

«ORDFORRÅDET» – EN LEKSIKALSK DATABASE OVER ET UTVALG NORSKE ORD



Fra venstre: Hanne Gram Simonsen, Marianne Lind, Bjørn-Helge Mevik, Pernille Hansen og Elisabeth Holm. Foto: Vibeke Enstrøm Nøking.

Marianne Lind er språkforsker og arbeider som seniorrådgiver i Statped sørøst, fagavdeling språk/tale. Hun er også tilknyttet Forskergruppe for klinisk lingvistikk og språktilegnelse ved UiO.

Hanne Gram Simonsen er professor i lingvistikk ved Institutt for lingvistiske og nordiske studier, UiO, og arbeider med språkutvikling hos barn og språkavvik hos barn og voksne. Hun er medlem av Forskergruppe for klinisk lingvistikk og språktilegnelse.

Pernille Hansen er stipendiat ved Forskergruppe for klinisk lingvistikk og språktilegnelse ved Universitetet i Oslo. Hun forsker på ordforråd og grammatiske ferdigheter hos enspråklige og tospråklige barn.

Elisabeth Holm er masterstudent i lingvistikk ved Institutt for lingvistiske og nordiske studier, UiO, og arbeider som forskningsassistent i Forskergruppe for klinisk lingvistikk og språktilegnelse.

Bjørn-Helge Mevik har en doktorgrad i statistikk fra Universitetet i Oslo, og arbeider med vitenskapelig IT-støtte ved Universitetets senter for informasjonsteknologi, UiO.

I logopedisk arbeid og i forskning innenfor ulike fagfelt (språkvitenskap, logopedi, psykologi) trenger vi kunnskap om forskjellige egenskaper ved enkeltord. Hittil har det ikke eksistert noen samlet oversikt over ulike egenskaper ved en større gruppe norske ord. I februar 2013 ble imidlertid «Ordforrådet» lansert. Dette er en leksikalsk database med drøyt 1600 norske ord der man kan finne opplysninger om en mengde ulike egenskaper ved ordene. I denne artikkelen presenterer vi «Ordforrådet» nærmere.

STRUKTURER OG PROSESSER I DET MENTALE LEKSIKONET

Alle ordene vi kan er lagret i det mentale leksikonet i hjernen vår, og vi henter dem fram når vi snakker, lytter, skriver og leser (Aitchison, 2003). Ord har mange og ulike særtrekk, knyttet til så vel form og betydning som bruksmønstre, og informasjonen om et ord i det mentale leksikonet inkluderer informasjon om alle disse særtrekkene. De forskjellige særtrekkene er med på å påvirke både språktilegnelse, blant annet hvor tidlig barn lærer seg ulike ord, og språkbruk, for eksempel hvor lett eller vanskelig det er for oss å hente fram ordene når vi trenger dem. I tillegg til de ulike særtrekkene ved de enkelte ordene finnes det mange forbindelseslinjer

mellom ordene, og det er dermed vanlig å anta at det mentale leksikonet er organisert som et nettverk, der ulike ord er koplet sammen gjennom forskjellige typer forbindelseslinjer. Disse forbindelseslinjene dannes, styrkes og svekkes gjennom hele livet. Noen ganger tilegner vi oss nye ord, andre ganger går noen av de gamle ordene i glemmeboka. Nye erfaringer knyttet til ord i det mentale leksikonet vil kunne skape nye forbindelseslinjer mellom ord og forsterke noen av de forbindelseslinjene som allerede er etablert.

Stort sett foregår prosessene med å hente fram ord fra det mentale leksikonet raskt og uproblematisk, og vi merker ikke egentlig at det er noe vi bruker kognitiv energi på. Men av og til kan noen hver av oss oppleve at vi står fast litt mer. Vi har ordet «på tunga», men vi greier ikke å hente det helt fram, i alle fall ikke like raskt som ellers. Samtidig vet vi ofte mye om det ordet vi leter etter, både om betydningen og de grammatiske særtrekkene ved ordet og om noen lydige særtrekk, for eksempel hvilken lyd ordet begynner med, og hvor mange stavelser det har.

