en historisk bakgrund. På så sätt kompletterar ordböckerna varandra och täcker därmed tillsammans en stor del av det offentliga samtida språkbruket.

3. Metod och materialbearbetning

I detta avsnitt presenteras de metoder och principer som har använts i studiens materialbearbetning.

3.1. Vad räknas som ett nyord?

I undersökningen ingår de nyord som publicerades i listorna från år 1986 till 2020. Som en del av undersökningen behöver några avvägningar göras om vad som, framför allt i äldre nyordslistor, ska räknas som ett nyord.

Ett nyord definieras här som a) en separat nyordspost som b) inte har en uppenbar överlappning med en annan nyordspost. Det första kriteriet utesluter ord inom en post. Därför räknas inte *genusforskning* som ett nyord, utan ingår i nyordet *genus*. På samma sätt räknas *portfoliometoden* och *portföljmetoden* som varianter av ett och samma nyord. Sådana varianter av ord och bildningar med ett visst ord är vanliga i de tidiga nyordslistorna. Generellt räknas också poster där flera närliggande betydelser är listade som ett nyord. Det andra kriteriet innebär att 22 nyordsposter inte räknas som nyord i denna undersökning, utan ingår i andra nyord. Dessa är av fem typer i förhållande till nyordet: ord som ingår i sammansättningar (som *muff-tankningsmuff*), stavningsvarianter (*beigel-bagel*), uttrycksvarianter (*emotionell intelligens-känslans intelligens*), ord som ingår i uttryck (*garderob-komma ut [ur garderoben]*) och förkortningar (*BO-Barnombudsmannen*).

Dessa avgränsningar minskar antalet tillfällen då nyord räknas dubbelt och gör urvalskriterierna mer enhetliga över tid (i mån-

ga listor listas t.ex. inte två ord där bara stavningsskillnad föreligger, och många gånger lyfts huvudordet i ett uttryck inte ut som ett eget nyord, till exempel i fallet *ligga på framkant*). Däremot är synonymer som listas som egna nyord också separata nyord i denna undersökning, exempelvis *gåtåg*.

Frågor om vad som ska räknas som ett nyord uppstår ofta till följd av utformningen av nyordslistorna. Som framkommit listas ofta likartade ord antingen inom samma post eller som egna poster. När det handlar om egna poster sker det med olika referenser till de likartade eller relaterade orden, genom markeringar som ”se”, ”jfr” eller ”= [annat ord]”. Då är det inte givet om orden ska räknas som separata nyord eller inte. Listorna är inte heller helt enhetliga. Till exempel är det inte självklart varför det på listan 2005 är två separata nyordsposter för *blogga* och *bloggare* medan *podda* och *poddare* ingår i samma nyordspost. Särskilt svårt är det att värdera vilka ord som ska räknas som nyord i nyordslistan 1998 som är skriven som en krönikeliknande text snarare än en lista. Där görs inte tydliga skillnader mellan vad som anses vara ett av årets nyord och vad som tas upp i samband med andra ord. I slutet av texten listas förvisso ”nya ord och uttryck” som ”behandlas i artikeln”. Där listas dock inte ord som *partism* och *högstres-arbete* som tas upp i texten, medan *outstanding* listas trots att det i texten står att ordet ”sedan länge ... förekommit som direktlån i svenskan”.

Avvägningar kring vad som ska betraktas som ett nyord förefaller alltså ha varierat över tid, vilket också framgår av flera av nyordslistornas förord, och detta försvårar möjligheten att genomföra en sådan här undersökning på ett konsekvent sätt. De första åren (1986–1988) skrevs att listan innehöll ett antal ord ”som vi uppfattat som relativt nya i svenskan” och att källan ”är huvudsakligen dagspressen”. Denna formulering vidgades till att mellan 1989 och 1999 (undantaget 1998) hänvisa till ”ord och fraser som uppfattats som relativt nya i svenskan och som varit aktuella i

svensk dagspress [vissa år: svenska medier] under [årtalet]”. Vidare står det att ”Anledningen till att dessa ord [som har belägg sedan tidigare] nu har kommit med kan vara att de märkbart ökat i användning eller att de används på ett annat sätt än tidigare”. Mellan 2001 och 2004 är formuleringen likartad, med skillnaden att den förklarande andra meningen inte finns med. Vad som anses och har ansetts vara ett nyord ger många frågor och ett arbete pågår för att göra det mer enhetligt och konsekvent (Karlsson 2021).

3.2. Sökning och inmatning

Undersökningen utgår från en sammanslagen lista över samtliga 1 504 nyord. För varje nyord gjordes en sökning på databasen SAOLhist, vilket möjliggör sökning i fem av ordböckerna samtidigt (liksom i tidigare upplagor av SAOL, men inget nyord förekommer i en tidigare upplaga än 1986). Förekomst i SO 2021 söktes på svenska.se eller i ordbokens app. I många fall räckte det med en sökning på själva uppslagsordet för att få fram det i ordböckerna. Vid andra tillfällen krävdes sökning på ett av orden i ett flerordsuttryck (t.ex. *politisk* för att hitta *politiskt korrekt*). Ibland måste sökning med jokertecken göras i någon del av ordet, exempelvis för att få träffar på flera stavningar (sökningen *bloddop%ing* ger träffar på både *bloddoping* och *bloddopning*) eller på flera ordformer (sökningen *geotagg%* ger träffar på *geotagga* och *geotaggning*; se vidare avsnitt 3.3 om bedömning av ordformer). På liknande sätt söktes ord i SO 2021. För varje nyord registrerades på så sätt huruvida det förekommer eller inte i var och en av de sex ordböckerna.

3.3. När räknas ett ord som medtaget i en ordbok?

Huruvida ett nyord ska anses vara med i en ordbok är inte alltid givet. Här ges exempel på när nyord räknas som medtagna i ordböckerna och vilka bedömningar som har gjorts för att avgöra det.

De enklaste fallen att kategorisera är nyord som är helt nya och överensstämmer med ordböckers poster. Ett sådant exempel är *maskrosbarn*, som var med på nyordslistan 1986 och finns i ordböckerna SAOL 1998, 2006, 2015 och SO 2009, 2021. I de fall flera uttryck listas som likvärdiga i en nyordspost i nyordslistan (som *störningsjour* och *störjour*) räcker det med att ett av dessa finns i en ordbok för att nyordet ska räknas som medtaget.

Nyordslistan tar ofta med flerordsuttryck (som *höja ribban*) som egna poster. I SAOL och SO brukar sådana uttryck inte ha egna poster, men om uttrycket nämns som språkprov eller idiom i ordboksposten för ett av de ingående orden (som *ribba*) räknas det som medtaget.

I SO listas många avledningar och sammansättningar av ord under annat uppslagsord, och dessa ord ger i sin tur träff i ordboken vid digital sökning. Ett exempel är *livsstilsmagasinet* som i båda utgåvorna av SO listas som en sammansättning under uppslagsordet *livsstil*. Dessa ord räknas som medtagna.

Om ett ord är med i nyordslistan i samband med att det har genomgått en betydelseförändring, är det förstås denna nya betydelse som ska finnas angiven även i ordböckerna. Ett exempel är ordet *klassresa*: i den äldre betydelsen 'resa med skolklass' finns ordet med i SAOL från 1950 års upplaga, men den nya betydelsen 'byte till högre socialgrupp' togs med i nyordslistan 1991 och i SAOL 2006. Även *global* (nyord 1990) finns i SAOL 1950 men den nya, bildliga, betydelsen tas upp i SO 2009 och 2021.

Ordet betraktas som detsamma även om stavningen eller skrivsättet skiljer sig mellan nyordslistan och ordboken: nyordslistans *air bag* är i ordböckerna *airbag*, nyordet *5:2-diet* är *fem-tvådiet* och *HIV* har blivit *hiv*.

Ordformen i nyordslistan behöver inte överensstämma exakt med den i ordböckerna för att räknas som förekomst. Som regel betraktas ett ord som medtaget även om det förekommer i en böjd eller avledd form. Exempelvis anses *kravallturism* från ny-

ordslistan ha motsvarighet i ordboksformen *kravalltulist*, och *glokalisera* motsvaras av *glokalisering* i SAOL 2015. På samma sätt anses *fejkade nyheter* täckas av ordboksposten *fejknyhet* i SO 2021. Undantag görs dock när relationen mellan ordformerna inte är entydig. Om det inte är uppenbart att den ena formen förutsätter den andra, räknas de som separata ord. Ett exempel är *rejva*, som förekommer som nyord, men inte anses belagt i SAOL 2006 eller 2015, trots att *rejvare* finns med. Eftersom ordbokens personbeteckning *rejvare* skulle kunna ha härletts från substantivet *rejv*, och inte förutsätter verbformen *rejva*, behandlas de som skilda lexikala enheter.

Andra fall som kan vara svårbedömda är när ett för- eller efterled listas som ett nyord. Exempel på det är *-faktor*, *mega-* och *-peng*, och de påminner om det som Svensén (2004:173–175) kallar för *affixoider*, alltså ordled som befinner sig mellan fria morfem och bundna affix. I några fall finns ledet som ett eget uppslagsord i ordböckerna (som *-peng*, vars funktion som efterled tas upp i SAOL 1998, 2006, 2015) och kan då enkelt bedömas som medtaget. Andra gånger anges till exempel att ordbildningar med ett visst för- eller efterled har ökat. Sökningar med jokertecken (t.ex. *mega%*) har använts för att identifiera konstruktioner med det aktuella ledet, och ett nyord som utgörs av ett led har i studien bedömts som etablerat om minst fem sådana konstruktioner återfunnits i ordboken.

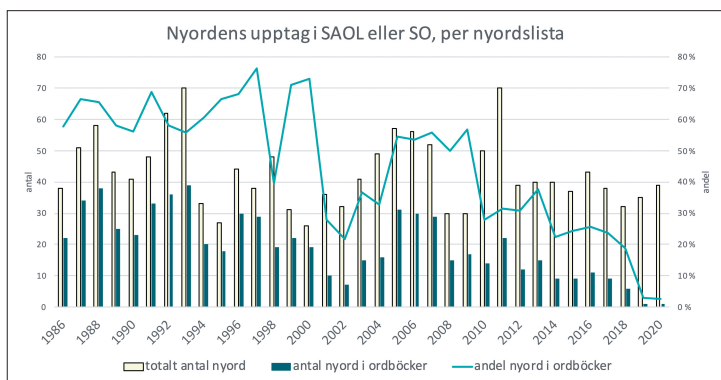
4. Resultat

Undersökningen visar att av de 1 504 nyord som listades mellan 1986 och 2020 återfinns 686 i någon utgåva av SAOL eller SO, motsvarande cirka 46 procent. Detta inkluderar 30 ord som har tagits med i en utgåva av SAOL eller SO och sedan strukits i senare utgåvor. 662 av nyorden, alltså 44 procent, finns med i den senaste

utgåvan av SAOL (2015) eller SO (2021). Totalt 6 av samtliga nyord var redan upptagna i SAOL eller SO när de listades som nyord.

En närmare undersökning av nyordens förekomst i SAOL får begränsa sig till listorna som hade publicerats före den senaste utgåvan från 2015. Av de 1 280 nyorden mellan 1986 och 2014 finns 549 med i någon utgåva av SAOL, vilket utgör 43 procent. I den senaste utgåvan av SAOL från 2015 återfinns 537 av nyorden, eller 42 procent.

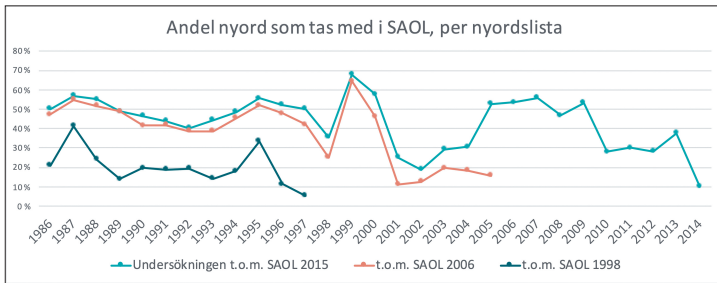
Med utgångspunkt i respektive lista kan undersökningen vidare ge en indikation om hur stor andel av orden i en nyordslista som brukar tas upp i Svenska Akademiens samtidsordböcker. Här uppvisar nyordslistorna stor variation över tid, med utfallet att mellan 3 och 76 procent av nyorden på listorna har tagits med i SAOL eller SO. Medelvärdet är 45 procent och medianen är 54 procent. Figur 1 visar antal respektive andel nyord i respektive lista som tagits med i SAOL eller SO.



Figur 1: Nyordens förekomst i SAOL eller SO, per nyordslista.

Undersökningen visar också att det för de 549 nyordslisteord som finns med i någon utgåva av SAOL, skiljer i genomsnitt nio år mellan året då ordet listas i nyordslistan och året det kommer med i

SAOL. Det innebär att ord från de nyordslistor som publicerats de sista åren före en ny upplaga av SAOL har mindre sannolikhet att inkluderas i den upplagan. Det innebär också att utfallet för de senaste nyordslistorna i föreliggande studies material sannolikt kommer att förändras när fler ordböcker utkommer. Figur 2 visar andel nyord från nyordslistorna som finns med i utgåvorna av SAOL från 1998, 2006 och 2015.



Figur 2: Andel nyord som tagits med i olika utgåvor av SAOL, per nyordslista.

Majoriteten av orden som är upptagna i SAOL finns också med i SO och vice versa, även om ordböckerna tillsammans täcker en större del av språkbruket än de gör enskilt. Det kan åskådliggöras med att undersöka nyorden fram till och med 2014, där 649 av 1 280 finns med i någon SAOL eller SO. Av dessa förekommer 396 i båda ordbokstyper, 153 bara i någon SAOL och 100 bara i SO. När ett nyord finns i SO men inte i SAOL beror det oftast på att det är ett flerordsuttryck som nämns som språkprov, något som inte förekommer i SAOL. När ett nyord finns i SAOL men inte i SO beror det oftast på att SAOL helt enkelt har betydligt fler uppslagsord.

5. Diskussion av gränsfall

Trots de avvägningar som har gjorts (se avsnitt 3) finns gränsfall i bedömningen av en del ord. Några exempel på det följer. I dessa gränsfall, som utgör de mest svårbedömda i undersökningen, har hänsyn tagits till att bedömningarna bör fördela sig relativt jämnt, så att exempelvis inte samtliga tveksamma fall kategoriseras som medtagna.

Ibland är nyordslistans förklaring så knapphändig att den inte enkelt kan jämföras med ordböckernas. Ett sådant ord är *arena* (1989) som har förklaringen ”ökad användning i bildlig betydelse” och exemplet ”Rhonedalen ... är en arena som kan bjuda på lika oväntade som behagliga upplevelser när det gäller skidorter”. Den bildliga betydelsen tas upp i SO 2009 och 2021 med exemplet ”NN lämnade hastigt den politiska arenan”. Det går inte att med säkerhet säga att dessa bildliga användningar är desamma. Något som talar för det är dock att SO anger att första belägg är 1985, alltså före nyordslistan, och nyordet får därför anses vara med i SO 2009 och 2021.

Vid andra tillfällen väcks frågor om vilka betydelser som ett ord kan anses täcka. Ett exempel är nyordet *subsidiaritet* (1990), som bara finns med i SAOL 2006. Däremot är *subsidiaritetsprincip(en)* med i alla ordböcker sedan SAOL 1998. I nyordslistan definieras *subsidiaritet* som ’principen att på central nivå behandla endast frågor som inte kan avgöras på lokal nivå’ (vår kursivering) vilket liknar ordböckernas definitioner av motsvarande ord med efterled *-princip(en)*. I Nyordsboken (Moberg 2000) räknas *subsidiaritet* och *subsidiaritetsprincip* som två olika ord, vilket gör att det förra är det enda som godkänns för att räknas som medtaget i en ordbok.

Ibland listas ord i nyordslistor som uppges ha fått ökad popularitet eller användning. Huruvida dessa ord ska räknas som medtagna i ordböcker är svårt att avgöra. Ett exempel är *bröllopsklänning*

(2011) som uppges ha fått ”ökad användning i stället för brudklänning”. Ordet är med i alla ordböcker från och med SAOL 1998. På liknande sätt är det med *spoilervarning* (2014) som vid tiden hade fått ”ökad användning” och finns med i både SO 2009 och 2021. I dessa fall är det svårt att säga något om nyordens genomslag i ordböcker, eftersom det för ordböckerna inte egentligen handlar om nya ord vid tiden för nyordslistans publicering. Fallen är dock få till antalet, och de räknas som nyord som också återfinns i ordböcker.

