

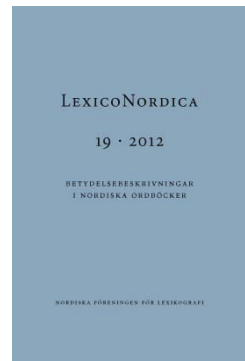
LexicoNordica

Titel: Ordbogsartikler, som ingen læser

Forfatter: Henning Bergenholtz & Bjarni Norddahl

Kilde: LexicoNordica 19, 2012, s. 207-223

URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

Betingelser for brug 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.

Ordbogsartikler, som ingen læser

Henning Bergenholtz & Bjarni Norddahl

Lemma lacunas in dictionaries are a traditional focus area for lexicographers, but the opposite problem, which we choose to call “lemma flooding”, has received very little attention. In this article we show that the study of this flooding is very relevant in order to save lexicographers spending thousand of hours producing dictionary entries which nobody reads. We examine nine possible reasons why a given word might not be of interest to users and could consequently be ignored in order to avoid lemma flooding. We demonstrate that while it is not possible to completely avoid lemma flooding, it can be reduced by implementing a relative simple rule.

1. Overflødige lemmaer

Når en ordbogsbruger slår op i en ordbog og ikke finder det søgte ord, må han eller hun prøve at søge i en anden ordbog. Når noget sådant sker flere gange i træk, vil vedkommende muligvis ikke bruge ordbogen mere. Der er for mange lemmahuller. En sådan erfaring kan ligge bag mange ordbogsanmelderes fokusering på lemmasektion og specielt på lemmahuller, hvor der nogle gange anføres lange lister over manglende ord.

Når anmelderen også anfører overflødige ord, ord, som efter anmelderens mening er mindre vigtige end nogle af de anførte lemmahuller, kan et sådant argument være relevant for en trykt ordbog. Men i en elektronisk ordbog vil et sådant argument være forkert af mindst to grunde:

1. Anmelderen kan ikke med sikkerhed vide, om nogle brugere ville søge på netop et af de “overflødige” ord.

2. De "overflødige" ord generer ingen brugere, selv om de aldrig slår op på dem. De tager heller ikke plads væk fra andre ord, for der er plads til alle de ord, leksikografen har tid og kræfter til at bearbejde leksikografisk.

Noget helt andet er så et leksikografisk tidsforbrugsargument: Overflødige ord tager lige så lang tid at beskrive som relevante ord, ordbogsbrugeren slår op på. Det er temaet for denne artikel, som modsat Bergenholtz (1989, 2003 og 2005) ikke fokuserer generelt på lemmaselektion, men på, om nogle ord i ordbøger er overflødige, fordi ingen ordbogsbrugere nogen sinde læser dem.

Noget sådant kunne man i princippet undersøge i trykte ordbøger, hvis man lod et og samme eksemplar af en ordbog blive brugt af alle brugere på et bibliotek med den opfordring, at der skulle laves et mærke ved hver artikel, brugerne har slået op på. Sådanne undersøgelser kender vi dog ikke¹. Vi kender heller ikke til tilsvarende undersøgelser af logfiler; i foreliggende bidrag om logfil-analyser fokuseres udelukkende på de ord, der søges efter, fx Ling et al. (2002), de Schryver/Joffe/Joffe/Hillewaert (2006), Bergenholtz/Johnsen (2007). Men sådanne logfilanalyser er mulige. Vi vil her præsentere nogle data for logfiler for Den Danske Netordbog i tiden 1.1.2009 til 13.10.2011. Den Danske Netordbog er en stor almensproglig danskordbog med mere end 110.000 forskellige opslagsord, som primært er tænkt som hjælpeværktøj i forbindelse med tvivl og usikkerhed, når man skriver en dansk tekst. I disse 31½ måned eller 1093 dage var der i alt 17.893.973