I løpet av den ordletingsprosessen som skjer når vi «har et ord på tunga», viser det seg at det ikke er tilfeldig hvilke andre ord som dukker opp. Disse andre ordene er ofte betydningsmessig relatert til det ordet vi leter etter (semantiske assosiasjoner), eller det kan være formmessige likheter (grammatiske og fonologiske) mellom det ordet vi leter etter, og ordene vi kommer på. Ecke (2009) diskuterer ulike gjenfinningsmønstre som er beskrevet hos personer som opplever «å ha ordet på tunga». Noen studier har vist at bruksfrekvens – hvor vanlig et ord er – spiller inn i leteprosessen på den måten at de ordene som dukker opp, ofte er mer frekvente enn det ordet vi leter etter. Andre studier har vist det motsatte, altså at de ordene som dukker opp, er lavfrekvente (sjeldne) sammenlignet med det ordet vi leter etter. Når det skjer, er tanken at språkbrukeren kanskje går ut fra at det ordet hun leter etter, er sjeldent, og at hun derfor starter leting blant de lavfrekvente ordene (Ecke, 2009). Hos personer som har en språkvanske – medfødt eller ervervet – er ordleting ofte mye vanligere enn hos personer uten språkvansker, og også hos personer med språkvansker kan vi se mønstre og forbindelseslinjer mellom de ordene vedkommende leter etter, og de

ordene som dukker opp underveis i leteprosessen (Laine & Martin, 2006).

«ORDFORRÅDET»: EN SØKBAR OVERSIKT OVER ULIKE EGENSKAPER VED ET UTVALG NORSKE ORD

I både forskning og praksis kan vi trenge opplysninger om ulike egenskaper ved enkeltord. Hittil har det ikke eksistert noen slik samlet oversikt for et større antall ord i norsk, og logopedier, språkforskere og psykologer har vært henvist til å samle inn og sammenstille informasjon om de ordene de er interesserte i, på egen hånd. Som resultat av et prosjekt i Forskergruppe for klinisk lingvistik og språktilegnelse ved Universitetet i Oslo (<http://www.hf.uio.no/iln/forskning/grupper/klinisk-ling/>) er denne situasjonen nå endret. Fra februar 2013 er databasen «Ordforrådet» fritt tilgjengelig på UiOs nettsider: <http://www.tekstlab.uio.no/ordforradet/>. I denne databasen finnes opplysninger om en mengde ulike egenskaper ved 1651 norske ord: 917 substantiver, 509 verb og 225 adjektiver. Grensesnittet for nettsiden er utarbeidet med hjelp fra Tekstlaboratoriet, UiO.

HVILKE ORD FINNES I «ORDFORRÅDET»?

Vi ønsket at ordene i «Ordforrådet» skulle være relevante i logopedisk praksis og valgte derfor å inkludere ord fra ulike typer tester og kartleggingsmateriell: *Foreldre-rapport for kommunikatív utvikling* (CDI) (Kristoffersen & Simonsen, 2012), *Verb- og setningstesten* (VOST) (Bastiaanse, Lind, Moen & Simonsen, 2006), *Psykolingvistisk kartlegging av språkprosessering hos afasi-rammede* (PALPA) (Kay, Lesser & Coltheart, 2009) og *Preteritumstesten* (Ragnarsdóttir, Simonsen & Plunkett, 1999). Fra PALPA har vi tatt med ord fra et utvalg av deltestene, primært de deltestene der billedlighet eller konkrethet inngår som en variabel. Fra PALPA og VOST har vi tatt med både målord og distraktorord fra de ulike deltestene.

I tillegg har vi tatt med ord fra semi-strukturerte bildebeskrivelser (data fra 60 normalspråklige og 40 afasi-rammede språkbrukere) (Lind, Kristoffersen, Moen & Simonsen, 2009; Korpijaakko-Huuhka & Lind, 2012) og ord som dannet grunnlaget for et verktøy for kartlegging av ordforråd hos tospråklige barn, et resultat av det europeiske samarbeidet Bi-SLI (<http://www.bi-sli.org/>).

Vi har også lagt til noen få substantiver og verb. Det gjelder slike ord som har en «tvilling», altså en homonym form, for eksempel '(å) danse' – '(en) dans', '(å) ønske' – 'et ønske'. Vi ønsket å ha muligheten til å studere egenskaper ved slike «tvillingord» nærmere, og derfor passet vi på å inkludere dem i basen.

Verbene i «Ordforrådet» er lagret i infinitiv med infinitivmerke ('å hoppe'), mens substantivene stort sett er lagret i ubestemt form entall med determinativ (ubestemt artikkel) ('en gutt'). Noen substantiver, nærmere bestemt de som kan referere til noe ikke-tellelig ('kaffe, vin, sjokolade, vann'), er lagret både med og uten determinativ. Noen tvetydige ord som finnes i det materialet vi har hentet ord fra, er ikke inkludert i «Ordforrådet». Det gjelder tvetydige ord der betydningene

ligger langt fra hverandre, og der det ikke er nok å føye til et infinitivmerke eller en determinativ for å vise hvilken betydning det er snakk om, for eksempel 'å sprette' (en ball eller en søm).

