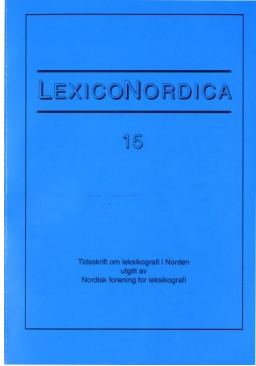


# LexicoNordica

Titel:	Analyse av søkelogger for bedre søkemuligheter i elektroniske ordbøker	
Forfatter:	Lars Nygaard og Ruth Vatvedt Fjeld	
Kilde:	LexicoNordica 15, 2008, s. 57-72	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive">http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive</a>	

© LexicoNordica og forfatterne

## Betingelser for bruk af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

## Søgbarhed

Artiklerne i de ældre LexicoNordica (1-16) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

*Lars Nygaard og Ruth Vatvedt Fjeld*

## **Analyse av søkelogger for bedre søkemuligheter i elektroniske ordbøker**

Analysis of search logs from on-line dictionaries can provide valuable insights into the behaviour, strategies and competence of dictionary users. This article presents a closer look at the search log for the electronic version of Bokmålsordboka from 11 months in 2007. We studied the logs by manually classifying random samples of unsuccessful queries. The data indicate that many users do not separate between dictionaries and other kinds of reference sources. Several of the unsuccessful queries would succeed if a spell cheque were added to the search interface. On the other hand, log files give valuable information to the lexicographers: identifying shortcomings in the lemma list and organising the information types in the dictionary.

### **1. Introduksjon**

Kunnskap om hvordan ordbøker brukes er nødvendig for å gjøre dem mer fullstendige og brukervennlige. Brukerundersøkelser er empirisk forskning som krever etterprøvbare metoder. De fleste kjente brukerundersøkelser hittil har benyttet metoder som introspeksjon, spørreskjemaer eller observert oppgaveløsning (jf. Svanepoel 2000, Svensén 2004). En gjennomgang av artikler om ordboksbruk i Euralex-rapporter fra 2000 til 2007 viser at de aller fleste undersøkelsene som er presentert der, er utført ved hjelp av en eller flere av disse metodene. Slike undersøkelser er beheftet med metodiske problemer, da det er de mentale prosessene forut for ordboksbruken som skal undersøkes, og det er vanskelig å kontrollere om de anvendte metodene er egnet for dette formålet. (Jf. ellers Tarp 2008 i dette nummer av *LexicoNordica*.)

En annen mangel ved tidligere brukerundersøkelser er at de stort sett har konsentrert seg om språklæringsordbøker (Svensén 2004:533), som regel tospråklige. Funn fra dem har mindre relevans for utarbeiding av enspråklige allmennordbøker.

Analyse av søkelogger gir en god mulighet til å kikke brukerne i kortene uten å måtte gå om indirekte informasjon fra brukerne selv, eksperimenter eller ytre observasjoner av indre prosesser. Selv om søkelogger selvsagt også innebærer en del metodiske mangler, har man i hvert fall der konkrete data å forholde seg til. I slike analyser kan man forholde seg til resultatene av brukernes tankevirksomhet i stedet for tanke-

virksomheten i seg selv, og slik komme nærmere hvilke problemer ordbøkene løser eller lar forbli uløst for brukerne.

Søkelogger er imidlertid lite undersøkt med tanke på ordbøkers bruksverdi, f.eks. ble det holdt en hel sesjon med brukerundersøkelser på Euralex-konferansen i Torino 2006, men ingen refererte fra logganalyser. I Barcelona 2008 var det bare én slik presentasjon (Mechura 2008), men i diskusjonen etter det innlegget sa flere at de arbeidet med det samme. Det tyder på at søkelogger vil bli et viktig redskap i utviklingen og utbedringen av moderne elektroniske ordbøker.

Av loggundersøkelser i leksikografien er en av de tidligste deSchryver & Joffe (2004), som har analysert loggfiler for tospråklige nettordbøker (mellom bantuspråket Sesotho sa Leboa og engelsk). De konkluderer med at søkelogger kan gi nyttig informasjon om lemmalakuner, men også om lakuner i informasjonstyper knyttet til lemmaene, dvs. både ytre og indre seleksjon.