Andra besvärliga fall handlar om små variationer mellan nyordslistans post och de uppslagsord som återfinns i ordböcker. Huvudsakligen utmanar uttrycken *göra en hel pudel* och *ligga på framkant*, vilka är upptagna i ordböcker som *göra en pudel* och *ligga i framkant*. Själva uttrycken har ändrats marginellt, men betydelseerna kvarstår. De har räknats som nyord som återfinns i ordböcker.

Denna undersökning innehåller tolkande inslag men resultatet är kvantitativt. Undersökningen är gjord på ett material som består av hela den datamängd som är av intresse – inget stickprov har gjorts. Den osäkerhet som föreligger är därför en följd av urvalet av ord (vilka nyord från listorna som ska vara med) och tolkningen av vilka ord som är upptagna i ordböcker (gränsfallen). En överväldigande majoritet av orden är inte gränsfall och de fall som använts som exempel i detta avsnitt påverkar inte resultatet med mer än ett par procentenheter på sin höjd.

6. Slutsatser

Den som i nyordslistetider vill ha ett svar på frågan om hur stor andel av årets nyord som har kommit för att stanna får ett antal indikationer av den här studien. Utslaget på hela samlingen nyord som publicerats fram till den senaste samtidsordboken visar stu-

dien att 46 procent av nyorden har tagits med i någon utgåva av Svenska Akademiens samtidsordböcker SO och SAOL. Vill man vara lite mer återhållsam kan man i stället lyfta fram att 44 procent av nyordslistornas nyord finns listade i den nu aktuella utgåvan av SO eller SAOL. Den högsta siffran som kan lyftas fram kommer av en närmare undersökning av andelen nyord som kommit med i ordböcker i varje publicerad nyordslista, vilket visar att medianvärdet på detta är 54 procent – med 76 procent av nyorden från 1997 års lista som högsta notering.

Undersökningen visar vidare en nedåtgående trend för andelen nyord som tas upp i de undersökta samtidsordböckerna. Sett till hur trenden sett ut precis före publicering av en ny ordbok är det dock sannolikt att denna trend kommer att vända uppåt igen i samband med den kommande publiceringen av en ny utgåva av SAOL. Detta tycks hänga ihop med att orden behöver förekomma i bruket betydligt längre innan de anses lämpliga för en ordbok, medan nyordslistan strävar efter att publicera ord som *inte* har förekommit brett i bruket särskilt länge.

Även om studien ger indikationer på hur stor andel av nyorden som kommer att tas med i ordböcker kan det vara problematiskt att dra likhetstecken mellan förekomst i ordböcker och förekomst i språkbruket. Att 46 procent av nyorden har kommit med i en ordbok betyder inte nödvändigtvis att 46 procent av nyorden hade ”kommit för att stanna”, i bemärkelsen att de inte snabbt kommer att glömmas bort igen. Det kan inte uteslutas att nyordslistorna används som underlag i framtagandet av ordböcker; i själva verket återkommer någon gång formuleringar från nyordslistorna när orden beskrivs i ordböcker. Risken för cirkelresonemang föreligger alltså. En mer omfattande studie skulle med fördel kunna använda korpusundersökningar för att undersöka nyordslisteordens etableringsgrad i det faktiska språkbruket.

Litteratur

- Abelsson, Ebba (2016): *Är din bror obror? En studie av ordförståelsen och användningen av nyord*. Kandidatuppsats, Uppsala universitet. <<https://www.diva-portal.org/smash/get/diva2:1134922/FULLTEXT01.pdf>> (januari 2025).
- Agazzi, Birgitta (2015): *Nyord i svenskan: Blogg, fulbryt, pudla, rondellhund och andra nytillskott från A till Ö*. Stockholm: Morfem.
- Bidemo, Kajsa (2017): *Feministpappa, selfie, kungspudel och vinstvarning: En enkätstudie av 54 nyord i svenska språket under 2000-talet gällande ordkännedom, användningsfrekvens och nyordens etablering i svenskan*. Kandidatuppsats, Mälardalens universitet. <<https://www.diva-portal.org/smash/get/diva2:1129729/FULLTEXT01.pdf>> (januari 2025).
- Gellerstam, Martin (2009a): Vad är Svenska Akademiens ordlista? I: Martin Gellerstam (red.): *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm: Norstedts. 11–30.
- Gellerstam, Martin (2009b): SAOL i många upplagor. I: Martin Gellerstam (red.): *SAOL och tidens flykt. Några nedslag i ordlistans historia*. Stockholm: Nordstedts. 53–84.
- Hanell, Linnea (2023): Förändringens ord: Samtidshistoriska nyord om miljö, klimat och omställning. I: M. Nelson, M. Michanek, M. Rydell, S. Sayehli, K. Skogmyr Marian & G. Sundberg (red.): *Språk i praktiken – i en föränderlig värld. Rapport från ASLA-symposiet. Stockholms universitet, 7–8 april 2022*. ASLA, Svenska föreningen för tillämpad språkvetenskap. Stockholm. 202–222. <<https://doi.org/10.17045/sthlmuni.24321526>>.
- Hänninger, Nils (1969): *Nämnden för svensk språkvård 1944–1969*. Stockholm: Nämnden för svensk språkvård.
- Institutet för språk och folkminnen (2024): *Nyordslistor*. <<https://www.isof.se/svenska-spraket/nyord/lar-dig-mer-om-nyord/nyordslistor>> (december 2024).

- Josephson, Olle (2022): *Språkpolitik*. Stockholm: Morfem.
- Karlsson, Ola (2021): Lesserwisser, lårs kav och läppstiftseffekt – presentation och problematisering av urvalskriterierna för den svenska nyordslistan. I: *Lexiconordica* 28, 101–120.
- Lönnfors, Linda (2021): *Nyord – passé eller succé? En undersökning om hur 2010 års nyord etablerat sig i dagens samhälle*. Kandidatuppsats, Mälardalens universitet. <<https://www.diva-portal.org/smash/get/diva2:1569420/FULLTEXT01.pdf>> (januari 2025).
- Mediearkivet = Tillgänglig: <<https://www.retriever.se/tag/mediearkivet/>> (augusti 2025).
- Moberg, Lena (2000): *Nyordsboken: Med 2000 nya ord in i 2000-talet*. Stockholm: Svenska språknämnden.
- NLO 1997 = Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén (red.): *Nordisk leksikografisk ordbok*. Oslo: Universitetsforlaget.
- Palmblad, Anna (2008): *Svenska nyord 1985–2007. En studie av 53 svenska nyord (1985–2007) gällande användningsfrekvens, ordförståelse och betydelseförändring*. Magisteruppsats, Göteborgs universitet. <https://gupea.ub.gu.se/bitstream/handle/2077/18942/gupea_2077_18942_1.pdf?sequence=1&isAllowed=y> (januari 2025).
- SAOL 2015 = *Svenska Akademiens ordlista över svenska språket*. Fjortonde upplagan. Stockholm: Svenska Akademien.
- SAOLhist = Tillgänglig: <<http://spraakdata.gu.se/saolhist/>> (augusti 2025).
- Sköldberg, Emma & Anna Helga Hannesdóttir (2017): Svenska ord – men vilka? Om uppslagsorden i Svensk ordbok utgiven av Svenska Akademien. I: Emma Sköldberg m.fl. (red.): *Svenskans beskrivning* 35. Göteborgsstudier i nordisk språkvetenskap 29. Göteborg: Göteborgs universitet. 329–340.
- Sköldberg, Emma (2022): Andra upplagan av Svensk ordbok: förutsättningar och redaktionella val. I: *LexicoNordica* 29, 139–152.

- SO 2009 = *Svensk ordbok utgiven av Svenska Akademien*. Stockholm: Norstedts/Svenska Akademien.
- Svanlund, Jan (2009): *Lexikal etablering: En korpusundersökning av hur nya sammansättningar konventionaliseras och får sin betydelse*. Stockholm Studies in Scandinavian Philology New Series 52. Stockholm: Stockholms universitet, institutionen för nordiska språk.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. 2:a uppl. Stockholm: Norstedts.
- Svenska Akademien (2024): *Om ordböckerna*. <<https://svenska.se/om/om-ordbockerna/>> (december 2024).
- Svenska språknämnden (1986): *Nyord i svenskan från 40-tal till 80-tal*. Solna: Esselte studium.
- Svensson, Anders (2017): Nyordens flippar och floppar. I: *Språktidningen* 3/2017. <<https://spraktidningen.se/artiklar/nyordens-flippar-och-floppar/>> (januari 2025).
- Svensson, Anders (2020): Svenskan växer ord för ord. I: *Språktidningen* 4/2020. <<https://spraktidningen.se/artiklar/svenskan-vaxer-ord-ord/>> (januari 2025).
- Wikström, Sanna (2017): *Nyorden som speglar nutiden: En korpusundersökning av hur 37 nyord har etablerat sig i svenskt skriftspråk*. Kandidatuppsats, Umeå universitet. <<https://www.diva-portal.org/smash/get/diva2:1111881/FULLTEXT01.pdf>> (januari 2025).

Andreas Brimberg
masterstudent, fil.mag.
Uppsala universitet

Linnea Hanell (korresponderande
författare)
språkvårdare, fil.dr
Språkrådet, Institutet för språk och
folkminnen
Box 20057
SE-104 60 Stockholm
linnea.hanell@isof.se

Den universelle First-Letter Law og dens anvendelse til statistisk undersøgelse af skandinaviske ordbøger

Poul Hansen

The article demonstrates how the mathematically based method First-Letter Law can utilize the statistical properties of lemmalists to analyse specific issues in Nordic lexicography. This offers a significant advantage compared to traditional approaches.

The practical implementation on three Scandinavian dictionaries and selected sub-materials (Swedish prefixed verbs) includes: 1) a clear model-based description of the frequency and relationships of initial letters in a dictionary, 2) calculation of the correlation between two dictionary materials, and 3) statistical documentation of diachronic changes in dictionaries.

The article reviews various aspects of the method and its limitations. The reader also gets practical instructions for applying the First-Letter Law in Excel.

1. Indledning

Enhver leksikograf er bekendt med, at fordelingen af begyndelsesbogstaver i en lemmaliste er ganske ujævn. Det er også velkendt, at det handler om et konsekvent mønster, der ikke adskiller sig meget fra ordbog til ordbog. Vi ved i dag, at mønstret kan beskrives tilfredsstillende med en eksponentiel funktion. Et lignende fænomen ses også hos cifre, og her udviklede Newcomb (1881) og senere Benford (1938) tidligt en matematisk funktion, der beskriver fordelingen af førstecifre. I litteraturen kaldes denne funktion Benfords lov, Newcomb-Benfords lov eller BL. Med hensyn til tekst skete modeludviklingen lidt senere og over en længere periode, men i dag er der en række modeller, som beskriver Ben-

ford-lignende tendenser og andre fænomener inden for korpuslingvistikken, herunder Zipfs lov (Zipf 1936), Heaps lov (Herdan 1960), Menzeraths lov (Altman 1980) og First-Letter Law (Xiao-yong et al. 2017). Nærværende artikel er baseret på First-Letter Law, som er en videreudvikling af Benfords lov. Den betegnes FLL-loven eller FLL her i artiklen.

Benfords lov er et statistisk værktøj af stor almen interesse for forskere. I litteraturen finder man eksempler på mange interessante og praktiske anvendelser inden for astronomi, biologi, datavidenskab og finansiel analyse. Der knytter sig særlig stor interesse til, at loven kan bruges til at kontrollere datakvalitet. Den er især velegnet til at afsløre afvigende eller manipulerede materialer, fx AI-genererede billeder (Bonettini et al. 2020). Den videreudviklede FLL for tekst har også disse kvaliteter. Et eksempel er en undersøgelse, hvor FLL benyttes til at afsløre forekomsten af falske indlæg på Donald Trumps Twitter-konto (Guillen, Yacone & Dai 2020).

Inden for leksikografi er der tilsyneladende ingen udbredt anvendelse af Benfords lov eller FLL til undersøgelse af ordbogsindhold. Få resultater er blevet publiceret i den forbindelse. I et arbejde af Shulzinger & Bormashenko (2017) har en Benford-lignende model været brugt til almene ordbøger på 10 sprog. Så vidt vides, foreligger der på nuværende tidspunkt endnu ingen publicerede resultater, hvad angår de nordiske sprog. Derfor er det måske ikke helt forkert at hævde, at vi i Norden endnu ikke har opdaget metodens deskriptive og frem for alt analytiske potentiale.

Denne artikel forsøger at råde bod på dette ved at undersøge metodens mulige anvendelser i nordisk leksikografi, dels som en overordnet deskriptiv model af en ordbogs samlede lemmabestand, dels som en analytisk metode til at teste og sammenligne udvalgte dele af en ordbogs indhold.

2. Målsætning

Følgende hovedspørgsmål danner udgangspunkt for arbejdet:

- Er FLL i stand til at forudsige begyndelsesbogstavernes kvantitative fordeling i tre større almene skandinaviske ordbøger, og hvordan forholder de tre ordbøgers FLL-modeller sig til hinanden?
- Har FLL praktisk anvendelse ved analyse af diakrone forandringer i ordbøger?

3. Data

3.1. Materiale 1

Ordbøger	Udgivelsesår	Sprog	Estimeret antal opslagsord
<i>Nudansk Ordbog</i>	1987	Dansk	50.000
<i>Riksmålsordboken</i>	1977	Norsk	50.000
SAOL	1982	Svensk	115.000

Tabel 1: De tre almene ordbøger, som indgik i undersøgelsen.

3.2. Materiale 2

Ordbøger	Sprog	Antal præfiksverber med <i>för</i> - og estimeret antal opslagsord i parentes
P. O. Welander: <i>Svensk-dansk-norsk lommeordbog</i> (1846)	svensk > dansk/norsk	380 (40.000)
K. Kristensen m.fl.: <i>Svensk-Dansk Ordbog</i> (2010)	svensk > dansk	324 (50.000)

Tabel 2: De to ordbøger, der blev brugt til diakron analyse.

4. Metode

4.1. FFL-loven udtrykt i generelle matematiske termer

Den generelle formel for FLL er følgende:

$$p_i = \frac{X - (X - 1)\log_x(X - 1) - i\log_x i + (i - 1)\log_x(i - 1)}{X(X - 1)\log_x\left(\frac{X}{X - 1}\right)} \quad [1]$$

p_i er FLL-værdien for et begyndelsesbogstav med hyppigheden i .

X er antallet af bogstaver i alfabetet eller antallet af forskellige begyndelsesbogstaver i et undersøgt delmateriale.

$\log_x$ betegner logaritmen med basen X .

4.2. Praktiske beregninger af FLL-værdier i MS Excel

Hvis man benytter MS Excel, kan FLL-værdierne beregnes for et alfabet med 28 bogstaver som følger:

For $i = 1$ (dvs. det begyndelsesbogstav, som forekommer hyppigst i materialet) beregnes FLL-værdien med:

$$p_1 = (1 + 27 * \text{LOG}(28/27; 28)) / (28 * 27 * \text{LOG}(28/27; 28)) \quad [2]$$

For $i = 2$ (dvs. det næsthypigste begyndelsesbogstav i materialet) beregnes FLL-værdien med denne formel:

$$p_2 = (28 - 27 * \text{LOG}(27; 28) - 2 * \text{LOG}(2; 28) + 1 * \text{LOG}(1; 28)) / (28 * 27 * \text{LOG}(28/27; 28)) \quad [3]$$

Sidstnævnte formel [3] kan derefter benyttes til alle de øvrige begyndelsesbogstaver med faldende hyppighed. For fx at få FLL-værdien for det sjette-hyppigste begyndelsesbogstav udskifter man de to 2-taller med 6 og de to 1-taller med 5.

Formel [2] og [3] kan kopieres direkte ind i Excel. For at få en tilpasning til frekvenserne i det aktuelle materiale multipliceres FLL-værdierne med det totale antal ord.

4.3. Tilpasninger af FLL-beregningerne til de aktuelle materialer

FLL-værdierne blev beregnet for 29 bogstaver for den danske og norske ordbog, fordi de indeholdt ord med begyndelsesbogstavet W, og for 28 begyndelsesbogstaver for den svenske ordbog, der ikke havde opslagsord, som begyndte med W.

FLL-værdierne for præfiksverberne med *för*- blev bestemt ud fra det første bogstav i andetleddet og blev beregnet for 23 bogstaver i materiale 2. Præfikser, der begynder med *för*- (fx *förut*-, *före*-, *förbi*-), er ikke medtaget i materialet.