HVA SLAGS SØK ER MULIGE

I «ORDFORRÅDET»?

«Ordforrådet» er lagt opp slik at man kan foreta tre forskjellige typer søk. For det første kan man søke på et enkeltord og få opp de opplysningene som finnes om ordet i databasen. For eksempel kan man velge å søke på ordet 'strømpebukse', og når man da trykker «søk», vil resultatet se slik ut (se illustrasjon 1).

Den andre typen søk man kan gjøre i «Ordforrådet», er å søke på bestemte egenskaper – enten én egenskap eller

Ord	ei strømpebukse
Ordklasse	? substantiv
Billedlighet	? høy (7,01)
Subjektiv tilegnelsesalder	? tidlig (3,38)
Bruksfrekvens	? lav (879)
Antall fonologiske naboord	? 3
Fonologiske naboord	strømpebuksen, strømpebukser, strømpebuksa
Ordlengde	?
Antall bokstaver	12
Antall lyder	12
Antall stavelser	4
Ordstruktur	?
Avledet ord?	nei
Sammensatt ord?	ja
Lydstruktur	?
Begynner med friksjonslyd?	ja
Begynner med konsonantgruppe?	ja
Egenskaper ved substantiv	?
Kjønn	hunkjønn
Brukes primært i flertall?	nei
Massesubstantiv?	nei

Illustrasjon 1: Resultat av søk på enkeltord.

← Tilbake

Last ned

Fant 22 ord med følgende egenskaper:

- bruksfrekvens: lav
- billedlighet: høy
- ordklasse: substantiv
- begynner med friksjonslyd: ja

▲ Ord	◆ Ordklasse	◆ Billedlighet	◆ Bruksfrekvens	◆ Begynner med friksjonslyd?	◆ Begynner med konsonantgruppe?
en fyrstikk	substantiv	høy (6,86)	lav (921)	ja	nei
en hane	substantiv	høy (7,05)	lav (1507)	ja	nei
en sandal	substantiv	høy (6,80)	lav (1087)	ja	nei
ei sandkasse	substantiv	høy (6,64)	lav (1060)	ja	nei
en sebra	substantiv	høy (6,72)	lav (447)	ja	nei
en sjiraff	substantiv	høy (6,84)	lav (290)	ja	nei
ei skilpadde	substantiv	høy (6,65)	lav (1352)	ja	nei
ei sklie	substantiv	høy (6,78)	lav (429)	ja	ja
en skuff	substantiv	høy (6,70)	lav (1145)	ja	ja
en smokk	substantiv	høy (6,64)	lav (875)	ja	ja
ei snekkerbukse	substantiv	høy (6,60)	lav (132)	ja	ja
en snømann	substantiv	høy (6,83)	lav (694)	ja	ja
spagetti	substantiv	høy (6,85)	lav (411)	ja	ja
en squash	substantiv	høy (6,72)	lav (856)	ja	ja
et stearinlys	substantiv	høy (6,64)	lav (1510)	ja	ja
ei stekepanne	substantiv	høy (6,66)	lav (1222)	ja	ja
en struts	substantiv	høy (6,86)	lav (664)	ja	ja
ei strømpebukse	substantiv	høy (7,01)	lav (879)	ja	ja
en støvsuger	substantiv	høy (6,90)	lav (1202)	ja	ja
syttetøy	substantiv	høy (6,65)	lav (1626)	ja	nei
ei symaskin	substantiv	høy (6,84)	lav (689)	ja	nei
ei såpeboble	substantiv	høy (6,80)	lav (247)	ja	nei

Illustrasjon 2: Resultat av søk på egenskaper.

en kombinasjon av flere egenskaper – og så vil man få opp alle de ordene i databasen som har den eller de egenskapene man har søkt på. Hvis man for eksempel søker på følgende kombinasjon «lavfrekvente, høy-billedlige substantiver som begynner på en friksjonslyd», får man opp følgende resultat (se illustrasjon 2).

Den tredje og siste typen søk er muligheten for å søke spesifikt på ord fra ett eller flere av de kartleggingsverktøyene ordene i databasen er hentet fra. Man kan med andre ord avgrense søket til å gjelde kun ord som er hentet fra *Foreldrerapport for kommunikatív utvikling* (CDI), VOST, PALPA eller *Preteritumstesten*.

HVILKE EGENSKAPER VED ORD ER TATT MED I «ORDFORRÅDET»?