1 Den eneste bemærkning til temaet er en yderst negativ indstilling til, at der foreligger nogen form for system, som kan beskrive, hvilke ordtyper der ikke søges på: "Furthermore, it is not possible to discern a distinct pattern on the basis of this examination, e.g. that certain types of words, such as semantic or orthographic variants, are never requested. A systematic description of the requested words compared with the non-requested words is thus not possible, and it remains unclear whether such an investigation would be of practical use to lexicographers." (Bergenholtz/Johnsen 2005:139f.)

opslag i ordbogen efter ord, der blev fundet. Ikke-fundne ord er ikke medtalt. Da ordbogen har en træfferkvote på mellem 70% og 71%, kan man se, at der har været en del flere opslag end de anførte knap 18 millioner, dvs. opslag med fejlskrivninger og efter ord, ordbogen ikke har, altså efter lemmahuller.

De knap 19 millioner enkelte søgninger i ordbogen udgør et så stort materiale, at det må have en ret stor empirisk tyngde, især sammenlignet med tidligere logfilanalyser, hvor der var mellem 21.337 opslag (Schryver/Joffe 2004) og 1.016.960 (Bergenholtz/Johnsen 2005).

Fra andre ikke-offentliggjorte logfil-analyser ved vi, at mellem 17% og 70% af artiklerne aldrig er blevet slået op. Vi ved også, at 80% af brugerne (eller mere nøjagtigt de enkelte IP-adresser) over en længere tidsperiode kun har søgt på 1, 2 eller 3 artikler. Det sidste fænomen er ikke temaet her. Her drejer det sig om andelen af ordbogsartikler, som ikke er blevet læst en eneste gang.

Vi vil vente til slutningen af artiklen med at røbe tallene for ikke-brugte søgeord. Med søgeord menes, at søgeprogrammet forudsætter, at man kan søge på lemmaet, dvs. grundformen, men også på de bøjningsformer, ordbogen angiver i sin grammatiske angivelse. Det vil for de fleste adjektiver være tre, fx *rød*, *rødt*, *røde*; nogle gange dog flere ved uregelmæssig komparation: *stor*, *stort*, *store*, *større*, *størst*. For de fleste verber vil det være fire fleksionsformer, fx *fnise*, *fniser*, *fniste*, *fnist*; nogle gange dog flere, fx ved *fnise*, hvor der alternativt anføres former som *fnisede*, *fnes*, *fneset*. Det vil for de fleste substantiver være fire former, fx *tavle*, *tavlen*, *tavler*, *tavlerne*; dog vil der ved fleksionsvarianter kunne være flere, fx ved *virus* ikke bare flertal *vira*, men også *virusser*. Grammatiske angivelser, der ikke fremgår af ordbogen, er ikke søgbare, fx genitivformer som *tavles*, *tavlens* mfl.

Det må her tilføjes, at brugerne ikke, som nogle måske kunne mene, udelukkende søger på grundformen. Vi har netop indført denne mulighed, fordi mange brugere søgte på bøjede former.

ICKE-TEMATISKA BIDRAG

Sammenlign hertil følgende liste med ret store søgetal på andet end lemmaet:

muligheder	3.897 opslag
mangler	3.765 opslag
omhandler	3.758 opslag
ordbogen	3.339 opslag
tænker	3.334 opslag
features	2.676 opslag
udgifter	2.294 opslag
svælger	2.483 opslag
generelt	2.463 opslag
kompetencer	2.358 opslag

Alt i alt har eller mere nøjagtigt havde databasen i 2011 i alt 365.911 potentielle søgeord i databasen (ordbogen udvides løbende, så tallet er en del større i marts 2012). Af disse mere end 360.000 søgeord er de 314.312 ikke blevet brugt en eneste gang. Det betyder, at der selv med 18.000.000 opslag kun er søgt på 51.599 lemmaer og fleksionsformer, dvs. 14,2% af alle mulige opslagsord. Der betyder ikke, at kun 14% af artiklerne er blevet læst, men med mere end 110.000 artikler i databasen kan man allerede nu se, at antallet af ikke-læste artikler må være mere end 50%. Det betyder, at mange leksikografer på sin vis har spildt tusindvis af timer på at skrive artikler, ingen havde brug for.