4.4. Optælling og estimering af antal opslagsord

De tre ordbøgers lemmabestand blev estimeret ved at optælle antallet af linjer per begyndelsesbogstav i ordbøgerne, og den gennemsnitlige forekomst af tydeligt fremhævede opslagsord per linje blev beregnet ud fra stikprøver. Det totale antal opslagsord per begyndelsesbogstav blev derefter beregnet ved at multiplicere det optalte antal linjer med det gennemsnitlige antal opslagsord per linje. Ordbogens totale antal opslagsord blev beregnet ved at summere alle begyndelsesbogstavers antal.

Præfiksverberne i *Svensk-dansk-norsk lommeordbog* og *Svensk-Dansk Ordbog* blev optalt manuelt.

4.5. Præsentationen af resultaterne

Resultaterne præsenteres i todimensionelle kurvediagrammer, hvor to linjer sammenlignes. Den ene linje repræsenterer den beregnede FLL-linje (FLL-modellen), og den anden linje repræsenterer ordantallet for den pågældende ordbog. Det kan også være to FLL-modeller, som sammenlignes, fx ved sammenligning af to ordbøger. På x-aksen angives begyndelsesbogstaverne i den rækkefølge, som bestemmes af ordantallet, i retningen fra højest til lavest. Den beregnede R^2 -værdi (forklaringsgraden) angives også.

4.6. Beregninger i Excel

Hvis en ordbog skal sammenlignes med dens FLL-kurve, opretter man først to kolonner: en kolonne med begyndelsesbogstaverne og en kolonne med det antal opslagsord, der har de pågældende begyndelsesbogstaver. Derefter sorteres de to kolonner efter ordantallet, fra højest til lavest. I en tredje kolonne lægger man derefter FLL-værdierne, som beregnes ifølge formel [2] og [3]. Disse FLL-værdier multipliceres nu med det samlede antal opslagsord for ordbogen. Cellerne i de tre kolonner markeres, og man tegner et diagram ved at vælge Excels funktion for todimensionelle diagrammer i menuen.

Diagrammer for sammenligning af to ordbøgers FLL-modeller kræver lidt mere håndtering af kolonner, men princippet er det samme. Man gennemfører ovenstående for begge ordbøger, men inden multiplikation af FLL-værdierne sorteres begge ordbøger alfabetisk hver for sig. Derefter markeres alle seks kolonner og sorteres efter FLL-værdierne for den ordbog, der skal udgøre udgangsortbogen (den ordbog, som den anden ordbog testes mod). Det næste trin er at multiplicere begge kolonner med FLL-værdier med det ordantal, som udgangsortbogen har. Nu kan man tegne et diagram, hvor de to ordbøgers FLL-modeller sammenlignes

med hinanden. Til dette benyttes kolonnen med begyndelsesbogstaverne for udgangsordbogen og de to kolonner med FLL-værdier.

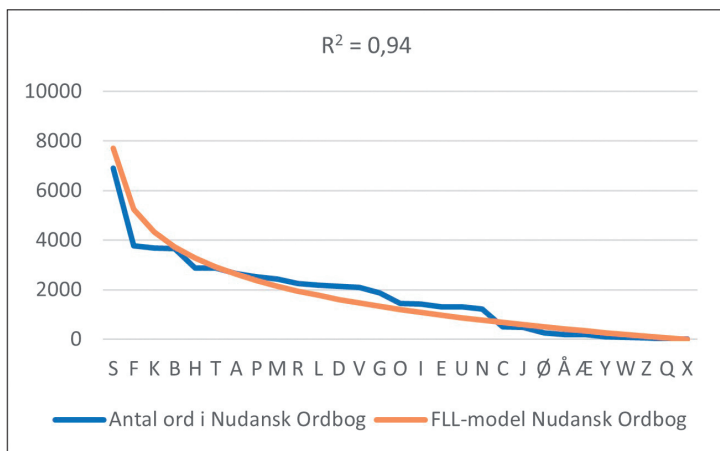
4.7. Beregninger af forklaringsgrad i Excel

Afvigelser i forhold til FLL-modellen i diagrammerne beregnes med forklaringsgraden R^2 , også kaldet variationskoefficienten. I MS Excel benyttes funktionen Forklaringsgrad. R^2 -værdien varierer mellem 0 og 1, hvor 1 svarer til 100 % korrelation. I nærværende artikel benyttes %-angivelsen.

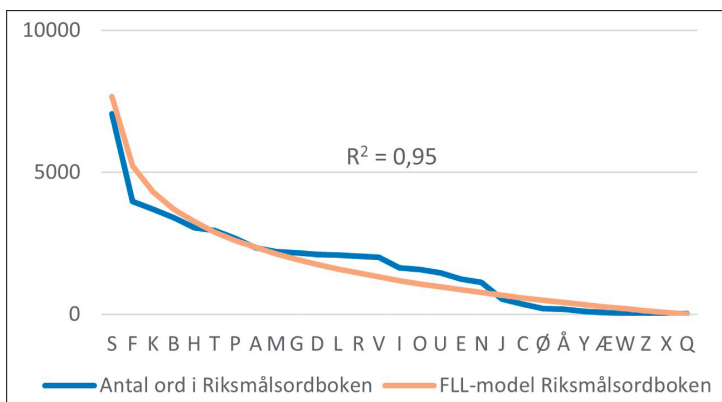
5. Resultater

5.1. *Nudansk Ordbog*, *Riksmålsordboken* og SAOL

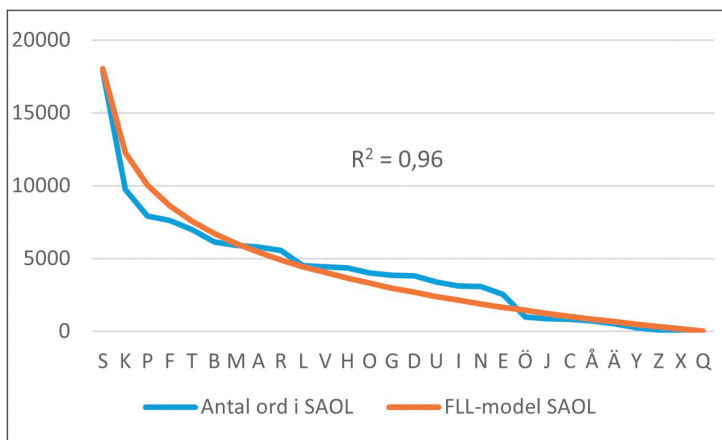
De beregnede FLL-modeller for de tre ordbøger (figur 1-3) forklarer en meget stor del af begyndelsesbogstavernes frekvenser med forklaringsgrader på 94-96 %. Der er tale om næsten identiske re-



Figur 1: *Nudansk Ordbog* og dens FLL-model.

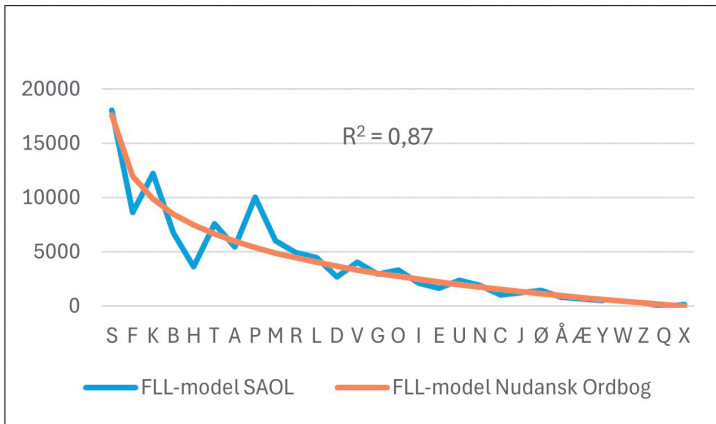


Figur 2: Riksmålsordboken og dens FLL-model.

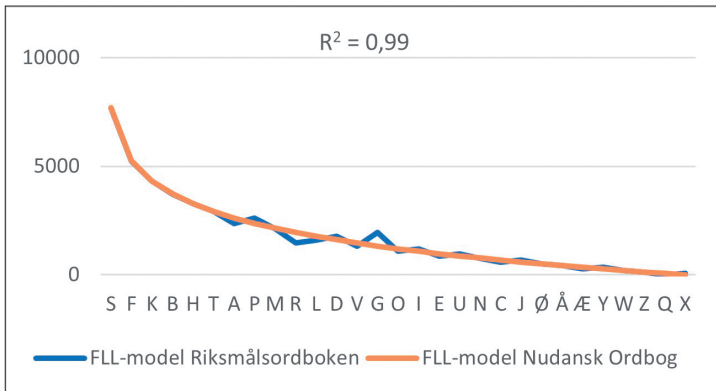


Figur 3: SAOL og dens FLL-model.

sultater, når man sammenligner diagrammerne. Selv de mindre afvigelser fra modellen mellem S og H (afvigelse under linjen) og mellem T og C (afvigelse over linjen) har det samme mønster i de tre diagrammer. Den store forskel i ordantal mellem SAOL (115.000) og de andre to ordbøger (50.000) påvirker ikke forklaringsgraden signifikant.



Figur 4: Sammenligning af FLL-modellen for SAOL med FLL-modellen for *Nudansk Ordbog*.



Figur 5: Sammenligning af FLL-modellen for *Riksmålsordbogen* med FLL-modellen for *Nudansk Ordbog*.

Det betyder dog ikke, at de tre ordbøger har eksakt samme struktur. Ved sammenligning mellem *Nudansk Ordbog* og SAOL (figur 4) fremkommer der signifikante forskelle samt en lavere forklaringsgrad (87 %). Det er indikeret, at nogle af de største forskelle ses ved bogstav *H* og *P*. Umiddelbart tænker man på, at *hv-* i dansk svarer til *v-* på svensk (fx *hvid/vit*). Derimod er det

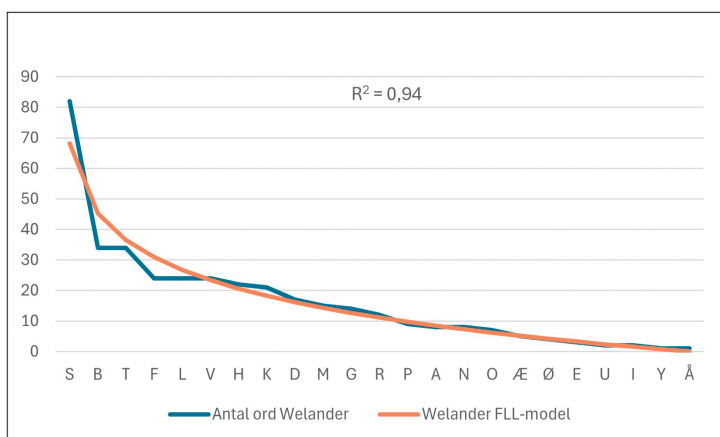
vanskeligere at forklare forskellen, når det gælder *P*. Det kan være interessant at notere, at det faktisk er muligt at sætte et tal (R^2 -værdien) på den relative lighed mellem dansk og svensk ordforråd og ortografi ved at sammenligne ordbøgernes lemmafordeling.

Mellem den danske og norske ordbog (figur 5) er der betydeligt mindre forskelle end ved sammenligning mellem den danske og svenske ordbog. Forklaringsgraden er på hele 99 %.

5.2. FLL-modellen anvendt til dokumentation af diakrone forandringer

I figur 6 vises antal præfiksverber med *för-* i *Svensk-dansk-norsk lommeordbog*. Den beregnede FLL-model dækker en stor del af variationen mellem begyndelsesbogstaverne med en forklaringsgrad på hele 94 %. Det er helt på niveau med nutidens ordbøger og indikerer, at Welander har udført et genuint indsamlingsarbejde i forbindelse med ordbogen.

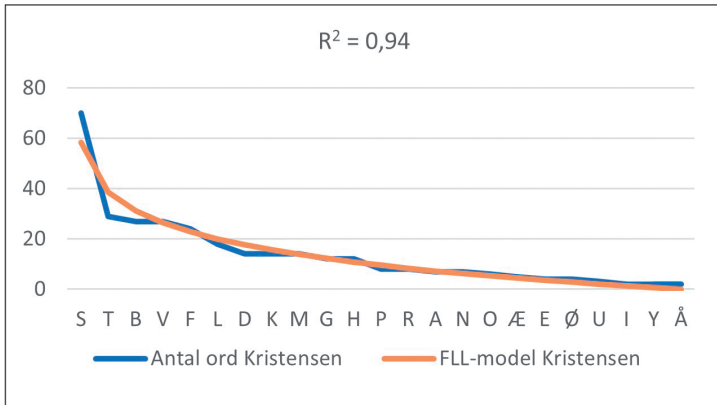
Resultatet er meget lig resultatet for den meget senere udgivne *Svensk-Dansk Ordbog* (figur 7), hvor man også ser en forklarings-



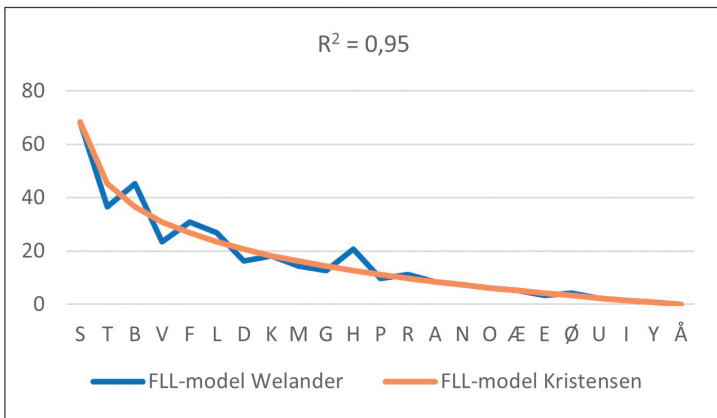
Figur 6: Antal svenske verber med præfikset *för-* i *Svensk-dansk-norsk lommeordbog* og den beregnede FLL-model.

grad på 94 %. I begge tilfælde er der kun forholdsvis små afvigelser fra FLL-modellen.

Ved sammenligningen mellem ordbøgernes FLL-modeller (figur 8) fremkommer der lidt flere og lidt mere markante afvigelser. Forklaringsgraden er dog stadig betragtelig høj, men visuelt er der større afvigelse mellem kurverne end i figur 6 og 7. Kigger



Figur 7: Antal svenske præfiksverber med *för-* i *Svensk-Dansk Ordbog* og den beregnede FLL-model.



Figur 8: Sammenligning af FLL-modellen for *Svensk-dansk-norsk lommeordbog* med FLL-modellen for *Svensk-Dansk Ordbog* med hensyn til svenske præfiksverber med *för-*.

man nærmere i de to ordbøger på de begyndelsesbogstaver, der har størst afvigelser, fx *H*, kan man se, at der her er et relativt stort indslag af forældede ord i *Svensk-dansk-norsk lommeordbog* (*förhyda*, *förhuttla*, *förhyra* og *förhåda*), der ikke er medtaget i *Svensk-Dansk Ordbog*. Denne omstændighed indikerer, at afvigelserne i diagrammet afspejler sprogforandringer i væsentlig grad.

6. Diskussion

6.1. FLL's rolle i leksikografi

Sammenfattende er det vigtigste, denne artikel viser vedrørende FLL's rolle i leksikografi, at danske, norske og svenske lemmalister har tydelige statistiske og testbare egenskaber. Det er også forbavsende ligetil at kontrollere disse egenskaber. Man behøver blot en matematisk formel og det samlede antal opslagsord. Beregninger kan dertil udføres med ringe arbejdsindsats, hvis man behersker regnearksprogrammer.

For nordiske leksikografer bør FLL være særlig interessant, fordi metoden kan give nye indfaldsvinkler i studiet af de skandinaviske nabosprogs parallelle og divergerende udvikling i ordbøger. Test af beskaffenhed og dækningsgrad kan også have betydning ved udvikling af nye ordbøger.

6.2. Der er store ligheder med andre sprog

I Shulzinger & Bormashenko (2017) har man beregnet den gennemsnitlige forklaringsgrad mellem begyndelsesbogstavers frekvens og beregnede modelværdier i ordbøger på 10 sprog (engelsk, tysk, lettisk, fransk, italiensk, spansk, hebræisk, russisk, polsk og tjekkisk) til 0,91 (91 %). Til sammenligning har de tre

her undersøgte ordbøger (dansk, norsk og svensk) et gennemsnit på 95 %, altså lidt over gennemsnittet for de 10 ovennævnte sprog. Men forskellen (4 %) er for lille til at drage nogen sikre konklusioner. I de publicerede diagrammer for tysk, russisk og fransk i samme artikel er der også en tydelig overensstemmelse med de tre skandinaviske ordbøger.