Bergenholtz & Johnsen (2007) gir en oversikt over forskjellige tilnærminger til hvordan logganalyser kan anvendes i ordbokskritisk sammenheng og i metaleksikografisk forskning. Gjennom en sammenlikning av søkelogger av fire svært forskjellige nettordbøker hevder de at funksjonsorienterte loggfilanalyser gir en sannere bilde av de undersøkte ordbøkene enn rene frekvensundersøkelser.

En tidlig nordisk logganalyse er gjort av Hult (2008). Hun har undersøkt en dags bruk av den svenske Lexin-ordboka *Svenska ord*, og finner at hele 41 % av søkene den dagen ikke ga det svaret brukerne søkte. Det er mye høyere nulltreff enn i andre logganalyser. Lexin-ordbøkene er riktignok spesielt laget for andrespråksinnlærere, og det kan hende at det er grunnen til den høye feilsvarprosenten. Men Lexin-ordbøkene har både mulighet til fritekstsøk og automatisk henvisning til liknende ord ved mislykket søk, så den høye feilsvarprosenten er overraskende. Hult peker på at denne ordboken har vært nettilgjengelig mye lenger enn andre ordbøker som har vært underkastet logganalyse, og det kan være en delforklaring. Det mest interessante i studien er en oppfølgingsanalyse av brukernes søkestrategier, der hun finner at de fleste som ikke får svar på sine søk med en gang, forsøker seg med flere stavemåter, eventuelt med søk på deler av ordet eller avledninger av det. Gjentatte søk fører da ofte til positive resultat.

I denne artikkelen vil vi bruke loggdata fra søk i Bokmålsordboka, en elektronisk norsk enspråklig allmennordbok, gratis tilgjengelig på Internett (<http://www.dokpro.uio.no/ordboksoek.html>). Her kan man søke på enten Bokmålsordboka eller Nynorskordboka, eller på begge bøkene samtidig. I norsk leksikografisk forskning er analyse av søke-

logger en ny type brukerdatabaser, bortsett fra den sammenliknende studien til Bergenholtz og Johnsen nevnt over, der også Bokmålsordboka er med. Hver måned gjør publikum om lag en million søk i nettversjonen av denne populære ordboka. Materialet for undersøkelsen er hentet fra søkelogg for perioden 1. januar–1. november i 2007.

## 2. Brukernes konsultasjonsferdigheter

Brukerundersøkelser kan rettes mot brukernes ferdigheter eller mot selve ordbokens egenskaper. En søkelogg egner seg spesielt godt for å se nærmere på konsultasjonsferdighetene, da den røper hvilke handlinger ordboksbrukeren faktisk gjør for å finne den ønskede informasjonen. Analyse av søkeloggen for Bokmålsordboka gjennom ti måneder, ga resultatet i tabell 1 nedenfor, som sier noe om hvor godt denne ordboka hjelper sine brukere.

TABELL 1. Antall søk i Internettversjonen av Bokmålsordboka jan.–nov. 2007

Søkeresultat	Antall søk
Treff	5.812.997
Treff (bøyde former)	1.128.945
Ingen treff	3.110.773
Antall søk totalt	10.051.300

Tabell 1 viser at det i undersøkelsesperioden ble gjort over ti millioner søk, og at vel halvparten av dem ga det svaret brukerne sannsynligvis lette etter. Ordboka inneholder også en funksjon for søk på bøyde former, og søkeloggen viser at det i tillegg ble treff på vel en million bøyde former, der søkeren automatisk ble vist fra en bøyd form til lemmaform. Men til tross for denne hjelpefunksjonen er det fortsatt nesten en tredjedel av søkene som ikke ga noen form for treff. I denne artikkelen vil se nærmere på årsaker til at søk i den elektroniske Bokmålsordboka mislykkes for brukerne. Det paradoksale i å bruke en rettskrivingsordbok er at man må kjenne stavemåten for å finne stavemåten. I papirordbøker kan man muligens lete litt i nærheten av det stedet der søkeordet er forventet å stå, og med flaks finne ordet selv om man bare omtrentlig kjenner stavemåten. Ved søk i elektroniske ordbøker er det vanskeligere, selv om det er mulig å søke på deler av ord ved å trunkere med %-tegn, og dermed få svar på søk selv om man ikke kjenner hele ordet. Dessverre er det forholdsvis få som bruker denne muligheten i sin fulle bredde.