Når det er sagt, taler vi stadigvæk om næsten 18 mio. gange, at en ordbogsbruger har søgt efter og fundet det ord, han eller hun søgte efter oplysninger om. De hyppigste søgninger er på følgende ord:

udgangspunkt	21.832
hej	18.450
samarbejde	18.001

forhold	17.568
som	15.299
mulighed	14.700
hest	14.689
sammenhæng	14.385
om	12.799
kollega	12.605
for	12.529
medføre	12.527
udføre	12.516
undersøge	12.383
derfor	12.195
ordbog	12.168
opnå	12.152
krav	12.145
udvikling	12.061
efterfølgende	12.028
opgave	11.859
forslag	11.619
forbindelse	11.547
aftale	11.361

Sådanne lister med “hyppigst søgte ord” er meget forskellige, hvilket vises i Bergenholtz/Johnsen (2005). Der sammenlignes der mellem søgninger på dansk, tysk, engelsk og norsk. Men også mellem forskellige logfilanalyser af danske monolingvale ordbøger er der store forskelle, kun *hest* går igen i alle lister. Man kunne sammenligne hyppighedsundersøgelser som Bergenholtz (1992). Det er dog ikke temaet her.

I det hele taget er der ikke nogen entydig relation mellem ord, der er søgt hyppigt, og dem, der er søgt sjældnere (således konkluderer også de Schryver/Joffe 2004). På grund af den måde, Googles søgealgoritme fungerer på, kan antallet af søgeresultater afvige

ICKE-TEMATISKA BIDRAG

en smule fra de angivne tal her i artiklen. Alle vores søgninger er foretaget den 21.2.2012. 100 gange har brugerne søgt på følgende ord, hvortil antal resultater for en Google-søgning på danske sider tilføjes, søgeordet angivet i anførselstegn i søgefeltet:

profileret	272.000
pomfrit	26.200
plejlstang	42.200
pladder	117.000
ommer	214.000
omgangssyge	19.200
om dirigere	101.000
olympiade	60.100
scenograf	69.300
safir	669.000
ressentiment	2.100
skolde	17.000
sidekammerat	24.300

Man bør her især lægge mærke til fx *ressentiment*, *skolde* og *omgangssyge*, som alle har ret lave Google-tal, særligt *ressentiment*. Forskellen til hyppighedsbrug er i alle tilfælde ikke påfaldende stor, når man sammenligner med en liste med ord, der er søgt på 10 gange, her også med tilføjelse af resultater ved en Google-søgning på danske sider den 21.2.2012:

svineheldig	21.400
vipbar	53.900
voksbehandling	74.800
voldgiftsdom	4.836
stabelstol	44.200
stillingsbetegnelse	141.000
storsnudethed	2.580

sultkatastrofe	99.900
svampemiddel	16.400
tidobling	57.400
tidsadverbium	3.930

Tendentielt er gennemsnittet af Google-resultater dog lidt mindre end ved resultater med 100 søgninger. Der er sjældent brugte ord som *tidsadverbium* og hyppigere brugte ord som *stillingsbetegnelse*. Med andre ord kan frekvenskriteriet godt spille en rolle for relationen til hyppig eller sjældnere ordbogsbrug, men brugsfrekvens kan ikke beskrives som den eneste eller den afgørende faktor, når man ser på antallet af ordbogssøgninger.

2. Lemmaoverflod

Termen **lemmahul** (synonym **lemmalakune**) er kendt og almindelig brugt i det leksikografiske fagsprog. Men hvad hedder, eller hvad kan vi kalde det modsatte af et lemmahul? *Lemmafryns*, *lemmafylt* eller *lemmaoverflod* kunne være første bud. Vi vil foreløbig vælge **lemmaoverflod** (eng. **lemma flooding**, ty. **Lemmaüberfluß**) for at betegne de mange ord og ordformer, ingen brugere søger efter, i vores undersøgelse ud af en mængde på knap 18 mio. opslag.