I «Ordforrådet» kan man finne opplysninger om følgende egenskaper ved ord: billedlighet, subjektiv tilegnelsesalder, bruksfrekvens, fonologisk naboskaps-tetthet, lydstruktur, ordlengde, ordstruktur og ordklasse. I tillegg er det noen egenskaper som bare gjelder for én av ordklassene. For verb kan man søke på om verbet inngår i en flerordsfrase, og man kan få opplysninger om transitivitet (om verbet tar objekt(er) eller ikke) og valens (hvor mange obligatoriske setningsledd verbet har knyttet til seg). For substantiver kan man søke på om substantivet er tellelig eller ikke, og man kan få informasjon om genus (grammatisk kjønn). Nedenfor beskriver vi hver av de søkbare egenskapene nærmere. På nettsidene kan man også få mer informasjon om de enkelte egenskapene ved å klikke på de små spørsmåls-tegnene som står ved siden av begrepene (se illustrasjon 1).

Billedlighet

Billedlighet dreier seg om hvor lett eller raskt et ord framkaller en sanseopplevelse eller forestilling, for eksempel et mentalt bilde eller en lyd. For eksempel framkaller ordet 'agurk' raskt en sanseopplevelse hos de fleste, mens 'urealistisk' ikke gjør det. Billedlighet er nært beslektet med konkrethet, men de to egenskapene er ikke helt synonyme. For eksempel finnes det ord som er lite konkrete, men allikevel høybilledlige, slik som ord for følelser (som 'sinne', 'kjærlighet'), og det finnes ord som er lavbilledlige, men konkrete (som 'hobbybonde'). Når man skal finne fram til et ords billedlighetsverdi, må man basere seg på vurderinger fra en gruppe språkbrukere. Man ber da den enkelte informant rangere ett og ett ord på en gitt skala etter hvor lett ordet framkaller en sanseopplevelse, og ordets billedlighet er gjennomsnittet av alle rangeringene. Det er vanlig å bruke en skala fra 1 til 7 der 1 tilsvarer «ingen forestilling», mens 7 tilsvarer «klar forestilling» (Paivio, Yuille & Madigan, 1968). Denne metoden har vært brukt i mange studier, blant annet for engelsk, fransk, italiensk, kinesisk og japansk.

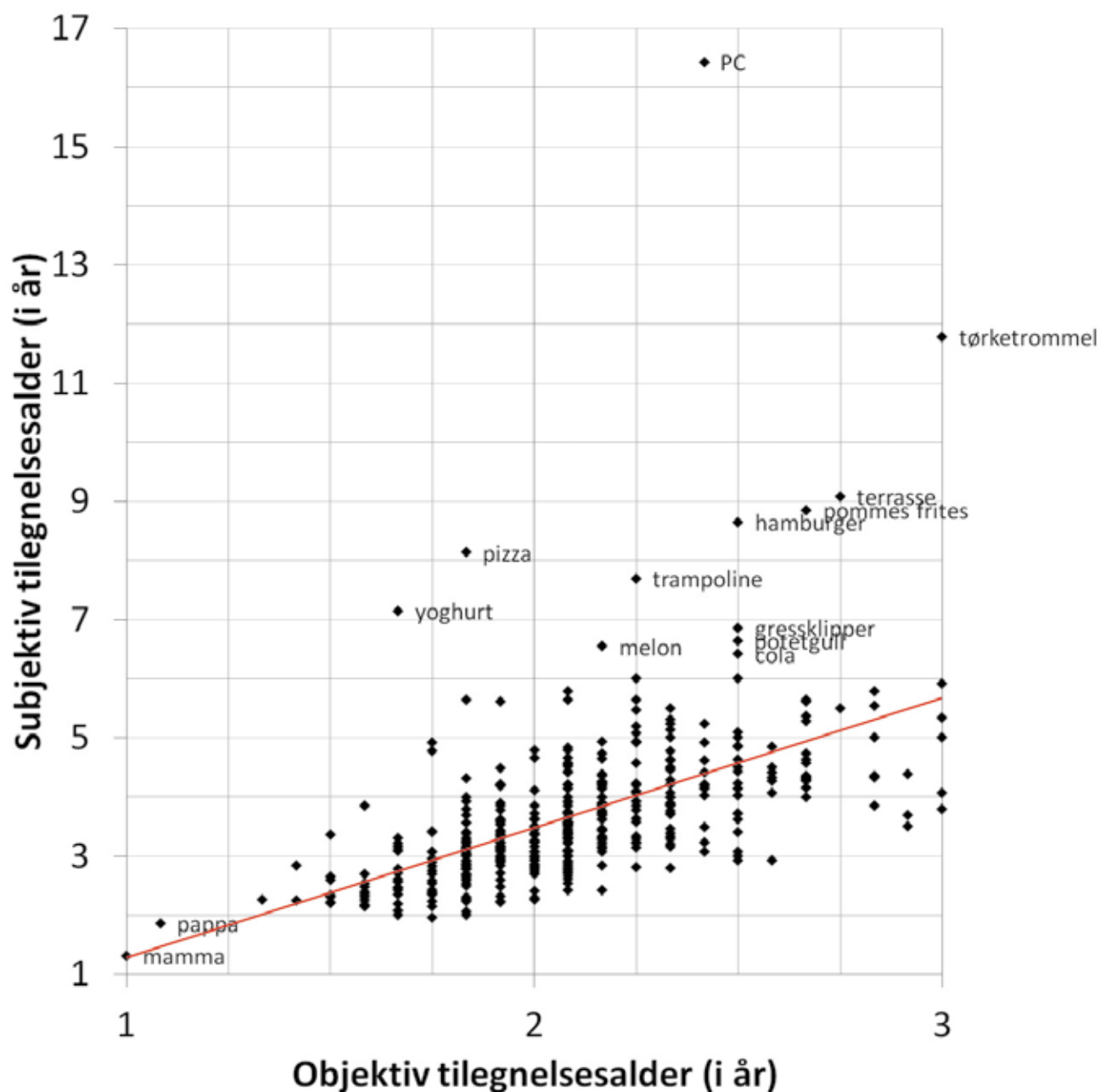
For å finne fram til billedligheten for ordene i «Ordforrådet» gjennomførte vi en nettundersøkelse der ca. 400 norskspråklige informanter rangerte 100 ord hver på en skala fra 1 til 7. I gjennomsnitt er hvert ord i «Ordfor-