6.3. Hvordan skal man tolke ensartede afvigelsesmønstre i forhold til FLL-kurven?

Det er ikke blot de generelle kurveforløb, der har store ligheder mellem sprogene. Afvigelsesmønstrene i forhold til FLL-modelens kurvelinje har også store ligheder, ikke bare mellem dansk, norsk og svensk, men også i relation til en række andre sprog. Det indikerer, at der mangler en universal komponent i modellen på tværs af sprogene. Her er der behov for bedre forståelse for, hvad der egentlig styrer den kvantitative fordeling af begyndelsesbogstaver og årsagen til, at FLL- og Benford-modeller er så effektive til at forudsige den. I mangt og meget opleves loven som et paradoks (Fellman 2014).

6.4. Viser FLL egentlig noget nyt?

En skeptiker kan måske hævde, at FLL viser det, man allerede ved. Men hertil kan man svare, at metoden først og fremmest tilfører overskuelighed, det vil sige, at man får et overblik over den samlede lemmabestand, og at man har en model at sammenligne med. Man kan også på en praktisk måde sætte tal på det, som man normalt kun vurderer med beskrivende ord. At arbejde modelbaseret med FLL fører desuden leksikografien nærmere de videnskabsgrene, hvor samspillet mellem teori og faktiske observationer spiller en stor rolle.

6.5. Top down-tilgang med FLL

En fordel med FLL er, at den giver mulighed for at arbejde ud fra en top down-tilgang, som indebærer en struktureret og systematisk analyse af et materiale, hvor man starter med det mest generelle og gradvis zoomer ind på det mere detaljerede. Et eksempel på dette er, at man forudsætningsløst starter med at se på de største afvigelser mellem to kurver og derefter går ned på detaljeniveau og ser, hvilke lemmaer det handler om. Det kan fx ske, at forklaringen er forekomst af forældede ord eller ortografiske forskelle. Brugen af FLL-modeller bidrager til, at man umiddelbart kan identificere de mest interessante afvigelser mellem ordbøger. Samtidig er top down-princippet med til at forhindre, at man overser noget vigtigt, fx i forhold til kun at basere analysen på nogle få udvalgte typeeksempler.

6.6. FLL er ikke begrænset til det første bogstav i et ord

Da denne artikel har brugt FLL til at analysere præfiksverber i ordbøger, har den også – på en måde – testet muligheden for at bruge FLL til at undersøge sammensatte og afledte ord. Der knytter sig måske en særlig interesse til dette i Norden, fordi man i de skandinaviske sprog jo frit kan danne nye sammensatte ord i modsætning til fx engelsk. Man skal blot være opmærksom på, at der medtages lemmaer nok til, at det bliver forsvarligt at gennemføre analysen, det vil sige, at der bliver tilstrækkeligt med data til at fastsætte en rangordning af begyndelsesbogstaverne. Ved for få data kan der opstå problemer med rangordning af de lavfrekvente begyndelsesbogstaver. Antallet af lemmaer kan dog være forholdsvis lavt. I denne artikel, hvor FLL-undersøgelsen vedrørende præfiksverber var baseret på andetleddets første bogstav, blev der opnået tilfredsstillende og stabile resultater, selvom det handlede om mindre end 1 % af det samlede antal opslagsord.

6.7. Ordbøger med divergerende antal opslagsord

De tre almene ordbøger, som blev undersøgt og sammenlignet, havde ikke det samme antal opslagsord. SAOL havde to gange så mange som hver af de to andre ordbøger. FLL-metoden er dog meget robust. Det skyldes netop, at metoden er modelbaseret. Det er ordbøgernes mønstre med hensyn til begyndelsesbogstavernes indbyrdes kvantitative relationer, der analyseres. De beregninger, der foretages, er på helt sammenlignelige værdier.

6.8. Begrænsninger i metoden

En udfordring med FLL-metoden er, at den kun arbejder med antallet af ord uden hensyn til, hvilke ord det drejer sig om. Hvis man sammenligner en ældre og nyere ordbog, der har samme antal opslagsord for et begyndelsesbogstav, indeholder de måske ikke helt de samme ord. I den ene ordbog kan der være tale om indslag af forældede ord, som ikke findes i den nye ordbog. I stedet har den nye ordbog måske fået tilføjet nye, men helt andre ord, der antalsmæssigt svarer til de manglende forældede ord. I sådanne tilfælde vil brug af FLL-modellen derfor underestimere de reelle forskelle.

Det kan også ske, at FLL-metoden overestimerer forskelle, hvis fx et ord har Q som begyndelsesbogstav i en ældre ordbog, men K i en nyere ordbog. Man er altså nødt til at sammenligne ordbøger mere detaljeret, inden man drager mere definitive konklusioner.

Dette arbejde har i lighed med tidligere arbejder med FLL-modellen, fx Xiaoyong et al. (2017), benyttet R^2 -værdien til at beskrive relationen mellem kurver. Det bør dog overvejes at finde en bedre måde at beskrive relationen mellem kurver. I figur 8 er der fx indikeret en større afvigelse visuelt mellem kurverne end i figur 6 og 7. R^2 -værdien afspejler ikke dette forhold.

Ved sammenligning mellem ordbøger, hvor interessen ikke er lagt på diakrone forandringer, bør der ikke være for stort tidsrum

mellem udgivelserne. I dette arbejde sættes en grænse på højst 10 år for sammenligning mellem nordiske ordbøger.

Med hensyn til sammenligningen af de valgte nordiske ordbøger kan der gøres den reservation, at sammenligningen er problematisk, fordi SAOL er en retskrivningsordbog og de to andre nordiske ordbøger er betydningsordbøger. I hvilken udstrækning det påvirker rangordningen mellem begyndelsesbogstaver er ikke blevet klarlagt i denne undersøgelse. Derimod er forskellene i lemmaantallet sandsynligvis af mindre betydning.

Foreliggende undersøgelse har først og fremmest været en pilotundersøgelse for at se, om metoden overhovedet fungerer. Der kræves mere målrettede oplæg, hvis man skal se nærmere på, i hvilken udstrækning rangordningen af begyndelsesbogstaver kan identificere ufuldstændigheder i en ordbogs lemmabestand, og hvad der i praksis kan udledes af forskelle.

7. Konklusion

Artiklen viser, at der er særdeles god grund til at interessere sig for FLL i nordisk leksikografi. Metoden tilbyder nogle klare fordele i forhold til traditionelt leksikografisk analysearbejde.

FLL kan forudsige begyndelsesbogstavets indbyrdes relation i en ordbog, både overordnet og i sammensatte ord, men kan ikke forudsige, hvilke bogstaver det drejer sig om. Men trods denne begrænsning bør begyndelsesbogstavets rangordning betragtes som en værdifuld vidensresurce, bl.a. fordi det giver mulighed for komparative analyser. Desuden kan metoden videreudvikles, især med hensyn til korrelationsberegningen. Eftersom det handler om rangordnede serier, vil ikke-parametriske metoder formodentlig kunne finde anvendelse, fx Kendalls tau eller Spearmans rangkorrelation.

Litteratur

Ordbøger, korpusser og digitale resurser

- Nudansk Ordbog* (1987). Christian Becker-Christensen, Lars Heltoft, Carol Henriksen & Peter Widell (red.). 13. udgave, 3. oplag. København: Politikens Forlag.
- Riksmålsordboken* (1977). Tor Guttu, Kåre Skadberg & Inge Wettergreen-Jensen (red.). Oslo: Kunnskapsforlaget.
- SAOL (1982) = *Svenska Akademiens ordlista över svenska språket*. 10 uppl. Stockholm: Norstedts.
- Svensk-dansk-norsk lommeordbog* (1846). Peter Olof Welander. Kjöbenhavn: G.H. Jægers skandinaviske Forlagshandel.
- Svensk-Dansk Ordbog* (2010). Kjeld Kristensen, Else Bojsen & Pernille Folkmann (red.). København: Det Danske Sprog- og Litteraturselskab & JP/Politikens Forlagshus.

Anden litteratur

- Altmann, Gabriel (1980): Prolegomena to Menzerath's law. I: *Glottometrika* 2, 1-10.
- Benford, Frank (1938): The law of anomalous numbers. I: *Proceedings of American Philosophical Society* 78, 551-572.
- Bonettini, Nicolò, Paolo Bestagini, Simone Milani & Stefano Tubaro (2020): On the use of Benford's law to detect GAN-generated images. I: *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE. 5495-5502. <<https://doi.org/10.1109/ICPR48806.2021.9412944>>.
- Fellman, Johan (2014): The Benford paradox. I: *Journal of Statistical and Econometric Methods* 3(4), 1-20. <https://www.scienpress.com/Upload/JSEM/Vol%203_4_1.pdf>.
- Guillen, Gabriel, Christopher Yacone & Zhuoran Dai (2020): Benford's Law Applied to Twitter Data. Working paper, Harvard

- University. <https://scholar.harvard.edu/files/gabrielguillen/files/final_paper.pdf> (februar 2025).
- Herdan, Gustav (1960): *Type-token Mathematics*. The Hague: Mouton.
- Newcomb, Simon (1881): Note on the frequency of use of the different digits in natural numbers. I: *American Journal of Mathematics* 4, 39-40.
- Shulzinger, Evgeny & Edward Bormashenko (2017): On the Universal Quantitative Pattern of the Distribution of Initial Characters in General Dictionaries: The Exponential Distribution is Valid for Various Languages. I: *Journal of Quantitative Linguistics* 24(4), 273-288. <<https://doi.org/10.1080/09296174.2017.1304620>>.
- Xiaoyong, Yan, Seong-Gyu Yang, Beom Jun Kim & Petter Minnhagen (2017): Benford's Law and First Letter of Words. I: *Physica A: Statistical Mechanics and its Applications* 512, 15. 305-315. <<https://doi.org/10.1016/j.physa.2018.08.133>>.
- Zipf, George Kingsley (1936): *The Psycho-Biology of Language. An Introduction to Dynamic Philology*. Oxford: Routledge.

Poul Hansen
Translatør, fil dr
Box 4158
SE-227 22 Lund
poul@danska-svenska.se

BOKMELDINGAR

Anmeldelse af *Retskrivningsordbogen*, 5. udgave

Holger Juul

Dansk Sprognævn: *Retskrivningsordbogen*. 5. udgave. Alinea, 2024.
1177 sider. Udgivelsesdato: 18/11 2024. Digital version: ro.dsn.dk.

1. Indledning

Retskrivningsordbogen (RO) er jo der man slår op for at få at vide hvordan ord staves. Så når der kommer en ny udgave af ordbogen, aktuelt 5. udgave (RO5), er det naturligt at spørge om der er ord der skal staves på en ny måde. Det vil jeg også gøre her – men først vil jeg opholde mig ved at RO ikke blot er en autoritet med hensyn til ortografi i snæver forstand, men også på flere andre områder.

Ordbogen dækker “rigssprogets almindelige ordforråd”¹, så det nye ved nye udgaver af RO er ikke mindst de nye ord som redaktionen har valgt at tage med. Ordbogens autoritet på dette område fremgår af mediernes omtale af nytilkomne ord, hvor det ikke sjældent bliver til at disse ord nu er blevet “optaget i det danske sprog”. RO bliver dermed indirekte normsættende som en liste over officielt anerkendte ord.

Ordbogen har også en betydelig autoritet når det gælder ord-dannelsen. Den giver oplysninger om ordenes form ved sammen-sætning og bøjning, hvilket klart nok er nødvendigt hvis brugerne skal kunne vide hvordan sammensatte og bøjede ord skal staves. Herved bliver RO indirekte normsættende også med hensyn til hvilke sammensætningsformer og bøjninger der overhovedet er “tilladt”.

1 <https://ro.dsn.dk/mere-om-retskrivningsordbogen/vejledning-til-indhold> (og den trykte ordbog s. 9).

Endvidere giver ordbogen regler for tegnsætning, som heller ikke hører til ortografien i snæver forstand. Her er kommareglerne blevet en slags føljeton, hvor hver ny udgave af RO ledsages af en ny runde offentlig debat om kommateringssystemer. RO5 er ingen undtagelse.

I det følgende vil jeg fremsætte nogle iagttagelser og vurderinger af RO5 der relaterer til de ovennævnte emner. Undervejs vil jeg i særlig grad interessere mig for hvordan ordbogen fungerer for urutinerede sprogbrugere såsom skoleelever. Allerførst vil jeg dog inspicere ordbogens fremtrædelsesformer, som omfatter både en digital udgave og en trykt.

2. Digital versus trykt ordbog

Som noget nyt er den primære udgave af RO nu den digitale udgave (ro.dsn.dk). Det vil sige at netversionen af ordbogen, inklusive løbende opdateringer, er den til enhver tid gældende.² Det er praktisk sådan, men så vidt jeg kan se, er det ikke en oplysning der flages med i den trykte ordbog, og ændringen er jo heller ikke just et argument for at købe et eksemplar til reolen. Et andet forhold der taler til fordel for den digitale udgave, er at den trykte ordbog er nået op på en vægt omkring 1,6 kg og en tykkelse på næsten 5 cm. Den er simpelthen svær at håndtere for brugere med almindelig fysik, og mange mindreårige elever vil stå helt af alle-rede her – men de vil nok under alle omstændigheder foretrække den digitale ordbog.

Layoutet i den trykte udgave er omtrent som i fjerde udgave (2012), men en forbedring er at alfabetet nu fremgår som et lodret indeks i højresiders højre margin med det aktuelle bogstav frem-

2 Dette fremgår her: <https://ro.dsn.dk/mere-om-retskrivningsordbogen/uddybende-laesning-om-ro-5/den-digitale-udgave-af-retskrivnings-ordbogen> – men ikke i forordet til den trykte ordbog.

hævet af en sort baggrund der går helt til kant og ses når bogen er lukket. På den måde kan man regne ud hvor man skal slå op for at finde et bestemt bogstav, og man får et indtryk af hvor meget de enkelte bogstaver fylder – med S-sektionen som den suverænt største.

limstift – Lissabon		
<i>med forskellige modersmål; historisk blandingssprog</i>	linolleums snit <i>sb.</i> , -tet,	A
lingvist <i>sb.</i> , -en, -er, -erne	linoleumssnit, -tene	B
<i>(person der beskæftiger sig med lingvistik)</i>	(linoleumstryk)	
lingvilstik <i>sb.</i> , -ken (<i>sprog- videnskab</i>)	linolleums tryk <i>sb.</i> , -ket,	C
	linoleumstryk, -kene	D
	(<i>en trykmetode; billede fremstillet ved linoleumstryk</i>)	
lingvilstisk <i>adj.</i> , <i>itk. d.s.</i> , -e	linolie <i>sb.</i> , -n	E
<i>(som vedrører lingvistik)</i>	linolsyre <i>sb.</i> , -n	
linilment <i>sb.</i> , -et, -er, -erne	linse <i>sb.</i> , -n, -r, -rne (<i>optik et skiveformet glas- el. plastic- legeme; et frø; en kage</i>), i	F
<i>(salve)</i>	<i>sms. linse-, fx linseform,</i>	G
linje <i>sb.</i> , -n, -r, -rne (<i>fork.: 1.); trække en linje; holde den slanke linje; tage linje 2 til stationen; knas på linjen</i>)	linsesuppe	H
linjeldommer <i>sb.</i> , -en, -e, -ne	linselus <i>sb.</i> , -en, linselus, -ene (<i>UFORMELT person som er overdrevent glad for at komme i medierne</i>)	I
linjefag <i>sb.</i> , -et, linjefag, -ene, i <i>sms. linjefags-, fx linjefagspraktik</i>	lipase <i>sb.</i> , -n, -r, -rne (<i>et enzym</i>)	J
linjeføring <i>sb.</i> , -en, -er, -erne; <i>motorvejens linje- føring</i>	lipgloss <i>sb.</i> , -en, lipglosser el. lipgloss, -erne	K
linjefilcer <i>sb.</i> , -en, -er, -erne	lipid <i>sb.</i> , -et, -er, -erne (<i>et organisk stof</i>)	L
	lipizainer <i>sb.</i> , -en, -e, -ne (<i>en hest</i>)	M
		N

Figur 1: Opslag i RO5, med alfabetet som et lodret indeks i højre side.