Loggfilene sier noe om hva brukerne faktisk gjør når de ikke får direkte svar, dvs. kjenner svaret før de søker.

### 3. Loggfiler som datakilde

Loggfiler er problematiske som datakilde av flere grunner. For det første inneholder de materiale som gjør at adgangen bør kontrolleres nøye. I prinsippet kan man hente ut opplysninger om privatlivet til enkeltbrukere fra dataene.

For det andre kan statistikken påvirkes i betydelig grad fra dataprogrammer som automatisk laster ned all informasjon på nettet. Disse programmene utfører søk på alle lenker de kan få tak i med uregelmessige, men ofte små mellomrom, og kan gi uheldige utslag på søkestatistikkene, spesielt hvis ordboksgrensesnittet inneholder interne lenker. Det er teknisk mulig å filtrere ut disse søkene, men det er ikke blitt gjort med materialet for denne artikkelen.

For det tredje er brukeridentitetsnummer basert på IP-adresser, noe som ikke gir sikker identifikasjon av unike brukere. Slike adresser kan deles mellom flere brukere, og en enkelt bruker kan over tid bruke flere adresser. Utviklere av ordboksgrensesnitt kan bruke ulike teknikker for å gjøre denne identifiseringen mer presis, men slike foreligger ikke for *Bokmålsordboka*. Vi undersøker i denne omgangen ikke forskjellige brukergruppers atferd, men forholder oss bare til de forskjellige typene mislykkede søk. Materialet for undersøkelsen er en loggfil fra perioden januar-november 2007, dvs. 11 måneders søkelogg. Ut fra denne loggen har vi trukket ut 500 tilfeldige søk som vi har analysert nærmere. Tabell 2 viser et utsnitt fra en slik loggfil i *Bokmålsordboka* som en illustrasjon av rådata for undersøkelsen.

TABELL 2. Tilfeldig utdrag fra den analyserte loggfile fra Bokmålsordboka

Søkestreng	Ant. treff	Trefftype	Brukerid
gle% <sup>1</sup>	463	ordliste <sup>2</sup>	1454543
nyttår	1	ordbok	1054387
imponere	8	ordbok	1461894
innsigelse	1	ordbok	1443708
nåme	0	–	1454543
gjemsel	1	ordbok	914397
skade	171	ordbok	1461894
ræl	0	–	839758
ræle	0	–	839758
vrøvl	1	ordbok	839758
irrgang	1	ordbok	1455819
lyrisk	1	ordbok	1239430
hesan	0	–	1462164
opdal ørnulf	0	–	1462169
eutym	0	–	818249
sebu	1	ordbok	1455819
orrekratt	0	–	1459817
orekratt	1	ordbok	1459817
sjønn	0	–	1202138
sedat	1	ordbok	1459817
samarbeid	1	ordbok	1462165
skjøn	0	–	1202138
delja	6	ordliste	1462138
vette	1	ordbok	1459817
skjønn	2	ordbok	1202138
laps	1	ordbok	1459817
tusja	6	ordliste	1014109
levende	98	ordbok	1461894
levende	7	ordliste	1461894
saran	0	–	1462153

Tabell 2 viser at 9 av 30 tilfeldige søk ikke gir noe treff, altså er hele 30 % av søkene mislykket. Av disse er det 6 som ikke ga treff fordi de ikke er lemmaer i Bokmålsordboka, mens 3 er feilstavet og dermed ikke leder brukerne til den informasjonen de trenger og burde kunne ha fått.

---

<sup>1</sup> %-tegn står for trunkert søk.

<sup>2</sup> Trefftypen ”ordliste” blir brukt der brukeren søker etter et ord i en bøydd form, altså noe annet enn den tradisjonelle oppslagsformen.

#### 4. Analyse av dataene

Loggfiler gjør det mulig å undersøke årsakene til at ordsøkene mislykkes. De mest frekvente søkene som ikke ga treff i den undersøkte perioden, er presentert i tabell 3. Høyre kolonne gir antall søk i perioden vi tok for oss.