rådet» vurdert av 23 personer. Grensene mellom høybilledlige, middels billedlige og lavbilledlige ord har vi definert slik: De 25 % minst billedlige ordene er lavbilledlige, og de 25 % mest billedlige er høybilledlige. Det innebærer at ord med en billedlighetsrangering fra 4,5 og nedover er lavbilledlige, og ord med en billedlighetsrangering fra 6,6 og oppover er høybilledlige. De resterende ordene utgjør mellomkategorien. En nærmere beskrivelse og drøfting av billedlighetsundersøkelsen finnes i Simonsen, Lind, Hansen, Holm og Mevik (2013).

Subjektiv tilegnelsesalder

Subjektiv tilegnelsesalder er et mål på hvor gammel man tror man var da man lærte et gitt ord. Rangeringen i «Ordforrådet» bygger på en nettundersøkelse med ca. 300 voksne informanter som skulle angi hvor gamle de mente de var da de lærte de ordene de ble presentert for. De skulle oppgi alderen i antall år på en skala fra 0 til 18, der 0 tilsvarte under ett år, 1 tilsvarte ett år og så videre, opp til 18 som tilsvarte 18 år eller eldre. Hvert ord ble vurdert av i gjennomsnitt ca. 20 informanter, og det er snittet av disse vurderingene som gir den subjektive tilegnelsesalderen for hvert ord. 25 % av ordene ble vurdert som tilegnet før 3 år og 7 måneder (3;7), og er kategorisert som lært tidlig. Tilsvarende ble 25 % av ordene vurdert som tilegnet etter 7;6 år, og disse er kategorisert som lært sent. De resterende ordene utgjør mellomkategorien.

For å undersøke hvor pålitelig dette målet er, har vi undersøkt samsvaret mellom subjektiv og objektiv tilegnelsesalder for 460 ord fra «Ordforrådet». Disse 460 ordene er hentet fra *Foreldrerapport for kommunikatív utvikling* (CDI) (Kristoffersen & Simonsen, 2012), hvor foreldrene til ca. 6500 norske enspråklige barn blant annet har fylt ut en sjekkliste over barnets ordforråd (se også Kristoffersen, Simonsen, Eiesland & Henriksen (2012) for en presentasjon av den norske CDI-studien). Vi har definert et ords objektive tilegnelsesalder som det første alderstrinnet da minst halvparten av foreldrene har krysset av for at barnet deres sier ordet. Generelt ligger de subjektive vurderingene av tilegnelsesalder noe høyere enn den objektive tilegnelsesalderen, men vi finner en signifikant korrelasjon mellom de to målene (se figur 1).



Figur 1. Korrelasjonen mellom objektiv og subjektiv tilegnelsesalder (i år). Ordene med lavest og høyest subjektiv tilegnelsesalder er merket med navn i figuren.