Den digitale udgave har fået valgmuligheder, sådan at man kan vælge om opslagsordene skal præsenteres med danske eller latinske termer (fx *navneord*, *fælleskøn* henholdsvis *substantiv*, *utrum*), og disse kan igen vælges i fuld form eller med forkortelser (*no.*, *fk.* henholdsvis *sb.*, *utr.*). Det er rart at kunne vælge selv, og ikkelingvister behøver så ikke kæmpe med kryptiske forkortelser. I den trykte udgave er det dog, givetvis af pladshensyn, forkortelserne der regerer, og her skal sprogbrugerne stadig lære at oplysningen i *sms.* ikke har noget med sms-beskeder at gøre, men betyder 'i sammensætninger'.

Et irritationsmoment er at opslag i den digitale ordbog ikke straks fører til alle oplysninger om det ord man har søgt på. Vil man se bøjningsoplysninger og sammensætningseksempler, må man klikke på ordet. Hvis man i stedet slår op i *Den Danske Ordbog* (DDO), spares man for dette klik og får dermed en lidt bedre brugeroplevelse. Det er en detalje, men det bliver til mange klik for den der bruger ordbogen meget.

På plussiden tæller det at søgemulighederne fra den nu nedlagte hjemmeside og app RO+ er tilgængelige, sådan at man kan søge med jokertegn. Det er en vigtig funktion for sprogforskere, og der er også mange pædagogiske muligheder i at søge efter ord med bestemte strukturer. Eksempelvis kan man med søgestrengen *\$\$\$\$| finde ord der har fem konsonantbogstaver (\$) i træk i slutningen af et ord (angivet med den lodrette streg). Der er kun et eneste hit, nemlig substantivet *borsjtj*. Det får mig til at spørge om der overhovedet er andre ord der ender på konsonant plus *j*. Svaret får man med søgestrengen *\$j|, men det gør en forskel om man har valgt *Søg i opslagsord* eller *Søg i bøjningsformer og sammensætningseksempler*. Med den sidste indstilling får man ud over *borsjtj* og forkortelserne *dj* og *kj* også tre verber med: *pulje*, *rutsje* og *talje*. Disse verber ender nemlig på *-j* i imperativ – men dette fremgår ikke af opslagene der ikke medtager imperativformer.

Søgninger kan med andre ord finde bøjningsformer som ikke vises i ordbogen, hvilket nok vil kunne forvirre en del brugere. Det gælder også genitiver, sådan at en søgning på *uis| foruden *croquis* finder substantivet *etui* – med den ikke viste genitivform *etuis*. Derimod finder denne søgning ikke verbet *cruise* hvis autoriserede imperativform således ikke kan være “cruis”, som ellers er den form jeg udleder af retskrivningsreglernes § 30.3.

Den digitale ordbog kommer ikke med forslag til hvilket ord der kan være ment, hvis man søger på en stavemåde der ikke er med i ordbogen. Søgninger på fx “cacao” og “prast” ender simpelthen blindt. Her er man igen lidt bedre hjulpet af DDO, hvor

“cacao” fører direkte til *kakao* og oplysning om at *cacao* ikke er en officiel stavemåde, og hvor “prast” fører til en række forslag, som måske kan hjælpe den urutinerede staver. Det oplagte ordforslag *præst* kommer dog først som DDO’s forslag nummer seks og efter det mindre sandsynlige forslag *prust*. For den usikre staver kunne RO, og til dels også DDO, gøres mere imødekommende på dette område.

En vis hjælp får man dog i RO, også i den trykte udgave, i form af henvisninger til forvekslingsord. Som bemærket af Torben Christiansen (Christiansen 2025) er henvisningerne dog ikke konsekvente. Eksempelvis henvises der fra *hviske* til *viske*, men ikke omvendt. Klassiske forvekslingsord blandt skoleelever såsom *ret-rat* og *man-mand* har ikke henvisninger. Indtrykket er således at henvisningerne til forvekslingsord primært er møntet på mere avancerede sprogbrugere. I lighed med en anmelder af 4. udgave (Theilgaard 2013) må man dog undre sig over at der ikke henvises fra *dosere* og *dossere* til *docere*, men kun indbyrdes mellem *dosere* og *dossere*. En støtte til at finde den rigtige stavemåde af lavfrekvente ord som disse har man dog i de betydningsangivelser som de fleste af dem er forsynet med.

3. Hvilke ord er med?

Den nye udgaves større omfang skyldes primært at antallet af opslagsord er forøget: ca. 67.500 i den trykte ordbog mod ca. 64.000 i 4. udgave. Det kan sådan set undre. Skyldes det at “rigssprogets almindelige ordforråd” er vokset? Eller er det redaktionens afgrænsning af hvad der er almindeligt, der er ændret? Måske, men forklaringen er også af mere teknisk art, idet eksempelvis forkortelsen *adm.* nu er fordelt på tre opslag, hvor de tre læsninger (*administration*, *administrerende* og *administrativ*) tidligere stod samlet. Udvidelserne afspejler muligvis også at det trykte format har sat

en grænse som man nu begynder at se stort på. I hvert fald læser man på ordbogens hjemmeside:

“Hvis ordbogen på et tidspunkt bliver udelukkende digital, er der ikke længere nogen pladsbegrænsning at tage hensyn til, og derfor vil der til den tid i mindre omfang blive fjernet opslagsord fra ordbogen. Antallet af ord der er fjernet i den nyeste udgave af *Retskrivningsordbogen*, er dog meget begrænset.”³

På samme side understreges det at når forældede ord som *A-post* og *nåddler* er forsvundet fra ordbogen, så betyder det ikke “at sådanne ord ikke (længere) findes i det danske sprog”.⁴ Spørgsmålet er så om man med den ubegrænsede digitale plads vil tage *A-post* og *nåddler* og andre bedagede ord med igen? Det håber jeg ikke, for det er vel stadig en væsentlig opgave at afgrænse og samle de mest *almindelige* ord i dansk? Jo, det må det være. Når man læser en redegørelse i *Nyt fra Sprognævnet* om RO5’s inklusion af *afholdskvinde* og flere andre ord på *-kvinde* (Andersen 2025), lægges der i hvert fald stor vægt på de nye opslagsords forekomst i almensproget.

Det er altid tankevækkende at se hvilke ord der er kommet til, fordi listen med nye ord⁵ jo tegner en slags portræt af samfundets udvikling i de seneste år. Men jeg har ikke kunnet finde nogen redegørelse for hvorfor antallet af *udgåede* ord i RO5 ikke er større, endsige en liste med disse. Og selve ordbogen har, så vidt jeg kan se, ikke beskrivelser eller henvisninger der forklarer hvordan redaktionen konkret har afgrænset det almindelige ordforråd. I betragtning af den autoritet RO tillægges, synes jeg dette er lidt ærgerligt. Hvis ordbogen gjorde nærmere rede for de redak-

3 <https://ro.dsn.dk/hvilke-ord-er-med-i-retskrivningsordbogen>.

4 *A-post* er faktisk heller ikke forsvundet helt da ordet optræder i forklaringen på opslagsordet *prioritaire*.

5 <https://dsn.dk/wp-content/uploads/2024/11/Nye-opslagsord-2024.pdf>.

tionelle principper og processer den bygger på, kunne det styrke brugernes forståelse for både ordbogens indhold og for sprogets udvikling.

4. Bøjninger og sammensætninger

Redaktionen af RO er, jævnfør citatet i forrige afsnit, tydelige om at ordbogen ikke bestemmer hvilke ord der findes i dansk. Derimod lader det til at redaktionen har det fint med at bestemme hvordan de ord der er taget med, skal bøjes og sammensættes. Eksempelvis fortæller retskrivningsreglernes § 30 at “[m]an danner imperativ af verber der i infinitiv ender på tryksvagt *-e*, ved at fjerne *e’et*”. Og på ordbogens hjemmeside forklares det at *slips* nu altid skal have fuge-*e* i sammensætninger. “Formen *slipsklædt* er dermed ikke længere korrekt.”⁶

Set fra brugernes synspunkt kan det være rart at få klar besked, det er jo gerne det man vil have når man slår op i en ordbog. Men der er tydeligvis sket noget siden Poul Lindegård Hjorth (formand for Dansk Sprognævn 1966-1973) for små tre årtier siden skrev:

“Men offentlige påbud om i almensproget at holde visse ord ude [...] er utænkelige. Det kan heller ikke tænkes, at staten skulle give sig af med at opstille regler for, hvordan danske ord skal dannes eller bøjes.” (Hjorth 1997)

Det der forekom Hjorth at være utænkeligt, er tilsyneladende nu blevet hverdag. Personligt finder jeg at den sproglige frihed begrænses uden god grund når brug af den (i hvert fald for mig) fuldt acceptable form *slipsklædt* skal være en slags civil ulydighed. Hvad angår imperativer, er det nyttigt med den klare regel, som tilsiger at man skal skrive *Cykl dog hjem!* fremfor **Cykel dog hjem!*

6 <https://ro.dsn.dk/oversigt-over-indhold-og-aendringer>.

Den ukorrekte form med *-el* ville være en lige så god gengivelse af udtalen, og den ville desuden undgå den påfaldende endelse på *-kl*, men en retskrivningsordbogs opgave er nu engang at udstikke en norm. Når det som her kan ske med en generel regel, er det fint.

På de mange områder hvor der i talesproget er variation i valget af bøjningsformer, ville jeg derimod som Lindegård Hjorth foretrække at man undlod at normere og nøjedes med at oplyse og eventuelt anbefale de mest gængse former. Ligesom *nåddler* fremdeles er et dansk ord, er ikkeinkluderede bøjningsformer som *dragede* (datidsform, diskuteret af Tarp 2002) og *kirsebærstene* (flertal ubestemt form, udgået i RO5) selvfølgelig også en del af det danske (bøjnings)ordforråd, selv om de ikke er udbredte nok til at være med.

Når de anførte bøjningsformer skal opfattes som de eneste tilladelige inden for den officielle norm, bliver det et irritationsmoment når man nogle gange skal lære at tage hensyn til forbud der forekommer inkonsekvente. Eksempelvis kan nogle af substantiverne i den voksende gruppe der har engelsk flertalsendelse som normeret mulighed (fx *boomers*, *daybeds*, *items*, *toasts*), bevare flertals-s'et i bestemt form (således *daybedsene*, *itemsene*), mens andre ikke kan (**boomersene*, **toastsene*). En artikel i *Nyt fra Sprognaevnet* medgiver at det ville være hensigtsmæssigt hvis man altid selv kunne udlede flertal bestemt form af flertalsformen, men forklarer at redaktionen har lagt mere vægt på at følge sprogbru- gen.⁷ Ordbogen bygger med andre ord på en beskrivelse af den mest udbredte sprogbrug, som så ophøjes til norm. Jeg er enig i at RO ikke skal indføre bøjningsformer uden empirisk grundlag, men jeg ville, som det er fremgået, foretrække at man afstod fra decideret normering af orddannelsen.

Der er dog ingen tvivl om at det er nyttigt for brugerne at få serveret opslagsordene ikke blot i deres grundform, men også med

7 <https://dsn.dk/nyt-fra-sprognaevnet/oktober-2023/normering-af-flertals-s-i-dansk-svensk-og-norsk-retskrivning>.

de gængse/anbefalede bøjninger skrevet helt ud. I 4. udgave kom verbernes nutidsformer med, og tak for det. Derimod er præsens participium (endnu) ikke en af de bøjningsformer der oplyses i RO – end ikke i retskrivningsreglerne. Som jeg har påpeget i anden sammenhæng (Juul 2025), er der ellers tale om en endelse der giver anledning til en betydelig mængde stavfejl. Og dog. Når det intetsteds i RO fremgår at *syng* har præsens participium-formen *syngende*, er “syngene” så egentlig en stavfejl?

5. Kommareglerne

Hvad angår reglerne for tegnsætning, må vi selvfølgelig en tur forbi kommareglerne. Her er den nye udvikling i RO5 at ordbogen ikke længere anbefaler udeladelse af komma før ikkeparentetiske ledsætninger.

Jeg har selv som daværende medlem af Dansk Sprognævns repræsentantskab været med til at vedtage denne ændring ud fra den betragtning at anbefalingen ikke lod til at have haft nogen videre effekt. Det store flertal af sprogbrugere ønsker tydeligvis at (forsøge at) markere starten af ledsætninger, selv om dette kræver betydelig grammatisk indsigt. Så lad dem dog. Helst så jeg at man helt droppede påbudt afgrænsning af ikkeparentetiske ledsætninger, for også slutkommaet kræver grammatisk indsigt på ganske højt niveau. Det er der ikke mange skoleelever der opnår uden et stort forbrug af undervisningstid, som nok kunne anvendes bedre. Så vidt jeg er orienteret, har vi ikke nogen tungtvejende dokumentation for at alle disse ledsætningsafgrænsende kommaer gør nogen væsentlig forskel for læsernes forståelse af tekster.

Med den droppede anbefaling har tilhængerne af startkomma klart nok vundet en sejr med RO5. Men lige for en ordens skyld: Det er fortsat tilladt at undlade at sætte startkomma, sådan som jeg gør i nærværende tekst.

Hvad jeg kan savne, er en mere pædagogisk rammesætning af de kommaregler som vi nu engang hænger på. De indledende bemærkninger (retskrivningsreglernes § 45) anfører at komma-reglerne “med enkelte undtagelser” er “identiske med reglerne for traditionelt grammatisk komma”. Den forudsætningsløse sprogbruger der slår op i RO5 for at lære at sætte komma, må føle sig fremmedgjort her, for hvad er så de traditionelle regler for grammatisk komma? Det står der ikke noget om i RO5. Man skal helt tilbage til ordbogens 1986-udgave for at finde regler for noget der kaldes grammatisk komma.

Som det fremgår, præsenterer § 45 muligheden for at udelade startkomma som en *undtagelse* fra noget der i sig selv må forekomme dunkelt (nemlig det grammatiske komma). Som jeg ser det, ville det være både enklere og bedre at forklare at der med udeladelse af startkomma er noget grammatisk man *ikke* behøver at beherske (nemlig identifikation af ledsætningers begyndelse). Og hvorfor ikke starte med at forklare lidt om kommaers kommunikative funktion – hvor skal man overhovedet bruge dem? Kort sagt må det være på tide at viske tavlen ren og etablere en ramme der taler til nutidens sprogbrugere, inklusive dem der ikke var født i 1986. Måske kunne dette endda bidrage til at hæve niveauet i den evige kommadebat?

Selve formidlingen af reglerne har jeg ingen indvendinger imod. På plussiden kan det noteres at Sprognævnet sammen med Syddansk Universitet tilbyder et glimrende online kommakursus (komma.sdu.dk).

6. Ændrede stavemåder

Nu er det endelig tid at vende tilbage til mit indledende spørgsmål: Er der ord der skal staves anderledes i RO5? Ja, det er der da, men det er mindre justeringer der er tale om. Dette følger af

bekendtgørelsen om Dansk Sprognævns virksomhed⁸, som siger at hovedprincipperne for dansk retskrivning er traditionsprincippet og sprogbrugsprincippet. RO skal med andre ord videreføre de stavemåder vi allerede bruger (traditionsprincippet), og kun indføre justeringer hvis såkaldt gode og sikre sprogbrugere er begyndt at afvige fra traditionen (sprogbrugsprincippet).

For skoleelever ville det oplagt være en fordel hvis RO tilpassede de talrige stavemåder der ikke stemmer særlig godt med udtalen i moderne dansk (Juul & Heegård Petersen 2021). Det giver den nævnte juridiske ramme bare ikke hjemmel til, medmindre de famøse gode og sikre sprogbrugere af sig selv er begyndt at bruge mere lydrette stavemåder. At dette ikke sker i større stil, skyldes formentlig at gode og sikre skriftsprogsbrugere i vid udstrækning kender og respekterer RO-normen – for ellers kan man vel ikke kalde dem gode og sikre?⁹

Nogle af justeringerne i RO5 går faktisk i retning af det mere lydrette.¹⁰ Således har verbet *at pibe* nu fået den udtaleoverensstemmende sideform *at pive*. Og hvor å-lyden i *med det vons* tidligere skulle skrives med *o*, kan man nu også bruge *u*, som er den langt mest almindelige danske stavemåde for kort å-lyd.¹¹ En lignende tilpasning fik man med 1986-udgaven hvor *nubret* blev indført som alternativ til *nopret*. I forhold til skoleelevers staveudvikling er disse justeringer dog i mikroafdelingen. Hvis det skulle batte, kunne man tage fat på nogle af de andre ord med kort å-lyd og tillade *u* som stavemåde, fx *onkel* > “unkel” (jf. *dunkel*), *ost* > “ust” (jf. *rust*). Eller man kunne gå mere radikalt og systematisk til værks og indføre *å* som standardstavemåde for kort å-lyd. Men det er der ingen udsigt til at man gør, så længe sprogbrugen hos

8 <https://www.retsinformation.dk/eli/ta/2024/1059>.