Vi citerer her et tilfældigt udvalg i den rækkefølge, logfilen viser dem i. Vi viser også antallet af Google-resultater på danske sider i en søgning med søgeordet skrevet i anførselstegn den 21.2.2012:

logesal	1.130
logeplads	1.420
logement	38.900
logemedlem	4.980
logaritmisk	186
logaritmetabel	4.270

ICKE-TEMATISKA BIDRAG

logaritmefunktion	26.100
loftsventilator	32.100
loftstige	2.600
loftsplade	12.800
loftmaleri	21.200
loftmaleriet	619
loftmalerier	15.300
loftmalerierne	1.460
loftslampe	191.000
loftslampen	3.890
loftslamper	142.000
loftslamperne	1.410
loftskammer	7.590
loftshummer	8
loftsetage	10.300
loftsatelier	367
loftplade	17.400
loftlys	21.800

Man kan her se hyppigt brugte ord som *loftslampe*, men også sjældent brugte ord som *loftshummer*.

Er der her noget system bag den store andel af lemmafyld efter en første analyse af denne første liste? Ja, meget sjældent brugte ord kunne man nok udelukke af lemmalisten. Grænsen for, hvad der er et sjældent ord, kan være svær at definere og må vurderes af den enkelte leksikograf til en given ordbog. Til denne ordbog vil vi sige, at tallet ud fra en Google-liste kunne være: Medtag ikke lemmer, som her har en hyppighed på under 1.000. Vi kunne forestille os følgende ni forklaringer på, at der ikke rigtigt er nogen brugere, der vil slå et ord op:

(1) Ikke noget system overhovedet

En mulighed er, at der ikke er nogen systematisk forklaring på problemet. Altså at man ikke kan finde en årsag til, at nogle ord bliver slået op, mens andre ikke bliver det. Det er den løsning, som var buddet i Bergenholtz/Johnsen (2005). Det er også den løsning, som er særligt utilfredsstillende, fordi den ikke giver leksikografen hjælp til at undgå at lave overflødigt arbejde.

(2) Ikke-relevante ord

Denne kategori er brugbar til specialordbøger, fx til musikordbogen *Betydning af musikudtryk* (2012). Hvis en sådan ordbog indeholdt medicinske termer, ville ingen vel søge efter sådanne i en musikordbog. De ville være lemmaoverflod og kan på forhånd udelukkes ved lemmasektionen. Dette argument kan dog ikke bruges ved denne almensproglige ordbog, som i princippet kan og skal indeholde alle ord, der forekommer eller skal bruges i danske almensproglige tekster og derfor kunne give anledning til ordbogsbrug.

(3) Ord, alle kender

Ved en ordbog for brugere med et andet modersmål end dansk kan et argument om, at man kan udelade ord, alle kender, ikke bruges. Men denne ordbog, *Den Danske Netordbog*, er en ordbog ment for brugere, som har dansk som modersmål. Alligevel holder argumentet ikke. Der findes masser af 0-søgninger på sådanne ord, fx *loftslampe*, *olieforurening* eller *blikkrus*, men de fleste ord, der søges på, er ord, alle danskere kender, vi kan bare henvise til listerne med 10 og 100 søgninger i foregående kapitel. På trods af hensigten med ordbogen så benyttes den desuden også af en del brugere uden dansk som modersmål.

(4) Komposita

Specielt når det drejer sig om de såkaldte “gennemsigtige” komposita, hvor betydningen for ordet svarer til summen af orddelenes betydning, er der ofte fremsat forslag om at udelade dem fra en ordbogs lemmabestand. Der findes sådanne komposita med en 0-søgning: *helsidesannonce*, *oliefyr*, *nordtysker*, *nordsvensker*. Men der er et (næsten) lige så stort antal komposita af denne type, der er søgt på mange gange, fx

problemformulering	4.911 opslag
barselsorlov	1.612 opslag
arbejdsopgave	985 opslag
videreudvikling	955 opslag
afdelingsleder	993 opslag
projektleder	889 opslag
bestyrelsesmedlem	787 opslag