Som vi ser i figur 1, skiller noen ord seg ut ved at avstanden mellom objektiv og subjektiv tilegnelsesalder er veldig stor. For eksempel har 'en PC' en subjektiv tilegnelsesalder på over 16 år, mens objektiv tilegnelsesalder er 2 år og 5 måneder. Alle ordene som skiller seg ut med stor avstand mellom subjektiv og objektiv

tilegnelsesalder, er ord som er relativt nye i språket vårt, og det er naturlig at dagens småbarn støter på dem mye tidligere i livet enn hva som var tilfellet for de voksne informantene som deltok i undersøkelsen vår om subjektiv tilegnelsesalder.

Bruksfrekvens

Hvor ofte vi leser, skriver, sier eller hører forskjellige ord påvirker hvordan vi bruker dem, hvor lett det er å hente dem fram. For å gi et mål på bruksfrekvens har vi brukt NoWaC, et korpus på 700 millioner ord som består av norske tekster på internett (Guevara, 2010). Bruksfrekvensen for et ord angir hvor mange ganger ordet forekommer i en eller annen form i korpuset. Det vil si at alle forekomster av 'være,' 'er', 'var' og 'vært' tilsammen teller som bruksfrekvensen til verbet 'være'. I «Ordforrådet» kan du se om et ord har høy, lav eller middels bruksfrekvens i forhold til de andre ordene i databasen. Den mest frekvente fjerdedelen av ordene i «Ordforrådet» er kategorisert som høyfrekvent, mens den minst frekvente fjerdedelen av ordene er kategorisert som lavfrekvent. Det innebærer at ord som forekommer mindre enn 1800 ganger i NoWaC, er definert som lavfrekvente, mens ord som forekommer over 34000 ganger i NoWaC, er definert som høyfrekvente. De resterende ordene er kategorisert som middels frekvente.

Vi har valgt å bruke NoWaC for å anslå bruksfrekvens fordi dette korpuset per i dag er det desidert største tekstkorpuset for norsk. Dette korpuset er imidlertid relativt nytt, og det var for eksempel ikke tilgjengelig da de norske versjonene av VOST og PALPA ble laget. Eventuelle avvik mellom hva som er kategorisert som henholdsvis høyfrekvent og lavfrekvent i «Ordforrådet» og i de ulike kartleggingsverktøyene, kan derfor delvis forklares med at vi har brukt ulike tekstkorpora for å anslå bruksfrekvens. I «Ordforrådet» har vi også valgt å ha en tredeling mellom høyfrekvente, middels frekvente og lavfrekvente ord i stedet for en todeling i høy- og lavfrekvente ord, slik det er for eksempel i VOST.

Fonologisk naboskapstetthet

Et ords fonologiske naboskap består av alle andre ord som skiller seg lydlig fra det med bare én lyd (ett fonem). 'Katt' er dermed fonologisk nabo med blant annet 'skatt og 'hatt', men 'skatt' og 'hatt' er ikke naboer med hverandre. I norsk kan ord også skille seg lydlig fra hverandre ved varierende vokallengde (som i 'mate'/'matte') og varierende tonelag (som i 'bønder'/'bønner'). Ord som skiller seg fra hverandre kun ved enten varierende vokallengde eller varierende tonegang, er også fonologiske naboer.

Hvor mange fonologiske naboer et ord har, har vist seg å spille en rolle både når vi skal hente fram ord og si dem, og når vi skal oppfatte og forstå ord vi hører. Når man skal si et ord, er det generelt lettere å komme på eller hente fram et ord som har mange fonologiske naboer, mens tendensen er motsatt når vi skal oppfatte og forstå et ord. Effekten av fonologisk naboskap og billedlighet på produksjon og forståelse av ord i norsk er undersøkt av Ribu (2012).

I «Ordforrådet» er antallet fonologiske naboer basert på det digitale leksikonet *NorCompLex*¹ utviklet ved NTNU i Trondheim. Man kan søke på ord som har mange, noen eller få fonologiske naboer, delt inn på samme måte som for billedlighet, tilegnelsesalder og bruksfrekvens. Det vil si at ord i den øverste fjerdedelen – i praksis ord med 19 eller flere naboer – er kategorisert som ord med mange naboer, mens ord i den nederste fjerdedelen – i praksis ord med 2 eller færre naboer – er kategorisert som ord med få naboer. Resten av ordene hører til i mellomkategorien (noen naboer).

Lydstruktur

I «Ordforrådet» kan man søke på ord som begynner med en frikativ (en konsonantlyd der det er hørbar friksjon, for eksempel *s*-lyden, *skj*-lyden, *kj*-lyden og *f*-lyden), og ord som inneholder en konsonantgruppe. Begge disse egenskapene øker den lydlig kompleksiteten til ordet.

Ordlengde

I «Ordforrådet» kan man søke på ordlengde målt på tre forskjellige måter: i antall bokstaver, i antall lyder (fonemer) og i antall stavelser.