TABELL 3. De 20 hyppigste mislykkede søk i materialet

<i>Søkeord:</i>	<i>antall søk:</i>
desverre	5958
etterhvert	3562
beijing	3448
nyskjerrig	2721
forøvrig	2674
tilsammen	2590
hvertfall	2539
blandt	2473
ihvertfall	2409
vere	2230
internett	2179
0	2094
værre	1808
interessant	1737
nysjerrig	1669
alikevel	1665
interessert	1605
anderlede%	1488
proaktiv	1446
idag	1444

Feilsøkene er av forskjellige typer, og kan systematiseres som i tabell 4:

TABELL 4. Feiltyper i materialet som gir mislykkede søk (Prosenttallene er tatt med for sammenlikningens skyld, materialet er altfor lite til at tallene har generell overføringsverdi.)

Stavefeil	11	55 %
Sær-/sammenskiving	6	30 %
Proprier	1	5 %
Lakuner	1	5 %
Søk på tall og tegn	1	5 %

En ren registrering av de hyppigste feilsøkene gir ikke uttømmende beskrivelse av årsakene til manglende treff. Ulike typer feilsøk kan ha ulik frekvensdistribusjon, og vil ikke nødvendigvis vises blant de mest vanlige over tid. For enkelt å få et noe tryggere materiale hentet vi derfor ut

500 tilfeldig valgte linjer fra feilsøkene i loggfilen og sorterte dem manuelt etter feiltyper. Det ga et noe annerledes bilde, jf. tabell 5:

TABELL 5. Analyse av 500 tilfeldig valgte feilsøk i materialet

Stavefeil	46 %
Lakuner	15 %
Feil språk el. målform	12 %
Sær-/sammenskriving	11 %
Proprier	6 %

Denne metoden viser imidlertid også at stavefeil er den klart vanligste årsaken til manglende treff i ordboka, særlig om sammenskriving av flerordsuttrykk regnes som stavefeil, men også lemmalakuner, søk på proprier eller søk på ord fra fremmedspråk utgjør en betydelig del av de mislykkede søkene, selv om de internt er så varierte at de ikke vises på oversikten over de mest vanlige. Vi skal se i mer detalj på feilsøkene innen de ulike kategoriene.

#### 4.1. Stavefeil

Tabell 2 og 3 antyder at de vanligste stavefeilene i materialet er enkel i stedet for dobbelt konsonant, sammenskriving av flerordsuttrykk og realisering av åpen e-lyd med 'e' eller 'æ'. Disse resultatene burde også være interessante i et normeringsperspektiv. Det høyfrekvente adverbet *dessverre* topper listene, både over feilsøk og søk med treff, og er det overlegent mest søkte ordet i hele ordboka. Det kan dermed kanskje kåres til det vanskeligste ordet i norsk bokmål. Når det er nesten 6000 søk på feilformen *desverre*, kan det komme av at trykktapet på første stavelse gjør brukerne usikre på stavemåten. Vi vet heller ikke hvor mange av dem som først har søkt på feilformen, som etterpå søker på den rette formen. Det ville sannsynligvis kommet fram ved en næranalyse av brukeridentitetene. Gode ordbøker bør ved slike ord i det minste gi brukerne hjelp ved å henvise til rett ortografi med spørsmål som "mente du *dessverre*?", slik mange ordbøker har i dag, f.eks. slik Hult refererer for Lexin-ordbøkene, jf. ovenfor.

#### 4.2. Lakuner

Hele 15 prosent av de mislykkede søkene kunne i prinsippet gitt tilslag i ordboka, siden de er korrekt stavede norske ord. Imidlertid er 5 av de 15



prosentene vanlige, ikke-leksikaliserte sammensetninger, 2 prosent er regelrette avledninger, og mange av de andre er sjeldne ord, terminologi, slang- og dialektord. Disse feilsøkene dokumenterer at en god elektronisk ordbok gjerne kan inneholde en funksjon som genererer avledninger, samtidig som den bør gi veiledning om hvilke avledninger som er vanlige og hvilke som er sjeldne.