Denne forklaring er derfor ikke brugbar. Heller ikke ved brug af de såkaldte affiksoider, fx *super-* *giga-*, *ultra-*, alle med en forstærkende betydning. Rigtigt er det, at følgende ord ikke er slået op en eneste gang: *gigafed*, *ultrahurtig*, *superhurtig*. Men *superflot* har haft ikke mindre end 29 opslag. Dette ord giver ganske vist mange resultater ved en Google-søgning, men det ikke søgte *superhurtig* har trods alt 74.300 resultater (21.2.2012) mod 568.000 for *superflot*.

(5) Fremmedord

Præcis det samme argument kan fremføres om fremmedord: Nogle søges, på andre ikke. Ved fremmedordene må andre regler gøre sig gældende, fx argumenterne om fagord og hyppighed.

(6) Helt nye ord, de såkaldte neologismer

Disse søges på, endda meget og straks efter deres første brug, fx:

smartphone	71 opslag
tablet-pc	71 opslag
ipad	30 opslag
iphone	62 opslag

Neologismer hører tværtimod til i toppen af ordbogens lemmahuller, så de risikerer næsten aldrig, når de er optaget i ordbogen, at blive til ord med 0-søgninger.

(7) Gamle eller forældede ord

Den Danske Netordbog har en korpusstøttet lemmaselektion, således at forældede ord i princippet ikke blev medtaget. Man må her skelne mellem, om tingen/sagen eller ordet er forældet. En båndoptager fx er som ting sikkert forældet, men ordet ikke. *Båndoptager* blev der i logfil-perioden søgt på 68 gange. Der blev til gengæld slet ikke søgt på følgende ord, hvortil der tilføjes fundne resultater ved en Google-søgning på danske sider den 21.2.2012 og derudover, hvor ofte ordet har været benyttet på danske hjemmesider ifølge Google i løbet af det sidste år:

Ord med 0-søgning	Google-resultater	Google-resultater sidste år
lispund	6.390	244
kattun	2.760	58
kattuntryk	352	9
karbidlygte	280	5
karbidlampe	1.260	21
kinematograf	1.680	47

Tabel 1: Forældede ord med 0-søgning.

Her begynder der at tegne sig et mønster: Om et ord, som leksikografen anser for at kunne være forældet, vil vi sige, at sådanne

ord ikke skal tages med i lemmabestanden, hvis det ved en Google-søgning har mindre end 2.000 resultater eller mindre end 100 resultater indenfor det sidste år. Denne regel, hvis den var blevet brugt til lemmaselektionen til Den Danske Netordbog, havde medført, at ingen af ordene fra den ovenstående liste (bortset fra *lispund*) var blevet selekteret til lemmabestanden. Ved ordbøger med en mindre lemmabestand eller med en anden empirisk basis end Google, dvs. et tekstkorpus, skal disse tal naturligvis justeres. Men princippet er, at man ved en kombination af absolut hyppighed og hyppighed i de nyeste tekster vil kunne undgå at medtage ord, som der alligevel ikke bliver slået op på.

(8) Fagord og håndværksmæssige udtryk

Her kan man i listen over de ord, der er søgt på, se, at der søges på mange fagord og håndværksmæssige udtryk. Men fra 0-listen, ud fra listen over ikke brugte søgeord, kan man lave en tilsvarende regel som for forældede ord. Sammenlign hertil følgende liste med ord, der ud af knap 18 mio. søgninger i ordbogen, har truffet et lemma i ordbogen:

Ord med 0-søgning	Google-søgning	Google-søgning sidste år
tetracyclin	16.700	665
tetralogi	7.580	219
spændkraft	63.700	59
gigabit	699.000	19.500
gigahertz	14.600	1.260
gigajoule	5.720	173

Tabel 2: Fagord med 0-søgning.