Ordstruktur

I «Ordforrådet» kan man søke på ord som er sammensatte, og ord som er avledet av andre ord. Et sammensatt ord er et ord som består av to eller flere ord som hver for seg kan fungere som selvstendige ord, for eksempel 'spisestue', 'hobbybonde', 'rødflekket' eller 'dagdrømme'. Et avledet ord er et ord som er dannet av et annet ord ved at en eller flere avledningselementer er lagt til rotordet, for eksempel 'venninne', 'kjærlighet', 'ufri', 'beslutte'. Slike avledningselementer (som -inne,

¹ NorCompLex – a computational lexical database of Norwegian. Hosted at and developed by Department of Language and Communication Studies (ISK), NTNU, Trondheim, Norway. Kontakt: lars.hellan@ntnu.no

-lig, -het, u-, be-) kan ikke fungere som selvstendige ord. Sammensetning og avledning øker den strukturelle kompleksiteten til ordet.

Ordklasse

Ordforrådet består som nevnt av 917 substantiver, 509 verb og 225 adjektiver. Man kan søke etter ord med bestemte egenskaper både på tvers av ordklassene og innenfor en bestemt ordklasse. Det finnes også noen søkbare egenskaper som bare gjelder ord innenfor én av ordklassene:

Spesielle egenskaper ved verb

«Ordforrådet» inneholder noen verb som består av flere ord, for eksempel 'trappe ned' og 'gjerde inn'. Man kan velge om søket skal inkludere eller ekskludere disse, eller man kan velge å søke etter bare slike verb. Søket kan også avgrenses etter transitivitet. Man kan søke etter verb som ikke tar objekt, verb som tar direkte objekt, verb som tar både indirekte og direkte objekt, eller verb som har vekslende transitivitet. Dessuten kan man søke etter verb som har et bestemt antall obligatoriske setningsledd knyttet til seg, for eksempel verb som bare tar subjekt, eller verb som tar subjekt og enten objekt eller adverbial.

Spesielle egenskaper ved substantiver

I «Ordforrådet» kan man velge å søke etter kun tellelige substantiver eller kun massesubstantiver. 'Hund' og 'spade' er eksempler på tellelige substantiver, mens 'smør' og 'kaffe' er massesubstantiver. En del substantiver brukes både om en masse ('vann') og om en tellelig enhet av denne massen ('et vann'). Man kan også søke etter substantiver som har et bestemt genus (grammatisk kjønn). Man kan søke etter hankjønnord ('bil'), hunkjønnord ('bok') og intetkjønnord ('hus'). Genusinndelingen i «Ordforrådet» er primært basert på *Bokmålsordboka* (www.bokmalsordboka.uio.no/).

HVA KAN MAN BRUKE «ORDFORRÅDET» TIL?

Målet med «Ordforrådet» er at denne leksikalske databasen skal bli et nyttig verktøy for praktikere – i denne sammenheng først og fremst logopedier – og for studenter og forskere innenfor logopedi, språkvitenskap og psykologi. En logoped kan for eksempel ønske å arbeide systematisk med visse typer ord, slik som høyfrekvente verb, lavbilledlige substantiver eller ord med mange

fonologiske naboer. Ved hjelp av «Ordforrådet» vil han eller hun da lettere kunne finne fram til relevante ord som kan være utgangspunkt for ulike typer oppgaver. Siden ordene i «Ordforrådet» for en stor del er hentet fra eksisterende kartleggingsverktøy (CDI, PALPA, VOST m.fl.), kan databasen også brukes til en grundigere analyse av resultater fra disse kartleggingsverktøyene.

Det er imidlertid viktig å huske på at mange av opplysningene i «Ordforrådet» gjenspeiler et gjennomsnitt av hva mange språkbrukere mener om de enkelte ordene. For eksempel er substantiver som 'afasi' og 'bevis' kategorisert som lavbilledlige, fordi gjennomsnittet av billedlighetsrangeringene for disse ordene var lavt. Og for de aller fleste språkbrukere stemmer nok denne kategoriseringen, men vi vil selvfølgelig alltid kunne møte enkeltindivider som har et tydeligere mentalt bilde av hva disse ordene viser til. For eksempel er det ikke utenkelig at en afasirammet person eller en som arbeider med afasirammede vil rangere ordet 'afasi' høyere på en billedlighetsskala enn gjennomsnittet i vår billedlighetsstudie. Tilsvarende kan det tenkes at ordet 'bevis' er mer billedlig for politifolk og jurister enn for andre. Dette gjelder også for subjektiv tilegnelsesalder og bruksfrekvens. I arbeid med enkeltindivider må man alltid ta den enkeltes erfaringsbakgrunn i betraktning.