Lakuner defineres her som en ufullstendighet i lemmalista. Det bør først og fremst være lånord og nylaginger som ikke er registrert ennå. Det kan selvsagt også være gamle ord som er uteglemt fordi ordboksbasisen har vært for dårlig, eller fagord som ikke hører hjemme i en allmennordbok.

Videre kan avledninger og sammensetninger betraktes som lakuner, selv om de er bevisst utelatt av redaktørene, fordi de er dannet i samsvar med vanlige orddanningsregler i språket. Det gjelder særlig nominaliseringer på *-ing*. Søkeloggen avslører tydelig forskjellen mellom redaktørens oppfatning av hva som bør lemmatiseres, og brukernes oppfatning av hva de kan finne i en allmennordbok. Kan hende kan begge oppfatningene justeres noe for databaserte ordbøker der plasshensyn ikke lenger veier så tungt som i papirbaserte ordbøker.

Tabell 1 viser at mange søker etter det relativt nye ordet *proaktiv*, som ennå ikke er inkludert i ordboka. Slike lakuner er det ikke mulig å unngå i en ordbok, lånord kommer inn og brer seg i et språk raskere enn ordbøker vanligvis revideres. Og det er naturlig at mange søker på et ord som er nylig kommet i allmenn bruk. Det viser at elektroniske ordbøker krever oppdatering oftere enn papirordbøker, og at det er viktig med nyordsfinning ved hjelp av mer moderne metoder enn manuell registrering i publiserte tekster. Moderne søkesystemer gjør det mulig å lokalisere nyord semiautomatisk, og det bør bli en fast oppdatering i ordbøker på Internett. Og selvsagt vil analyse av søkelogger være et naturlig hjelpemiddel.

Mange søker på tekniske termer, spesielt fra domenet medisin. Det viser at brukerne ikke har nok kunnskap om skillet mellom ordbøker og andre oppslagsverker, f.eks. mellom allmennordbok og en fagordbok. Det er kanskje ikke så rart, siden allmennordbøker inneholder termer som er på vei inn i allmennspråket. Siden elektroniske ordbøker ikke har plassbegrensninger, kan man argumentere for at mer terminologi bør inkluderes i allmennordbøker. I hvert fall representerer disse feilsøkene en verdifull tilbakemelding ved forbedring av lemmautvalget i ordboka.

Man kan også regne ikke-lemmatiserte sublemmaer som lakuner, dvs. avledninger og sammensetninger, og de analyserte loggfilene tyder på at man i større grad bør ta med avledede former. For eksempel søker

mange på nominalavledningen *redigering* uten å få noen treff. Det er lite sannsynlig at alle tenker på at de i stedet bør søke på verbet *redigere*.

Vi har laget en egen analyse av alle mislykkede søk for å identifisere ord som sannsynligvis er ekte lakuner (dvs. som *ikke* er stavefeil, proprier etc.). Denne analysen er ikke perfekt, men den utgjør likevel et utgangspunkt for å finne ord som bør vurderes inkludert i ordboka. Høyre kolonne angir antall søk.

TABELL 6. Automatisk generert liste over tilsynelatende lakuner i ordboka

Søkeord:	antall søk:
avspasering	466
utagerende	463
curriculum	462
reliabilitet	446
selvinnsikt	432
komposisjonsprinsipper	430
kredibilitet	418
hvalp	390
ifm	389
imorgen	386
semi	381
troverdighet	373
forhåndsregel	373
nemmelig	363
avveining	356
inkrementell	351
inntresert	344
prevalen%	339
marginalisering	338
granskning	331
forutsigbarhet	331
prokrastinere	329
slåssing	327
prinsippielt	326
nysjereg	325
notoritet	325
iløpet	324
pseudo	320
suicidal	317
orddeling	316
sjenerelt	313
stusselig	311
imidlertidig	311
alumni	302

Tabell 6 inneholder forskjellige typer feilsøk der proprier og antatte feilstavinger er fjernet automatisk. Ved en nærmere analyse ser vi at mange av dem er avledninger som kan genereres ut fra grunnord i ordboka.