Ud fra disse 0-søgninger og under brug af den ovenstående regel “skal ikke medtages i lemmabestanden, hvis det ved en Google-

søgning har mindre end 2.000 resultater på Google eller mindre end 100 resultater det sidste år” vil kun *spændkraft* være blevet udeladt. Det rene frekvenskriterium, hvor man ser på det samlede korpus hhv. her hele Google, fører til selektion af ord, som i første gang kunne udelades. Den ovenstående regel er dog heller ingen hjælp i tilfældet ved fagord.

(9) Ikke-hyppige ord

Hyppighed er det oftest brugte kriterium ved lemmaselektion. Vi vil her gentage, at der ikke er en direkte relation mellem hyppighed i tekster og hyppighed i søgning efter ord i en ordbog. Et af de bedste eksempler er det netop nævnte *gigabit*, et andet *loftslampe*. Vi vil afslutningsvis bringe et sidste uddrag af 0-listen, af listen over ord og ordformer, ingen brugere ud af de knap 18 mio. opslag nogensinde har søgt på. Man kan se, at mange af disse ord er ret hyppige, fx *bleskift*:

Ord med 0-søgning	Google-søgning	Google-søgning sidste år
blegevand	9.520	28
blegfed	15.900	130
blegfedt	189	4
blegfede	21.600	350
blegfiset	3.970	14
bleggul	11.200	143
bleghvid	3.970	93
bleglilla	559	7
blegning	112.000	6.990
blegselleri	39.900	1.640
blegsot	5.480	42
blegsotig	8.000	45
blegvand	5.970	4

ICKE-TEMATISKA BIDRAG

bleskift	124.000	5.560
bleskifte	6.640	202
blesnip	3.120	6
blevask	4.380	43
blidelig	9.040	23
blidhed	34.500	1.160
blikhus	1.600	9
blikkrus	1.160	15

Tabel 3: Andre ord med 0-søgning.

Under brug af begge frekvenskriterier ville flertallet af ordene i denne 0-liste ikke være blevet medtaget. Dog ville *blegfed*, *blegfede*, *bleggul*, *blegning*, *blegselleri*, *bleskift* og *blidhed* være kommet med.

3. Kan man undgå at lave ikke-læste artikler?

Svaret er nej! Vi taler i denne undersøgelse om en andel af ikke-søgte lemmaer og ordformer på 85,8%. Når man kun tæller ordbogsartikler, er tallet 66,6%. Der er søgt på 37.238 artikler, dvs. at der IKKE er søgt på 74.254 artikler i Den Danske Netordbog. En regel, som ikke blev brugt til Den Danske Netordbog, men som vil blive brugt til kommende ordbøger af denne størrelse, vil lyde, at potentielle ord ikke medtages som lemmaer, hvis:

de ved en Google-søgning har mindre end 2.000 resultater eller mindre end 100 resultater i løbet af det sidste år

Hvis en sådan regel bruges, kan man risikere at udelukke ord, som nogle ordbogsbrugere virkelig gerne ville vide noget specielt om. Ud fra breve fra sådanne brugere kan man i en trykt ordbog først udbedre lemmahullerne i en ny udgave. I elektroniske udgaver

kender vi to løsninger: Den ene bliver brugt i fx Ordbogen.com Dansk-Engelsk (2012). Her tilføjes i løbet af en dag alle ord, som brugere har søgt på, men som ikke var i lemmalisten. I Den Danske Netordbog (2012) er der en del artikler, som kun indeholder lemma og grammatik, men ikke andre angivelser (mere præcis drejer det sig om 34.000 artikler). Når brugere henvender sig om manglende data i ordbogen, bliver de tilføjet straks.

Den nævnte regel er kun undersøgt for Den Danske Netordbog, men kan sandsynligvis bruges i udarbejdelsen af lignende almensproglige ordbøger. Den kan derimod med sikkerhed ikke bruges ved fagordbøger og andre typer ordbøger kræver yderligere undersøgelser.

Litteratur

Ordbøger

Betydning af musikudtryk = Inger Bergenholtz i samarbejde med Henning Bergenholtz. Database: Richard Almind og Martin Gyde Poulsen: *Betydning af musikudtryk*. Odense: Ordbogen.com 2012. <www.ordbogen.com>.