Studenter og forskere kan også ha nytte av «Ordforrådet», for eksempel hvis målet er å finne ut om noen typer ord er vanskeligere å produsere eller forstå enn andre. Da trenger man et arsenal av ord å velge fra, og man trenger kunnskap om ulike egenskaper ved ordene.

«Ordforrådet» er nyskapende på flere måter. For det første er en stor del av opplysningene om egenskaper ved ordene ikke tidligere kartlagt for norsk. For det andre er det uvanlig å finne et så stort sett av opplysninger på ett brett – det er helt nytt for norsk, men også svært sjeldent i internasjonal sammenheng. Og endelig er databasen tilrettelagt for søk på en slik måte at den skal være lett tilgjengelig for så vel studenter og forskere som logopedier.²

² Vi vil gjerne takke studenter og kollegaer fra ILN og Statped sørøst, fagavdeling språk/tale som velvillig deltok i en brukerundersøkelse av «Ordforrådet» for nettstedet ble lansert. Vi fikk mange nyttige tilbakemeldinger og gjorde noen endringer som vi håper forbedret grensesnittet ytterligere.

Selv om vi stadig får mer kunnskap både om språk og språkbruk og om hva som er hensiktsmessig og mindre hensiktsmessig i logopedisk praksis, er det mange ubesvarte spørsmål. Vi håper at «Ordforrådet» kan bli en

nyttig ressurs som kan være med på å bidra til en videreutvikling av kunnskapen vår om det mentale leksikonet, om språkbruk og om tilegnelse og rehabilitering av språk.

Referanser

- Aitchison, J. (2003). *Words in the mind. An introduction to the mental lexicon*. 3rd edition. Malden, MA: Blackwell Publishing.
- Bastiaanse, R., Lind, M., Moen, I. & Simonsen, H.G. (2006). *Verb- og setningstesten (VOST)*. Oslo: Novus.
- Ecke, P. (2009). The tip-of-the-tongue phenomenon as a window on (bilingual) lexical retrieval. Pavlenko, A. (red.) *The bilingual mental lexicon. Interdisciplinary approaches*. Bristol: Multilingual Matters (185-208).
- Guevara, E. (2010). NoWaC: A large web-based corpus for Norwegian. In *Proceedings of the NAACL HLT 2010 6th web as corpus workshop*, Los Angeles, California, June 2010 (pp. 1–7). Association for Computational Linguistics. <http://www.rekstlab.uio.no/nowac/>
- Kay, J., Lesser, R., & Coltheart, M. (2009). *Psykolinguistisk kartlegging av språkprosessering hos afasirammede (PALPA)*. (Bredtvet kompetansesenter, Logopedtjenesten Helse-Bergen, Statped Vest & Øverby kompetansesenter, overs.). Oslo: Novus.
- Kristoffersen, K.E. & Simonsen, H.G. (2012). *Tidlig språkutvikling hos norske barn. MacArthur-Bates foreldre rapport for kommunikativ utvikling*. Oslo: Novus.
- Kristoffersen, K.E., Simonsen, H.G., Eiesland, E.A. & Henriksen, L.Y. (2012). Utvikling og variasjon i kommunikative ferdigheter hos barn som lærer norsk – en CDI-basert studie. *Norsk tidsskrift for logopedi*, 58, 1 (34–43).
- Korpjääko-Huuhka, A.-M. & Lind, M. (2012). The impact of aphasia on textual coherence: Evidence from two typologically different languages. *Journal of Interactional Research in Communication Disorders*, 3, 1 (47-70).
- Laine, M. & Martin, N. (2006). *Anomia. Theoretical and clinical aspects*. Hove: Psychology Press.
- Lind, M., Kristoffersen, K.E., Moen, I. & Simonsen, H.G. (2009). Semi-spontaneous oral text production: measurements in clinical practice. *Clinical Linguistics & Phonetics*, 23 (872-886).
- Paivio, A., Yuille, J.C. & Madigan, S.A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology*, 76, 1 (1-25).
- Ragnarsdóttir, H., Simonsen, H. G. & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children: an experimental study. *Journal of Child Language*, 26, (577–618).
- Ribu, I.S.B. (2012). An image is worth a thousand sounds? On imageability and phonological neighborhood density effects in speech processing. Masteroppgave, Institutt for lingvistiske og nordiske studier, UiO.
- Simonsen, H.G., Lind, M., Hansen, P., Holm, E. & Mevik, B.-H. (2013). Imageability of Norwegian nouns, verbs and adjectives in a cross-linguistic perspective. *Clinical Linguistics & Phonetics*. DOI:10.3109/02699206.2012.752527