9 <https://dsn.dk/wp-content/uploads/2021/01/september-2015-pdf.pdf>.

10 En oversigt over justeringerne kan ses på <https://ro.dsn.dk/mere-om-retskrivningsordbogen/udbydende-laesning-om-ro-5/normændringer-paa-enkeltordsniveau>.

11 Dette kan man verificere ved opslag på Center for Læseforsknings netsted bogstavlyd.ku.dk.

de gode og sikre følger traditionen. Det er betegnende at afsnittet om hvordan dansk ortografi kunne gøres mere regelmæssig i Christian Becker-Christensens værk *Dansk Retskrivning: Bogstav – Lyd – Bogstav* (Becker-Christensen 2021, afsnit 2.6.6-2.6.11) har overskriften “Nogle (u)mulige generelle reguleringer”.

Bekendtgørelsen om Dansk Sprognævns virksomhed åbner mulighed for tilpasninger, selv når sprogbrugen ikke af sig selv er gået den vej, når det gælder franske lån som *kalkule* og *kostume*. Disse skal dog fortsat skrives med *u*, selv om rimende låneord (*molekyle* og *kutyme*) har lydt *y*. Hvis man stemte om stavemåderne, ville tilpasninger nok også nærmere gå i den modsatte retning, altså til “molekule” og “kutume/koutume”. Gennemgående lægger sikre sprogbrugere nemlig ikke den store vægt på stavemåders lydrethed. Det ses i en række sprogbrugstilpassende justeringer i RO5, hvor *papaya* nu skal have *y* (RO4 havde *papaja* med *j*), og hvor *odør* nu igen kan staves *odeur*.

De nye stavemåder *papaya* og *odeur* gør ikke livet lettere for skoleelever, men de afspejler ordenes fremmede oprindelse. Ordene får med de nye stavemåder en duft af noget eksotisk. Hvad angår *odeur*, kan endelsen *-eur* desuden ses som et nyttigt signal om at sidste stavelse udtalemæssigt adskiller sig fra domesticerede ord som *chauffør*, *redaktør*, *kulør* som i modsætning til *odeur* har stød på sidste stavelse. Spørgsmålet er så om *charmør*, der heller ikke har stød i standardudtalen, skal have sideformen *charmeur*. Men nej, det skal den nok ikke, for *charmør*, der ligesom *odør* blev indført i 1986, lader til at have fået de gode og sikre (og inkonsekvente) sprogbruges accept.

7. anbefalinger

Jeg ejede i flere år en hæftet, uopsprættet udgave af *Ordbog over det danske Sprog*. Jeg endte med at splitte et af bindene ad og ta-

petsere en væg med de udfoldede ark. Måske kan jeg engang finde en lignende kreativ anvendelse af mit trykte eksemplar af RO5. Jeg tror ikke jeg kommer til at bruge det så længe der er adgang til en digital udgave som er hurtigere at slå op i, og som endda vil være mere kurant når der kommer opdateringer.

Til skoleelever og andre urutinerede skriftsprogbrugere vil jeg til de fleste formål anbefale netudgaven af DDO fremfor RO eftersom man her får det samme (oplysninger om normerede stavemåder) og mere til (betydningsbeskrivelser, flere sprogbrugseksempler, oplysninger om ikkenormerede stavemåder etc.). Og tilmed har DDO et lidt mere brugervenligt interface, der også som en nyttig håndsrækning omfatter en "mente du"-funktion ved opslag på ikkegenkendte ordformer.

RO har dog noget DDO ikke har, og som skoleelever bør stifte bekendtskab med, nemlig retskrivningsregler. Disse er efter min vurdering solidt formidlingshåndværk, men i hvert fald på kommaområdet kunne jeg, som det er fremgået, ønske mig en mere pædagogisk rammesætning, der ikke stiller større krav end højst nødvendigt til brugernes forudsætninger.

RO har også en retskrivningshistorisk hjemmeside, rohist.dsn.dk (ROhist), hvor man kan følge ordenes normerede stavemåder fra 1872 og frem til 2012 (dvs. RO4). Denne søgefunktion er måske mest interessant for ortografinorder, men den har også et stort potentiale i pædagogisk sammenhæng, fx når man læser tekster af ældre dato, når man vil vise at (skrift)sproget er en foranderlig størrelse, og når man vil sætte staveundervisning i perspektiv. Jeg håber derfor at ROhist bliver opdateret så RO5 kommer med. Det fremgår ikke af hjemmesiden om der er planer om dette.

Afslutningsvis vil jeg fremhæve at RO opfylder en vigtig funktion ved at identificere rigsmålets *almindelige* ordforråd. Jeg kunne ønske mig mere tydelighed om hvordan dette begreb defineres og operationaliseres – men grundlæggende er det nyttigt at få udpeget de ord man vil kunne støde på i almensproget, sådan at de for

sprogbrugerne mest relevante ord ikke forsvinder i et hav af sjældent brugte ord og fagtermer. Måske kunne man endda forestille sig en funktion der tillod brugerne at vælge mellem forskellige afgrænsninger af det almindelige.

En praktisk anvendelse af RO-redaktionens afgrænsningsarbejde er krydsogtværsspillet Wordfeud. Kun hvis en ordform er med i RO, er den acceptabel og pointgivende. Uden en sådan stringent afgrænsning mister spillet noget af sin appel. Jeg ærgrer mig dog stadig over en spilsituation hvor jeg kunne lægge samtlige syv bogstavbrikker (en såkaldt *bingo*) og stove ordet *boxplot*. Det ord er nemlig ikke med i RO.

Litteratur

Ordbøger, korpusser og digitale resurser

DDO = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. <<https://ordnet.dk/ddo/>> (april 2025).

Ordbog over det danske Sprog 1-28 (1919-1956). Det Danske Sprog- og Litteraturselskab. København: Gyldendal.

RO4 (2012) = *Retskrivningsordbogen*, 4. udgave. København: Dansk Sprognævn/Alinea.

RO5 (2024) = *Retskrivningsordbogen*, 5. udgave. København: Dansk Sprognævn/Alinea. Online: <<https://ro.dsn.dk/>> (april 2025).

ROhist = *Retskrivningsordbøger gennem historien*. Dansk Sprognævn. <<https://rohist.dsn.dk/>> (april 2025).

Anden litteratur

Andersen, Margrethe Heidemann (2025): *Afholdskvinde, ankerkvinde og redningskvinde – hvorfor skal sådanne ord være i Retskrivningsordbogen? I: Nyt fra Sprognævnet* 2025(1), 18-22.

- Becker-Christensen, Christian (2021): *Dansk Retskrivning. Bogstav – Lyd – Bogstav. En grafemisk beskrivelse af dansk ortografi*. Odense: Syddansk Universitetsforlag.
- Christiansen, Torben (2025): Ny udgave af Retskrivningsordbogen. I: *Læsepædagog* 73(1), 28-29.
- Hjorth, Poul Lindegård (1997): Ret og galt at stave. I: *Fra Egtved-pigen til Folketinget*. København: Det Kongelige Danske Videnskabernes Selskab. 109-132.
- Juul, Holger (2024): 1,6 kilo danske ord – og en manglende bøjningsendelse. En bemærkning til den nye retskrivningsordbog. I: *Mål og Mæle* 45(3), 5-7.
- Juul, Holger & Jan Heegård Petersen (2021): Om analyse af stavfejl. Årsager, korrespondensregler og udtalevariation. I: *NyS* 60, 112-149. <<https://doi.org/10.7146/nys.v1i60.129352>>.
- Tarp, Sven (2002): Kritiske bemærkninger til den officielle danske retskrivningsordbog. I: *LexicoNordica* 9, 195-208. <<https://doi.org/10.7146/ln.voi20.20883>>.
- Theilgaard, Liisa (2013). Anmeldelse af Retskrivningsordbogen, 4. udgave. I: *Danske Studier* 2013, 197-202.

Holger Juul
 lektor, ph.d.
 Center for Læseforskning
 Institut for Nordiske Studier og Sprogvidenskab
 Emil Holms Kanal 2
 DK-2300 København S
juul@hum.ku.dk

En rikhaltig ordbok om dialekterna i Thy

Eva Thelin

Torben Arboe: *Thy Ordbog. Dialekter i Thy i mands minde. På grundlag af Torsten Balles optegnelser i 1950-1970'erne*. Del 1 A-L, del 2 M-Å. Thisted: Forlaget Knakken. 2. oplag 2024. 686 s.

Thy Ordbog. Dialekter i Thy i mands minde är ett mäktigt verk om totalt 686 sidor, fördelade på två band i stort format. I ordboken beskrivs det dialektala ordförrådet i området Thy, beläget på nordvästra Jylland. Ordboken utgavs 2023 och nådde tydligen snabbt ett stort antal användare, för denna recension utgår från ett exemplar av den andra, oförändrade upplagan, tryckt 2024.

Redan på titelbladet upplyses man om att ordboken baserar sig på uppteckningar gjorda på 1950-1970-talet av Torsten Balle. Vem var då denne man? Av den minnesruna över honom som återges i ordboken framgår att han var lärare, infödd i Thy och själv dialekttalare. Hans kunnighet i att använda ljudskrift och förmåga att särskilja fina nyanser i uttal, gjorde honom till en enastående dialektupptecknare och han samlade in material åt Institut for Jysk Sprog- og Kulturforskning, vilket sedan kommit att användas i redigeringen av *Jysk Ordbog* (JO; 2000-). Cirka 17 000 maskinskrivna ordlappar blev det inalles och det är den samlingen som denna ordbok baserar sig på. Orden hämtade han huvudsakligen från sin hemsocken, Torsted, men han tillförde uppgifter också från andra socknar i Thy. I sitt uppteckningsarbete utgick Balle från två ordböcker, som han fick sig tillsända från institutet. Till att börja med arbetade han utifrån *Morsingmålets Ordforråd* (1948-1954) av A. C. Skyum, en ordbok som beskriver målet på ön Mors, belägen intill Thy. Senare gick han igenom sitt material utgående från H. F. Feilbergs *Bidrag til en Ordbog over jyske Al-*

muesmål (1886–1914) och han kompletterade också med ord från andra källor. Efter att 1962 ha inköpt en bandspelare börjar Balle spela in äldre dialekttalare och ur inspelningarna exciperar han ytterligare ord. 25 av inspelningarna transkriberar han och citat från transkriptionerna ingår i ordboken.

Ordbokens redaktör, Torben Arboe, är också han från Thy och har sedan 1979 varit verksam som redaktör på JO. Arboe berättar i ordbokens inledning att han 1980 hade som uppgift att kontrollläsa excerpter för JO som gjorts av de transkriptioner Balle åstadkommit och att han redan då slogs av tanken att Balles ordsamling borde ges ut som ordbok. Det skulle dock vara ett ytterst tidskrävande projekt. För att åtminstone göra uppgifterna allmänt tillgängliga påbörjades 2013 en databas, där uppgifterna från Balles ordlappar skrevs in av flera medhjälpare. 2020 var hela materialet infört från A till Å, sänar som på språkproven, som det för A–L registrerats maximalt ett av per ord, för M–Å inga alls, allt för att spara tid. Möjligheten att faktiskt ge ut ordboken verkade nu inom räckhåll och efter att ha fått ett löfte om utgivning från Lokalhistorisk Arkiv i Thisted gick man vidare i arbetet. Resterande språkprov registrerades och alla data fördes över från databasen till Word-filer, varpå redigerandet kunde ta sin början.

I den här recensionen sägs först något om ordbokens megastruktur, därefter något om vilka urvalsprinciper som tillämpats och hur artiklarna har indelats. Sedan följer avsnitt om normaliseringen av uppslagsord och olika typer av hänvisningar, om artiklarnas innehåll och struktur samt om de illustrationer som finns i ordboken. Avslutningsvis görs en sammanfattande värdering.

1. Megastruktur

Ordbokens första band inleds, efter titelbladet, med en innehållsöversikt (s. 5–6) följd av förord och en karta som visar de socknar

det finns språkprov ifrån (s. 7), inledning och användarvägledning m.m. (s. 8–50) samt minnesrunan över Torsten Balle (s. 51). Därefter vidtar från och med s. 53 själva lemmalistan från A till L. Första delen avslutas av en karta över samtliga socknar i Thy cirka 1921. Det andra bandet innehåller resterande ord från M till Å. På verkets allra sista sida, under de avslutande orden på *å*-, finner man en mycket kortfattad presentation av redaktören.

Att ordboken är försedd med en innehållsöversikt är välkommet, främst i och med att avsnittet Inledning, användarvägledning m.m. är relativt omfattande och har nio numrerade avdelningar med ytterligare underavdelningar. ”Inledning” är egentligen rubriken för hela avsnittet, medan ”användarvägledning” är den tredje av underavdelningarna. Den hade gärna fått vara sidoställd med inledningen, inte underordnad den, för att ytterligare förenkla för användarna att hitta fram till den. Viss information kunde också ha presenterats i mer samlad form. Den användare som undrar över hur en viss förkortning ska uttydas får nu söka svaret antingen i avsnitt 3.2 där först ordklassernas förkortningar är listade och längre ned övriga grammatiska förkortningar m.m. eller i avsnitt 3.5 där förkortningar som används i Balles betydelseangivelser är förtecknade. Innan man känner till vad en förkortning står för kan det vara svårt att veta vilken av kategorierna man ska söka sig till. I vägledningen redogör annars Arboe förtjänstfullt för hur han i redigeringsarbetet förhållit sig till Balles uppteckningar. Utgångspunkten tycks vara att uppteckningarna återges som de befins. Balles val av rättskrivning följs nästan överallt och understrykningar han gjort har behållits. Ibland har det emellertid funnits behov att supplerat uppgifter av olika slag och sådana tillägg har då satts inom hakparentes.

Större delen av inledningen ägnar Arboe åt att beskriva karakteristiska drag i dialekterna i Thy med exempel hämtade ur ordboken. Systematiskt genomgår han ordklass för ordklass och uppmärksammar därefter särskilda förhållanden när det gäller

dialekternas fonetik, grammatik och syntax samt fasta ordförbindelser och liknande. Också här redovisar han om något lagts till eller ändrats i Balles uppgifter.

2. Urvalsprinciper och artikelindelning

Av ordbokens inledning framgår att Balle helst ville koncentrera sig på de dialektala orden och att han var kritisk till Skyums ordbok för att den innehåller "[a]lle disse moderne ord, som fylder så megen kostbar plads og tager så megen tid at sigte fra" (återgivet ur ordboken s. 8). Av allt att döma har samtliga ord, såsom de befins i Balles samling, tagits med i ordboken. Vilka ord är det då han i praktiken valde att teckna ned? För att få svar på det har jag gått igenom partiet *h-*, en relativt omfattande del som utgörs av 926 lemman, dessutom en bokstav som jag själv är bekant med genom att jag nyligen redigerat orden på *h-* för *Svenskt dialektlexikon*.

Ord i dialekterna kan vara särdialektala, det vill säga inte existera i standardspråket överhuvudtaget. Många ord är dock gemensamma för dialekter och standardspråk, men uppvisar skillnader i uttal, böjning, betydelser, syntaktisk användning osv. i varierande grad. Att Balle inte bara samlat in särdialektala ord framgår ganska snart av det stora antal ord som har betydelser som definierats med " = rm." (dvs. lika med rigsmål).

Av de 926 lemnarna på *h-* har 359 sin enda eller alla sina betydelser gemensamma med riksdanskan. För många av dessa ord anförs därutöver bara uttalsformer. Orden är dessutom ofta påfallande moderna i den meningen att de bör vara fullt gångbara och frekvent använda i dagens Danmark. Det gäller ord som *halsbånd*, *halvsøster*, *handske*, *hund*, *hvid* och *hviske* för att nämna några. Ett avvikande uttal har uppenbarligen varit tillräckligt för att Balle skulle finna dessa ord värda att uppteckna. Han har

även uppptecknat egennamn, såväl personnamn som ortnamn, och här är det namnens lokala uttalsformer som gjort att de finns med.

Som svensk är det givetvis svårt att göra några språkliga bedömningar av danskan. Några gånger undrar man dock om inte vissa ord saknar uppgiften ”=rm.”. Bland annat gäller det de för oss svenskar så speciella danska räkneorden. I ordboken uppges inte att *halvferds*, *halvfems* och några till är riksdanska.

I en hel del av de 359 artiklarna upplyses dock om mer än ordens uttal. Det redogörs exempelvis för uttryck, talesätt o.d. som orden ingår i och man får ta del av språkprov samt realiauppgifter av skilda slag. Dessa upplysningar kan säkert ha bidragit till att Balle valt att dokumentera orden.