Den Danske Netordbog = Henning Bergenholtz under medvirken af Filip Odgaard Bodilsen m.fl. Database: Richard Almind, Rasmus Theodor Styrk, Peter Christensen: *Den Danske Netordbog*. Odense: Ordbogen.com 2009-2012. <www.ordbogen.com>.

Ordbogen.com Dansk-Engelsk = Anja Becher Andresen, Elisabeth Hedegaard Kristiansen og Jacqueline R. Levin i samarbejde med Bjarni Norddahl og Michael Walther under medvirken af Ann-Christine Weber Brandt m.fl. Database: Michael Walther og Bjarni Norddahl: *Ordbogen.com Dansk-Engelsk*. Odense: Ordbogen.com 2006-2012. <www.ordbogen.com>.

Anden litteratur

- Bergenholtz, Henning (1989): Probleme der Selektion im allgemeinen einsprachigen Wörterbuch. I: *Wörterbücher. Dictionaries. Dictionnaires. Ein internationales Handbuch zur Lexikographie. An International Encyclopedia of Lexicography. Encyclopédie internationale de lexicographie. Erster Teilband*, hrsg. von Franz Josef Hausmann, Oskar Reichmann, Herbert Ernst Wiegand, Ladislav Zgusta. Berlin/New York: de Gruyter, 772-779.
- Bergenholtz, Henning (1992): *Dansk frekvensordbog. Baseret på tekster fra danske romaner, ugeblade og aviser fra 1987-1990*. København: Gad.
- Bergenholtz, Henning (2003): Die Entwicklung der Lemmaselektion. I: Herbert Ernst Wiegand (Hrsg.): *Untersuchungen zur kommerziellen Lexikographie der deutschen Gegenwartssprache I. "Duden. Das große Wörterbuch der deutschen Sprache in zehn Bänden"*. Tübingen: Niemeyer, 83-98.
- Bergenholtz, Henning (2005): Lemmaselektion. I: *Schreiben, Verstehen, Übersetzen und Lernen: Zu ein- und zweisprachigen Wörterbüchern mit Deutsch*, hrsg. von Irmhild Barz/Henning Bergenholtz/Jarmo Korhonen. Frankfurt a.M./Bern/New York/Paris: Peter Lang, 147-164.
- Bergenholtz, Henning/Mia Johnsen (2005): Log Files as a Tool for Improving Internet Dictionaries. I: *Hermes* 34, 117-141.
- Bergenholtz, Henning/Mia Johnsen (2007): Log Files Can and Should Be Prepared for a Functionalistic Approach. I: *Lexikos* 17, 1-20.
- de Schryver, Gilles-Maurice/David Joffe/Pitta Joffe/Sarah Hillewaert (2006): Do Dictionary Users Really Look Up Frequent Words? — On the Overestimation of the Value of Corpus-based Lexicography. I: *Lexikos* 16, 67-83.

- de Schryver, Gilles-Maurice/David Joffe (2004): On How Electronic Dictionaries are Really Used. In: Geoffrey Williams/Sandra Vessier (Eds.): *Proceedings of the Eleventh EURALEX International Congress, Euralex 2004, Lorient, France. July 6-10, 2004*. Volume I. Lorient: Université de Bretagne, 187-196.
- Ling, Charles X./Jianfeng Gao/Huajie Zhang/Weining Qian/Hongjiang Zhang (2002): Improving Encarta Search Engine Performance by Mining User Logs. I: *International Journal of Pattern Recognition and Artificial Intelligence* 16, 1101-1116.

Henning Bergenholtz
professor
Center for Leksikografi
Institut for Erhvervs-
kommunikation
Århus Universitet
Fuglesangs Allé 4
DK-8210 Aarhus V
hb@asb.dk

Bjarni Norddahl
bestyrelsesformand
Ordbogen A/S
Billedskærervej 8
DK-5230 Odense M
btn@ordbogen.com