### *Regelrette avledninger*

Vanligvis lemmatiseres ikke regelrette avledninger om de ikke også har en leksikalisert betydning eller spesielle bruksrestriksjoner. Men søkeloggen viser at mange søker på slike avledninger, og vi vet ikke om de selv klarer å gå tilbake til grunnformen, så de finner den informasjonen de trenger. En hjelpefunksjon som antydnet nedenfor, ville vært av stor nytte:

avspasering – se *avspasere*  
 utagerende – se *utagere*  
 troverdighet – se *troverdigg*  
 avveining – se *avveie*  
 inkrementell – se *inkrement*  
 marginalisering – se *marginalisere*  
 forutsigbarhet – se *forutsigbar*  
 slåssing – se *slåss*

### *Selvforklarende sammensetninger*

Likeledes er det en del sammensetninger som ikke er leksikaliserte og dermed ikke forsvares en lemmaplass i vanlige ordbøker:

selvinnsikt  
 komposisjonsprinsipper  
 orddeling

Det betyr at en komposisjonsfunksjon eller sammensetningsgenerator kunne kanskje ha hjulpet brukerne til å finne det de leter etter. Selv om en sammensetning er selvforklarende, kan den gjerne gi treff med henvisning til definisjon av enkeltdelene av ordet. Særlig for andrespråksinnlærere vil en slik funksjon være nyttig. Men slike søkeord kan også tyde på at brukerne ser på ordboka som en encyklopedi der de vil finne saksforklaringer.

### *Feilstaving/feil orddanning*

Selv om det var forsøkt å rense ut feilstavinger automatisk, er rettskrivingen i norsk så vanskelig at det kom med en del feilstavinger likevel, i tillegg til såkalte ”vanskapninger” i språket, forvanskninger, folkeety-mologiske nydanninger og andre feil. Mange av dem er så vanlige at det

burde vært en hjelpefunksjon ved dem med henvisning til rett ord, som ”mente du *forholdsregel*” når noen søker på ordet *forhåndsregel*.

hvalp (jf. *valp*)  
 ifm (jf. *i forbindelse med*)  
 forhåndsregel (jf. *forholdsregel*)  
 nemmelig (jf. *nemlig*)  
 inntresert (jf. *interessert*)  
 granskning (jf. *gransking*)  
 prinsippielt (jf. *prinsipielt*)  
 nysjerrig (jf. *nysgjerrig*)  
 pseudo (jf. *pseudo-*)  
 sjenerelt (jf. *generelt*)  
 stusselig (jf. *stusslig*)  
 imidlertid (jf. *midlertidig*)

### *Sammenskriving*

Som i analysene for øvrig viser det seg at sammenskriving av flerordsuttrykk er vanlig. Men i disse tilfellene kan det hende at det rett og slett også er tastefeil, som i ”i løpet”, siden det siste leddet mangler, frasen er ”i løpet av”.

imorgen  
 iløpet

### *Nyord/ekte lakuner*

Når disse feiltypene er sortert fra, sitter vi igjen med en rest som leksikografisk sett kan klassifiseres som ekte lakuner. Det vil si nye ord som ennå ikke er blitt registrert og redigert inn i ordboka, både lånord og nylaginger:

curriculum  
 reliabilitet  
 kredibilitet  
 semi  
 prevalen%  
 prokrastinere  
 notoritet  
 suicidal  
 alumni

Dette er ord som må vurderes redaksjonelt og eventuelt tas opp som lemma. En søkelogg er dermed også en verdifull informasjonskilde til å oppdage nye ord. Vi har rett nok en søketjeneste på Internettversjonen der man kan melde fra om man ikke får tilslag for det man søker etter, kalt Ordvakta. Det kommer jevnt og tett inn forslag der, både ord som er ekte lakuner i Bokmålsordboka, men også all verdens rariteter av dia-

lektord, familieord, barnespråk og språklige misforståelser. Alle henvendelser blir registrert i en database, men det er langt mellom forslag som blir tatt inn som lemmaer. Søkeloggen er et mer pålitelig redskap, antall søk viser jo også hvor mange som har søkt på ordet, og altså hvor utbredt det er blitt.