Såväl simplex som avledningar och sammansättningar representeras av egna artiklar i ordboken. Det gäller även fasta verbförbindelser i de fall det finns ett stort antal av dem till samma verb. Som exempel följs **holde** av knappt två spalter med artiklar från **holde af** till **holde øje med**. När förbindelserna är få till antalet har Balle nöjt sig med att redovisa dem i verbets artikel. *Harve af* nämns i artikeln **harve**² och *hamle op* har ett eget moment i artikeln **hamle**. Också homografer redovisas i skilda artiklar, där uppslagsformerna följs av upphöjda siffror. Olika ordklasser ingår i samma nummerserie och orden redovisas i en fastlagd ordningsföljd efter ordklassstillhörighet.

3. Normalisering och hänvisningar

En aspekt att särskilt beakta för dialektlexikografer är att dialekterna i huvudsak är talspråkliga. För den ordboksanvändare som bara vet hur ett ord låter, kan det vara svårt att föreställa sig hur ordet skulle se ut i skrift. Har ordboken dessutom valt att ha uppslagsord som är normaliserade, dvs. i en konstruerad standard-

språklig form, kan det vålla problem för användaren att hitta fram till ordet. Normaliserade uppslagsord förekommer främst i dialektordböcker som täcker in ett större geografiskt område med rik variation i ordens uttal. Att låta alla uttal representeras av egna uppslagsformer ter sig i sådana fall ofta ogörligt. För ordböcker som behandlar dialekten i mindre områden är det ofta möjligt att ha dialektnära uppslagsformer och så sker också ibland.

I Thy-ordboken återges artiklarnas uppslagsord i sin standard-språkliga form när det rör sig om ord som finns i riksdanskan. Särddialektala ord som ingår i JO har normalt getts den uppslagsform som används där. I regel överensstämmer den med Balles val av uppslagsord. När Balle någon gång har en annan uppfattning anges hans form med JO:s alternativ tillagt i hakparentes (se figur 1). Stundom har han föreslagit flera möjliga uppslagsformer, vilka då vanligen är åtskilda med snedstreck.

hjelle s. *hjel̥* – *hjel̥ər* = tørrestativ til fisk eller tøj.
 1331 Rær. [JO: *hield* s.] – Om vasket tøj: *de kam*
'te' å 'heŋ po ə hjel̥ det kom til at hænge på hjel-
 len.

Figur 1: Artikeln *hjelle* med JO:s uppslagsform i hakparentes.

Trots den normalisering som gjorts i Thy-ordboken ligger uttalsformerna relativt ofta så pass nära uppslagsformen att användaren bör kunna hitta rätt. Ibland kan det dock vara lite svårare. Hur går det exempelvis att hitta fram till *hyw'* som är uttalet av substantivet respektive adjektivet **høj** eller till *hòw'* som är ett verb med uppslagsformen **hue**?

Ett sätt att kompensera för nackdelen med normaliserade uppslagsord är att supplera med hänvisningar från uttalsnära former till uppslagsformer. I Thy-ordboken har det bara skett undantagsvis, såsom för **hvævs** där man leds vidare till **hveps**, som har *hwæws* som en av uttalsformerna. Det hade varit till avgjord fördel om fler hänvisningar av detta slag hade lagts till. Andra gånger går

hänvisningarna i stället i motsatt riktning; från **hvilken tid** uppmanas man gå till det uttalsnära **hwitter**.

Flera andra typer av hänvisningar finns. Vanligast är de som görs från dialektala ord som ingår i definitioner, realiaupplysningar eller språkprov till de artiklar där dessa ord behandlas (se figur 2). Relativt ofta hänvisas också till ord med samma eller liknande betydelse. Andra kategorier av hänvisningar uppträder mer sporadiskt och är olika utformade, men tillför ändå något för användarna.

hækkeri s. *hèkə'ríx* = netværk af lister, som danner underlag for sejlene på møllens vinger. Består af lister [→ liste¹ 2], der går langs med vingebjælken, og spoler [→ spole 2], der går på tværs.

Figur 2: Från lister och spoler i artikeln **hækkeri** hänvisas till liste¹ 2 och spole 2.

4. Artiklarnas innehåll och struktur

Utifrån de inregistrerade uppgifterna i databasen har Arboe strukturerat upp Balles formuleringar i artiklar och moment. Ibland kan det ha lett till mycket detaljerade uppdelningar, särskilt vid formord som prepositioner och pronomen, där Balle ofta redovisat många nyanser och användningar. Generellt sett är dock artiklarna enhetligt uppbyggda.

Efter det inledande uppslagsordet följer ordklassangivelse i förkortad form. Undantagsvis kombineras den med en grammatisk upplysning, såsom s.pl. för substantiv som bara förekommer i plural. För verbet **hamle** anges att det är transitivt med v.tr., vilket aldrig skett för andra transitiva verb på *h-* i ordboken.

Efter uppgiften om ordklass återges ordets uttal i kursiv stil med den ljudskrift Balle använt sig av, vilket är en förenklad form av Dania, nästan densamma som används i JO. Uttalsuppgifter

kan ibland även inleda betydelsemoment i de fall ett visst uttal är kopplat till en viss betydelse. Uttalsformerna återges alltid i sin helhet. Överlag ger det ett tydligt intryck, men när varianter med mycket små skillnader i uttal anføres, kan det ibland vara svårt att upptäcka vad som egentligen skiljer dem åt.

Också böjningsformer av substantiv, adjektiv och verb redovisas med sina uttal i fullform. Många ord saknar dock uppgift om böjning. I synnerhet gäller det sammansättningar. I de fall senareledens ord ingår som simplex i ordboken finns alltid möjligheten att ta del av böjningen där.

Efter uttalsformer och eventuella böjningsformer följer betydelsebeskrivningen. Som redan nämnts har de betydelser som överensstämmer med riksdanskan definierats med ”=rm.”. I flera fall följer ändå en utförd definition och stundtals ges ytterligare upplysningar som inte förefaller helt nödvändiga (se figur 3).

hed adj. *hiːə / hiːð – hiːə / hiːð* = rm.; meget varm, så varm, at varmen føles ubehagelig. Væsker: over ca. 50° C. Luft: over ca. 30°. Om metal, kakkelovne, strygejern: så varmt, at det kan spytte - over 100° C. – *han bløw ˈhiːð i ə ˈhuːə* han blev hed i hovedet (vred, flov, forlegen). – *æ ˈæ så ˈhiːð i mi ˈhuːə* jeg er så varm i hovedet. – *ən hiːə / hiːð såmər* en [hed] meget varm sommer.

Figur 3: Adjektivet **hed** har samma betydelse som i rigsmål, men har även försetts med en utförd definition samt en beskrivning av vilka olika temperaturer det rör sig om.

För att definiera särdialektala betydelser används riksdanska ekvivalenter i den mån sådana finns, annars längre definitioner. Inte sällan kompletteras betydelsebeskrivningarna av realiauppgifter och ibland även av illustrationer (se nedan).

För synonyma ord finns alltid möjligheten att förse bara ett av dem med en utförd betydelse och sedan hänvisa från de övriga orden till den. Detta har aldrig skett för *h*- och kan väl ha att göra med att betydelserna fått en materialnära redovisning.

Proveniensen uppges endast för de uppgifter som inte är från Balles hemsocken Torsted och det görs med ett fyrsiffrigt "topografiskt nummer" följt av socknens namn, t.ex. 1331 Rær (se figur 1). Uppgiften är oftast placerad efter betydelsebeskrivningen. Nödvändigheten att använda numren undandrar sig min bedömning, men de står onekligen ut i texten och gör det lättare att få syn på dessa inte så frekventa uppgifter. De ord som är belagda från flera socknar har dessutom delats upp på moment, där vart och ett av momenten innehåller betydelser och övriga förekommande uppgifter från en av socknarna. Arboe har med denna uppdelning velat göra det lättare för läsaren att urskilja uppgifter från de andra socknarna, och så kan nog vara fallet, men detta sätt att gruppera materialet får ofta till följd att till synes samma betydelse upprepas i flera moment. Exempelvis gäller det verbet *hæfte*², vars artikel innehåller fyra moment med betydelseuppgifter från varsin socken och där definitionerna lyder: 1) fæstne, gøre fast, hæfte; om et søm (en skrue), som fæstnes med spidsen i træet, inden det slås (skrues) i, 2) fæstne, gøre fast, 3) fæstne (med søm, nagler, pløkke), slå fast, 4) fæstne, fastgøre.

Efter betydelseangivelsen följer eventuella språkprov. Balle har varit generös med sådana och de tillför också något till betydelsebeskrivningen. Ofta exemplifierar de ordets användning i uttryck och talesätt och visar typiska kollokationer. Särskilt flödande är språkproven för de mest frekventa orden som redovisas i längre artiklar. Hit hör exempelvis verbet *have*, pronomenet *hvad*, adverbet *hvor*, adjektiven *halv* och *hel* samt substantiven *hus* och *hånd*. För att användaren ska kunna identifiera orden som ingår i språkproven, har redaktören fattat det vällovliga beslutet att förse samtliga med ordrädda översättningar till riksdanska.

5. Illustrationer

Att artiklar är försedda med illustrationer är ett inte ovanligt inslag i dialektordböcker. Från svenskt håll kan man som exempel nämna *Ordbok över folkmålen i övre Dalarna* (1961–2022) och *Ordbok över laumålet på Gotland* (1972–1986), vilka båda innehåller ett större antal teckningar och foton samlade i särskilda häften till vilka det hänvisas i artiklarna. Kanske har det fallit sig naturligt för dialektlexikografer och upptecknare av dialektord att lägga till illustrationer, eftersom arbetet ofta inneburit att man dokumenterar en försvunnen eller försvinnande kultur, där de föremål som är knutna till den inte längre är allmänt bekanta och ibland svåra att beskriva till utseendet.

Också Balle har gjort teckningar till en del av sina ord och dessa har införlivats i ordboken efter att ha renritats av en illustratör. Denne har även gjort illustrationer av växterna i ordboken, men då delvis utgått från *Danmarks vilde planter* (Skytte Christiansen 1958–1966). Teckningarna är genomgående välgjorda och tydliga penn-teckningar i svartvitt. Växterna är dessutom sparsamt färglagda.

Illustrationerna är placerade inuti spalterna, vanligen direkt efter den definition av ordet som teckningen hör samman med. Ibland, väl av utrymmestekniska skäl, har illustrationerna placerats på kort avstånd från artikeln eller längsmed flera artiklar. Inuti artiklarna signaleras med "ILL." att det finns en tillhörande illustration. Av någon anledning används "ILL." aldrig när illustrationen föreställer en växt. Några problem att hålla samman illustration och artikel bör dock aldrig uppstå. Inom illustrationerna nämns dessutom alltid uppslagsordet (se figur 4). Ett fåtal gånger är det två närbesläktade eller synonyma ord, såsom *hag* och *hage* respektive *hjørne* och *krog*, som nämns. I det senare fallet återfinns samma illustration vid båda orden. Även vid de avbildade växterna förekommer ibland två benämningar, men då är den första dialektal, den andra växtens florabenämning.

hollænder s. 'hò,læn'ər = kyse af tyndt, hvidt eller blomstret stof med stivere og med nedhængende krave. Brugtes ved høstarbejde. ILL.

u

HOLLÆNDER



Figur 4: Artikeln **hollænder** med tillhörande illustration.

De föremål som försetts med illustrationer är väl valda; det är sällan en illustration tycks överflödig. I enstaka fall skulle man ha önskat sig ytterligare förtydligande bilder, men avsaknaden av illustrationer kan väl antas bero på att det inte funnits några i Balles material och att under redigeringsarbetet komplettera med teckningar av olika, idag kanske svåråtkomliga föremål, har säkert tätt sig alltför tidsödande eller besvärligt.

Illustrationerna av växter har ju tillkommit i efterhand, men här har man kunnat utgå från en flora. Varför de lagts till motiveras inte, men den tilltagande så kallade biologiska analfabetismen, där man inte kan förvänta sig att alla vet hur en gullviva (*himmerignøgle*) eller fläder (*hyld*) ser ut, kan vara ett tänkbart skäl. För vissa växtbenämningar är illustrationerna särskilt klargörande, som när samma benämning förekommer både i dialekt och i danska floror, men avser olika arter. Det gäller dialektens *havtidsel* som avser strand-mandstro, *Eryngium maritimum* (dvs. svenskans martorn) och dialektens *havtorn* som avser havtidsel, *Hippophaë rhamnoides* (dvs. svenskans havtorn).

6. Sammanfattande värdering

I Thy-ordboken presenteras ett fylligt ordförråd, där man får rikhaltiga upplysningar om vart och ett av orden. Man kan inte annat än imponeras av Torsten Balles idoghet och noggrannhet som upptecknare. Definitionerna är välformulerade och begripliga och de tillhörande illustrationerna utgör ett fint tillskott. Språkproven bidrar till förståelsen av ordens användning och de har gjorts lättillgängliga genom att redaktören försett samtliga med översättning.

Den främsta invändningen rör de normaliserade uppslagsorden och bristen på hänvisningar till dem från uttalsnära former. Här kan användarna stundtals få problem att hitta ord som ändå finns i ordboken. Att lägga till hänvisningar till en eventuell tredje tryckt upplaga är kanhända inte realistiskt att tänka sig. Förutom det arbete som skulle krävas, är de som redan köpt ordboken kanske inte beredda att införskaffa nya exemplar, även om det skulle innebära förbättrade sökmöjligheter. Skulle ordboken dock komma att publiceras digitalt kan ju fler sökmöjligheter erbjudas och användarna få förslag på närliggande alternativ till ord man valt att söka efter men inte fått träff på.

I artikeln lyfts enskilda onöjaktigheter fram, men det får inte skymma det allmänna intrycket av det gedigna arbete som Thy-ordboken är. Den är ett gott exempel på att ordboksarbete kräver uthållighet, samarbete och beslutsamhet för att man ska nå fram till målet. Och resultatet har blivit till *en sand fornøjelse* för alla som är intresserade av dialekterna i Thy.

Litteratur

Feilberg, H. F. (1886–1914): *Bidrag til en Ordbog over jyske Almuesmål*. Kjøbenhavn: H. H. Thiele.

- JO = *Jysk Ordbog* (2000-). Peter Skautrup Centret for Jysk Dialektforskning, Institut for Kommunikation og Kultur, Aarhus Universitet.
- Levander, Lars & Stig Björklund (1961-2022): *Ordbok över folkmålen i övre Dalarna*. Uppsala: Dialekt- och folkminnesarkivet.
- Klintberg, Mathias & Herbert Gustavson (1972-1986): *Ordbok över laumålet på Gotland*. Uppsala: Lundequistska bokh.
- Skytte Christiansen, Mogens (1958-1966): *Danmarks vilde planter: med et udvalg af vore nordiske nabolandes flora*. København: Branner og Korch.
- Skyum, A. C. (1948-1954): *Morsingmålets Ordforråd* 1-2. Aarhus: Universitetsforlaget.
- Svenskt dialektlexikon*, under arbete vid Institutet för språk och folkminnen i Uppsala.

Eva Thelin
 forskningsarkivarie, fil. kand.
 Institutet för språk och folkminnen
 Box 135
 SE-751 04 Uppsala
 eva.thelin@isof.se

MELDINGAR

Nyt fra bestyrelsen for Nordisk Forening for Leksikografi

Einar Freyr Sigurðsson

Nordisk Forening for Leksikografi (NFL) arrangerer symposier hvert år og konferencer hvert andet år. Symposierne samler et mindre antal specialister fra de forskellige nordiske landes konferencerne henvender sig til et bredere publikum og samler leksikografer, sprogforskere, sprogteknologer og andre leksikografiinteresserede fra hele Norden.

Det 32. *LexicoNordica*-symposium blev afholdt på Voksenåsen i Oslo d. 23.-25. januar 2025. Temaet var *Nordiske ordbøger – opdatering, udvikling og tilgængeliggørelse*. Skriftlige versioner af bidragene er udgivet i den tematiske del af dette bind af *LexicoNordica*. Symposiumet og tidsskriftet er en meget vigtig del af NFL's aktiviteter. Stor tak til hovedredaktørerne Henrik Hovmark og Terje Svardal og landsredaktørerne Helga Hilmisdóttir (Island), Louise Holmer (Sverige), Caroline Sandström (Finland), Bente Selback (Norge) og Liisa Deth Theilgaard (Danmark). Henrik, som har lagt en meget stor indsats i *LexicoNordica*, træder tilbage som hovedredaktør, og Helga slutter sig til Terje. Vi byder ligeledes Islands nye landsredaktør, Ellert Jóhannsson, velkommen, i lighed med Bente Selback som allerede afløste Kjetil Gundersen ekstraordinært ved det 32. symposium. Temaet for næste års symposium er *Brugerundersøgelser og -involvering i nordiske ordbøger* og vil blive afholdt på Voksenåsen d. 21.-23. januar 2026.