### **4.3. Feil språk el. målform**

Tabell 5 viser at hele 12 % av de mislykkede søkene er på feil språk eller målform, og av disse er 8 % søk på feil målform, altså søk på nynorskord i Bokmålsordboka. Dette tallet er nok så høyt fordi en og samme nettside gir tilgang til begge ordbøkene, og knappene for de to ordbøkene ligger tett inntil hverandre og ser helt like ut, så det er fort gjort å trykke på feil knapp. Dette kan forklare mesteparten av de mislykkede søkene, men samtidig kan det være et signal om at kunnskapen om grensene mellom målformene ikke er kjent for alle. Nettopp da er en samordning av ordbøkene god hjelp, og en tredje knapp gjør at man kan søke i begge ordbøkene samtidig.

Et problem for denne statistikken er imidlertid at disse fellessøkene registreres som søk på feil målform for den ordboka som ikke gir treff. Tallene her er dermed bare indikasjoner på manglende kunnskap om målformene hos brukerne.

I tillegg er det en rest på 3 % som sannsynligvis er søk på ord i fremmedspråklige tekster, mest engelsk, svensk og dansk, som vist i tabell 6 nedenfor.

TABELL 8. Oversikt over søk på fremmedspråklige ord i norsk

dansk:	kartofel
svensk:	elitidrott kudde resenær
engelsk:	sealing allegiance% reciprocal recruitment prurigo sepulture past the lower part of your face wieu nope cheer%
spansk:	tú me tengo acabado

Av 16 søk på utenlandske ord er 11 sannsynligvis engelske ord, 3 svenske og ett fra dansk og ett fra spansk. Det er rimelig at de aller fleste er fra engelsk, da det er dette fremmedspråket nordmenn flest mest må forholde seg til. Tabellen viser også at en del brukere ikke skiller klart mellom enspråklige og flerspråklige ordbøker. Men noen av disse feilsøkene kan også vise nye lånord i norsk eller utenlandsk slang på vei inn i norsk, som *nope*. I tillegg ser vi at mange av de utenlandske ordene er stavet feil.

Det er enkelt å hjelpe disse brukerne ved å ta med lenker til ordlister for noen andre språk og gi brukere beskjed om at ”Dette ordet ser ut til å være på svensk; du kan finne informasjon om svenske ord fra følgende kilder ...”.

#### **4.4. Sær-/sammenskriving**

Tabell 5 viser at 11 % av de analyserte feilsøkene kan kategoriseres som sammenskriving av flerordsuttrykk. Den undersøkte søkeloggen viser at det ikke er selvsagt for alle hvor et ords grenser går. Norsk er et språk med mange sammensetninger, og analysene viser at det er flere søk i sammenskrevne flerordsuttrykk enn omvendt, altså at man særskriver sammenskrevne former. Det er særlig adverbialfraser som består av preposisjonsuttrykk mange oppfatter som så leksikaliserte at de sammenskrives: *imorgen*, *etterhvert*, *tilsammen*. En medvirkende årsak til det kan være at eldre norsk – som dansk – hadde mer sammenskriving enn det som har vært Språkrådets normeringslinje i etterkrigstiden. Den ortografiske standarden riksmål har beholdt flere sammenskrevne fraser, og siden landets største avis Aftenposten har fulgt riksmål, er mange vant til flere sammenskrivinger enn den offisielle standarden tillater. Dermed blir nok folk flest lett usikre og ønsker å slå det opp.

#### **4.5. Proprier**

Proprier utgjør 6 % av de mislykkede søkene. Et utvalg av de vanligste er vist i tabell 8 nedenfor. Feilsøkene viser at mange brukere ikke har klart for seg skillet mellom ordbok og leksikon. Dette skillet kan imidlertid være vanskelig å trekke selv for leksikografer, og en del ordbøker har også med en del navn, særlig stedsnavn.

TABELL 9. *Proprier som søkeord i det undersøkte materialet.*

Stedsnavn:	aserbajdsjan beijing INDIA rælingen skandinavia
Adresser:	dalaneveien 80 dronningensgate granveien 8, trondheim Lillebergveien Oslo Tvetenveien 5
Personnavn:	are JONAS lenin nikolai
Sykdom:	Stevens-Johnson Syndrome
Firma-/varenavn:	Brødmat octavianu% revitalisme
Institusjonssnavn:	TAS

Dette resultatet kan tilsi at proprier tas med i ordbøkene, spesielt siden nettpublisering gjør at plassbegrensninger ikke lenger er relevant. Mange søker på navn som er vanskelig å stave (Beijing, Azerbajdsjan), og det virker dermed som de er ute etter ortografisk snarere enn ensyklopedisk informasjon. En annen løsning kan være å vise søkerne til en encyklopedi, til generelle navneordbøker eller Geografilista utgitt av Språkrådet, der de vil finne mer informasjon om navnet og stavemåten. En tredje mulighet kunne være å lage ordbaser der alle typer ord er med, men kategorisert som allmennord, termer eller navn i en type allordbok. Man kunne tenke seg at søk på navn er et tegn på deproprialisering, men ingen av de ettersøkte navnene i dette materialet er eksempler på det.

## 5. Konklusjon

Vi har i denne artikkelen presentert en metode for analyse av loggfiler, samt resultater fra bruk av Bokmålsordboka ved å ta spesielt for oss tilfeller der brukerne ikke får tilslag på sine søk. I mange tilfeller antar vi at brukeren får en viss informasjon fra disse mislykkede søkene, siden manglende svar enten betyr at ordet ikke eksisterer eller er stavet feil. I andre tilfeller mislykkes søkene selv om søkeordet er korrekt stavet, fordi det er en ikke-leksikalisert sammensetning eller avledning.

Resultatene tyder på at brukerne har behov for mer veiledning i bruk av referanseverk, og at slike henvisninger gjerne kunne inkorporeres i selve ordboka. Vi har forsøkt å antyde metoder for å gi brukerne mer presis og relevant informasjon der søkene mislykkes. Vi vet imidlertid ikke om feilsøk skjer fordi brukeren er overbevist om at den formen de søker på, er rett. Det kan like gjerne være at de søker en bekreftelse på at en stavemåte er feil, som en overbevisning om at sammenskriving er rett. Hva som egentlig er ordboksbrukerens motivasjon, kommer ikke fram i en søkelogg og bør følges opp med andre typer observasjon av ordboksbrukerne, noe som også er påpekt av Almind (2008), som gjør en analyse av loggfilene til en dansk fraseologisk ordbok. Konklusjonen ser ut til å være at det bør legges til en stavekontrollfunksjon til grensesnittet som henviser brukeren til den korrekte stavemåten. Denne funksjonen bør tilpasses de typene av stavefeil nordmenn ser ut til å gjøre, noe man greit kan lese ut av søkeloggen.

Dessuten viser vi at analyse av loggfiler kan gi verdifulle data for identifisering av lakuner i lemmalista.

## Litteratur

- Almind, Richard 2008: Søgemønstre i logfiler. I: *LexicoNordica* 15, 33–55.
- Bergenholtz, Henning & Mia Johnsen 2007: Log Files Can and Should Be Prepared for a Functionalistic Approach. I: *Lexikos* 17, 1–21.
- de Schryver, Gilles-Maurice & David Joffe 2004: On How Electronic Dictionaries are Really Used. I: Geoffrey, W & S. Vessier (eds.), *Proceedings of the eleventh EURALEX International Congress*. Lorient, France, July 6–10, 2004. Lorient: UBS, 187–196.
- Hult, Anki 2008: Från ord till handling – en studie i ordboksanvändning på nätet. I *Nordiske studier i leksikografi nr. 10*. Reykjavík (in press).
- Mechura, Michal Boleslav 2008: Giving Them What They Want: “Search Strategies for Electronic Dictionaries”. I: Bernal E & J. De Cesaris (eds.), *Proceedings of the XIII Euralex International Congress*. Barcelona, 1295–1299.
- Svanepoel, Piet 2000: Providing lexicographic support for SL vocabulary acquisition: What kind, under what conditions, for whom? I: Heid, V. & al (eds.): *Proceedings of the Ninth Euralex International Congress*. Stuttgart, 403–417.



- Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Nordstedts Akademiska Förlag.
- Tarp, Sven 2008: Kan brugerundersøgelser overhovedet afdække brugernes leksikografiske behov? I: *LexicoNordica* 15, 5–32.

Ruth Vatvedt Fjeld  
professor  
Universitetet i Oslo  
r.e.v.fjeld@iln.uio.no

Lars Nygaard  
cand.philol.  
Kaldera språkteknologi,  
ln@kaldera.no