En fælles konference som kombinerede Den 18. Konference om Leksikografi i Norden og Nordisk Sprog møde 2025 blev afholdt i Vejle d. 21.-23. maj 2025. Vi er konferencekomiteen taknemmelig for en vellykket konference. På NFL's generalforsamling

i forbindelse med konferencen blev bestyrelsen i perioden 2025-2027 valgt med følgende sammensætning:

Formand:	Einar Freyr Sigurðsson
Næstformand:	Thomas Widmann
Kasserer:	Laila Höglund
Sekretær:	Margunn Rauset
Medlem:	Zakaris Svabo Hansen
Suppleanter:	Carina Nilstun og Solveig Arle

Mange tak til afgående kasserer Pär Nilsson, sekretær Kristin M. Magnussen og medlem Caroline Sandström for uselvisk arbejde for foreningen. Heldigvis vil afgående formand Thomas Widmann fortsætte i bestyrelsen, nu som næstformand.

Konferencerne finder sted på skift i landene i Norden, og de arrangeres af et eller flere fagmiljøer i det respektive land – konferencebidragene publiceres i *Nordiske Studier i Leksikografi*. Den 19. konference om leksikografi i Norden vil blive afholdt i Reykjavík d. 19.-21. maj 2027.

Information om bl.a. kommende konferencer, symposier og publikationer og referater fra NFL's generalforsamlinger kan findes på NFL's hjemmeside <<https://nordisk-leksikografi.com/>> og på NFL's Facebook-side, *Nordiska Föreningen för Lexikografi – NFL*). Medlemmer modtager også et nyhedsbrev med aktuel information. Samtlige numre af *LexicoNordica* og *Nordiske Studier i Leksikografi* publiceres digitalt på <<https://tidsskrift.dk/>> i samarbejde med Dansk Sprognævn. Betaling for medlemskab foregår digitalt, således at hvert medlem betaler direkte via NFL's hjemmeside, og de registreres i et fælles medlemsregister. Man kan kontakte bestyrelsen direkte via e-mail: nordisk.lexikografi@gmail.com.

De administrative opgaver i NFL varetages af bestyrelsesmedlemmerne og af *LexicoNordica*'s to hovedredaktører og fem landsredaktører. Mange personer bidrager også med deres arbejdsind-

sats, ikke mindst i forbindelse med udgivelse af *LexicoNordica* og konferencerapporterne.

NFL har grundlæggende en sund økonomi, men der er fortsat behov for ekstern økonomisk støtte for at kunne fortsætte med at arrangere resursekrævende initiativer som symposier og konferencer. Herudover bidrager værtsinstitutionerne med arbejdstid og andre resurser. NFL takker for alle eksterne økonomiske bidrag til konferencer, symposier og publicering af konferencerapporterne – ikke mindst Nordplus Nordens Sprog. Uden denne støtte ville disse aktiviteter ikke være mulige.

I sommer kom den triste nyhed om at NFL's mangeårige sekretær Rikke Hauge døde d. 4. juni 2025, 82 år gammel. Rikke var en meget vigtig del af NFL og tæt sammenvævet med foreningens mere end 30-årige historie. Hun var meget central i foreningens arbejde – f.eks. klarede hun i mange år og frem til for få år siden opgaver vedrørende økonomi og administration. Hun arrangerede også alt det praktiske ved symposierne, bidrog ved konferencerne og havde desuden mange kontakter gennem Nordisk Sprogkoordination og Språkrådet i Oslo. Rikke var en stor hjælp for dem der opsøgte hende, hvilket let kan ses når man gennemgår selskabets publikationer fra forskellige år og læser om alle dem som takker Rikke Hauge for hendes indsats. Vi vil mindes Rikke Hauge med stor taknemmelighed – og huske hende for hendes karakteristiske latter og altid gode humør.

Einar Freyr Sigurðsson

forskningsdocent

Árni Magnússon-instituttet for islandske studier

Edda, Arngrímsgötu 5

IS-107 Reykjavík

einar.freyr.sigurdsson@arnastofnun.is

REDAKTIONSANVISNINGER

1. LexicoNordica udkommer hvert år i november. Tidsskriftet indeholder leksikografiske bidrag som er skrevet på et af følgende nordiske sprog: dansk, finsk, færøsk, islandsk, norsk (bokmål eller nynorsk) og svensk. Bidrag på engelsk kan også optages hvis særlige forhold taler for det.
2. **Bidrag** sendes til det medlem af redaktionskomitéen som repræsenterer bidragsyderens land:
 - Danmark: Liisa Deth Theilgaard, Helgesensgade 1, 2. tv., DK-2100 København Ø. <liisa.theilgaard@gmail.com>.
 - Finland: Caroline Sandström, Institutet för de inhemska språken, Hagnäskajen 6, FI-00530 Helsingfors. <caroline.sandstrom@sprakinstitutet.fi>.
 - Island: Helga Hilmsdóttir, Árni Magnússon-institutet for islandske studier, Arngrímsgötu 5, IS-107 Reykjavík. <helga.hilmsdottir@arnastofnun.is>.
 - Norge: Bente Selback, Institutt for lingvistiske, litterære og estetiske studium, Universitetet i Bergen, Postboks 7805, NO-5020 Bergen. <bente.selback@uib.no>.
 - Sverige: Louise Holmer, Institutionen för svenska, flerspråkighet och språkteknologi, Göteborgs universitet, Box 200, SE-40530 Göteborg. <louise.holmer@svenska.gu.se>.

Fristen for aflevering af bidrag er **den 1. april** hvis artiklen skal kunne trykkes i det nummer af tidsskriftet som udkommer samme år. Bidraget indleveres digitalt i både tekstbehandlingsformat (.docx) og i PDF-format. Dette gælder også evt. reviderede versioner.

3. **Illustrationer** der skal medtages i artiklen, indsættes i manuskriptet og vedlægges som separate billedfiler, helst i JPG-format og minimum 300 ppi. Tabeller udført i Word indsættes

i manuskriptet og kræver ikke særskilt billedfil. Der refereres eksplicit til figurer og tabeller undervejs i teksten, fx ”jf. figur 1”, ”se tabel 3” e.l., men IKKE ”se følgende figur/tabel”. De vedlagte billedfiler nummereres tydeligt og i overensstemmelse med den rækkefølge og den angivelse som bruges i manuskriptet. Dette gælder også evt. reviderede versioner af artiklen, hvor antal og rækkefølge af illustrationer/billedfiler (og tabeller) kan være ændret.

4. Bidraget skal forfattes i LexicoNordicas **stilark** (.docx), der kan rekvireres ved henvendelse til redaktionen. Stilarket er på forhånd opsat med korrekte marginer og forhåndsdefinerede typografier. Når man har modtaget stilarket, tages der en kopi af stilarket, som herefter omdøbes efter følgende model: ”[forfatter(e)]_LN30”. Artiklen udfærdiges herefter i det omdøbte stilark, og der vælges kun foruddefinerede typografier fra stilarket. Ved evt. problemer eller tvivlsspørgsmål rettes henvendelse til redaktionen.
5. **Manuskriptet** indledes med titel på artiklen og forfatterens navn. For tematiske og ikke-tematiske bidrag følger et **abstract** på engelsk på op til 10 linjer og dernæst selve artiklen, som opdeles i afsnit. Bidrag på engelsk tilføjes et abstract på dansk, norsk eller svensk. **Afsnit** nummereres efter følgende model: 1.; 1.1.; 1.1.1. (højst tre niveauer; ved henvisninger i teksten udelades slutpunktum, fx ”jf. afsnit 2.1”; der henvises eksplicit til afsnit i teksten, fx ”jf. afsnit 2.4”, ”se videre afsnit 4”, men IKKE ”se ovenstående/følgende afsnit”). Bidraget afsluttes med angivelse af forfatterens navn, titel samt post- og e-mailadresse. Bidrag kan normalt have et omfang på højst 20 sider inkl. litteraturliste. Bidraget afleveres **uden orddeling** af nogen art. Jf. i øvrigt stilarket mht. typografi, blanke linjer o.l.

6. **Citater:** Kortere citater (op til 3 linjer) bringes som en del af teksten med dobbelte anførselstegn omkring, mens længere citater eller fremhævelser af større vigtighed gives i et afsnit for sig selv **uden** anførselstegn (vælg typografien LN-citat i stilarket).
7. Vi anbefaler en meget tilbageholdende brug af **fodnoter**. Evt. nødvendige noter gennemnummereres i teksten med højtstillet angivelse uden parentes. Der anvendes fodnoter, ikke slutnoter.
8. **Litteraturhenvisninger** foretages i teksten efter følgende model:

som det fremgår af Herbst (2009)
 som det fremgår af Borin & Forsberg (2011:18)
 (se Herbst 2009:158ff.)
 (jf. Borin & Forsberg 2011:49–52)

For kilder med tre eller færre forfattere anføres efternavnene på alle forfattere i henvisningen, fx ”Gudiksen 2009”, ”Gudiksen & Hovmark 2020”, ”Gudiksen, Hovmark & Monka 2015”. For kilder med fire eller flere forfattere anføres kun det første efternavn efterfulgt af ”et al.”, fx ”Gudiksen et al. 1999”. Forfatternavnene adskilles af komma, undtagen de to sidste navne, som adskilles af ”&” (tilsvarende hvis der kun er to forfatternavne). I litteraturlisten skrives alle forfatternavne imidlertid ud, også hvis der er flere end tre forfattere.

9. I den løbende tekst angives IKKE hele **internetadresser**, men et forfatternavn eller en angivelse af titlen på internetbidraget, som herefter bruges i litteraturlisten. I litteraturlisten angives internetadresser i deres fulde form omgivet af < >, og måned og årstal for sidste tidspunkt for opslag på adressen anføres umid-

delbart efter i parentes, fx ”<<https://ordnet.dk/ddo/>> (april 2025)”. Hvis der er oprettet en unik reference til en digital kilde (DOI), anføres denne altid i litteraturlisten. DOI-adressen indsættes også i sin fulde form omgivet af < >, fx <<https://doi.org/10.7146/ln.v1i31.151283>>. Internetadresser, som har vundet indpas som titler af propriel karakter (fx ”svenska.se”), kan undtagelsesvis anføres som sådan i den løbende artikeltekst.

10. **Særlige angivelser:** Vær meget tilbageholdende med brug af fede typer; **sprogksempler** markeres med kursiv, fx: ”ordet *ungkarl* har synonymet *alenemand*”; **betydninger** af sproglige enheder angives ved hjælp af enkelte anførselstegn, fx: ’en ugift mand’; dobbelte anførselstegn bruges ved **citater** eller **forbehold**, fx: De er vokset op i de ”glade” tressere. Tegnsætningsreglerne, bl.a. for brug af komma, tankestreg, bindestreg (i betydningen ’fra ... til’), er forskellige i de nordiske lande, og forfatterne skal følge reglerne for det sprog som bruges i artiklen. Titler på ordbøger o.l. sættes i kursiv, fx ”*Den Danske Ordbog* er ...” og gives evt. en introduktion første gang titlen nævnes. Hyppigt anvendte titler kan evt. erstattes af en forkortelse, der indsættes i parentes første gang titlen nævnes, og som herefter anvendes, fx ”*Den Danske Ordbog* (DDO) er ...”.

11. Litteraturangivelser

I litteraturlisten anføres forfatternavne efter følgende model:

Gudiksen, Asgerd ([årstal])

Gudiksen, Asgerd & Henrik Hovmark ([årstal])

Gudiksen, Asgerd, Henrik Hovmark & Malene Monka
([årstal])

Ved mere end ét bidrag fra samme forfatter anføres bidragene i kronologisk rækkefølge. Alle bidrag hvor en person er enefor-

fatter anføres før bidrag hvor samme person er førsteforfatter sammen med andre forfattere, fx ”Nielsen 1999, Nielsen 2020, Nielsen & Krogh 2010”.

I tilfælde af en længere litteraturliste kan den inddeles i to dele i lighed med nedenstående eksempel. Hvad angår angivelser som *red.*, *eds.*, *Hrsg.*, anbefales det så vidt muligt at bruge originalsproget. Det vigtigste er dog konsekvens inden for samme liste.

I tvivlstilfælde rettes henvendelse til redaktionen.

Litteratur

Ordbøger, korpusser og digitale resurser

ALD (1948) = A.S. Hornby, E.V. Gatenby & H. Wakefield: *A Learner's Dictionary of Current English*. London: Oxford University Press.

BÍN = *Beygingarlýsing íslensks nútímamáls*. Kristín Bjarnadóttir (red.). Árni Magnússon-instituttet for islandske studier. <<https://bin.arnastofnun.is>> (marts 2021).

COBUILD (1987) = *Collins COBUILD English Language Dictionary*. Editor in Chief: John Sinclair, Managing Editor: Patrick Hanks. London/Glasgow: Collins.

DDO = *Den Danske Ordbog*. Det Danske Sprog- og Litteraturselskab. <<https://ordnet.dk/ddo>> (april 2021).

Italiensk-Dansk Ordbog (1999). Knud Andersen & Giovanni Mafera. København: Gyldendal.

Jarvad, Pia (1999): *Nye Ord. Ordbog over nye ord i dansk 1955-1998*. København: Gyldendal.

LBK = Leksikografisk bokmålskorpus. Tekstlaboratoriet, Institutt for lingvistiske og nordiske studier, Universi-

- tetet i Oslo. <<https://tekstlab.uio.no/glossa2/bokmal>> (august 2020).
- Norstedts stora engelska ordbok* (2000). Stockholm: Norstedts.
- Oxford-Hachette French Dictionary* (1994). Oxford: Oxford University Press.
- Risamálheildin (2017-2018). Stofnun Árna Magnússonar í íslenskum fræðum. <<https://malheildir.arnastofnun.is>> (februar 2020).
- Språkbanken Text. <spraakbanken.gu.se/> (marts 2021).
- Svenska.se = Svenska Akademiens ordboksportal. <<https://svenska.se/>> (april 2021).

Anden litteratur

- Delvin, Jacob, Ming-Wei Chang, Kenton Lee & Kristina Toutanova (2019): BERT: Pre-training of deep bidirectional transformers for language understanding. I: *Proceedings of the 2019 Conference of NAACL: Human Language Technologies*, Volume 1. Minneapolis, Minnesota: Association for Computational Linguistics. 4171–4186.
- Faarlund, Jan Terje, Kjell Ivar Vannebo & Svein Lie (1997): *Norsk referansegrammatikk*. Oslo: Universitetsforlaget.
- Haiman, John (1980): Dictionaries and Encyclopedias. I: *Lingua* 50, 329–357.
- Jónsdóttir, Halldóra & Þórdís Úlfarsdóttir (2011): ISLEX – en flersproget nordisk ordbog. I: Birgit Eaker, Lennart Larsson & Anki Mattisson (red.): *Nordiska studier i lexikografi* 11. Lund: Nordisk förening för lexikografi. 353–366. <<https://tidsskrift.dk/nsil/article/view/19351>>.
- Lakoff, George & Mark Johnson (1980): *Metaphors we live by*. Chicago/London: The University of Chicago Press.

Mugdan, Joachim (1985a): Grammatik im Wörterbuch: Wortbildung. I: Herbert Ernst Wiegand (Hrsg.): *Studien zur neuhochdeutschen Lexikographie IV*. Hildesheim/Zürich/New York: Olms. 237-308.

Nikula, Kristina (2012): Samspelet mellan text och bild i enspråkigt svenska ordböcker. I: *LexicoNordica* 19 (dette bind).

Nordenstorm, Leif (2017): Tro och tradition enligt Svenska Akademiens Ordlista. I: *Svensk kyrkotidning* 11/2017. <<https://svenskkyrkotidning.se/recension/tro-och-tradition-enligt-svenska-akademins-ordlista/>> (april 2021).

NRG = *Norsk referansegrammatikk*, se Faarlund et al. (1997).

12. LexicoNordica udkommer både som trykt tidsskrift og i en **digital udgave** på open access-plattformen Tidsskrift.dk. Ved indsendelse af et bidrag til redaktionen erklærer forfatterne sig derfor indforstået med både en trykt udgave og en digital udgave på open access-plattformen Tidsskrift.dk.