

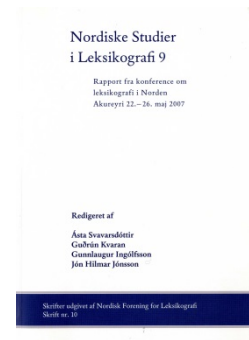
# NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Från ord till handling - En studie i ordboksanvändning på nätet

Forfatter: Ann-Kristin Hult

Kilde: Nordiska Studier i Lexikografi 9, 2008, s. 243-257  
Rapport fra Konference om leksikografi i Norden, Akureyri 22.-26. maj 2007

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



© Nordisk forening for leksikografi

## 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.

## Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) 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.

ANN-KRISTIN HULT

## Från ord till handling

### En studie i ordboksanvändning på nätet

The article presents the results of a log file analysis of the searches in *Lexin Svenska ord*, a free Swedish monolingual Internet dictionary. It focuses on the non-successful searches, called not-found. The results show, firstly, that a considerable amount of the not-found are correctly spelled Swedish words, but they are not included in the lemma selection of the dictionary. Users also often search for two (or more) word expressions. Secondly, the dictionary's programmed response in the case of not-found gives far too many suggestions for it to be user-friendly; sometimes the user is given as many as forty suggestions on words that he or she could have been searching for. Thirdly, users tend to repeat their searches in order to obtain a successful search, e.g. by trying another spelling. It is shown that log files can be a helpful tool in the lexicographers' work improving electronic as well as paperback dictionaries.

#### 1. Inledning

Som en följd av att ordböcker har blivit tillgängliga i elektroniskt format har möjligheten uppstått att registrera alla sökningar som utförs i så kallade loggfiler. Via de automatiskt registrerade loggfilerna är det möjligt att få tillgång till stora material, i detta fall användarnas sökningar i en Internet-ordbok. Vid traditionella användarstudier av pappersordböcker är informanterna färre och brukarsituationen (oftast) konstruerad. Via loggfilerna kan man få reda på vilka ord användarna faktiskt slår upp och, då mer avancerade sökningar är möjliga, vilken typ av information de är ute efter. Det är således dels ett relativt nytt medium som undersöks, Internetordböcker, dels ett nytt verktyg som används för att studera ordboksanvändning, loggfiler. Befintliga studier visar också att loggfiler kan fungera som ett effektivt hjälpmedel i arbetet med att förbättra såväl pappersordböcker som elektroniska ordböcker.

I denna artikel presenteras resultaten av en studie av loggfilerna från den svensk-svenska delen av *Lexin*-seriens ordböcker, *Svenska ord*, som ligger fritt tillgänglig på Internet. Syftet med studien har varit att ta reda på vilka typer av ord som användarna söker efter och hur de gör då de inte får någon träff på det ord som efterfrågas. Detta undersöks med hjälp av de uppemot 17 000 registrerade sökningarna i loggfilerna från den första oktober 2006. Andelen icke framgångsrika sökningar visar sig vara mycket hög, hela 41 %. Det är därför

intressant att undersöka vad användarna vill men inte kan få reda på, samt vilka strategier användarna verkar ta till då de inte får någon träff. Studien avgränsas således till de icke framgångsrika sökningarna, hädanefter kallade icke-träffar. Målet med undersökningen har varit att få kunskap om användarnas faktiska sökbeteende och därigenom också vad de förväntar sig att (e-)ordboken ska innehålla.

Artikeln är disponerad på följande sätt: i kapitel 2 redogör jag för *Lexin*-serien och i kapitel 3 ger jag en inblick i forskningen kring ordboksanvändning där jag också behandlar tidigare loggfilsstudier. I kapitel 4 presenteras materialet. Resultaten redovisas sedan i tre avsnitt under kapitel 5. Till sist sammanfattas resultaten i kapitel 6 där jag också blickar framåt mot vidare användarforskning.

## 2. *Lexin*-serien

*Lexin*-serien består för det första av en enspråkig svensk ordbok, *Svenska ord*. Den har tagits fram av Myndigheten för skolutveckling tillsammans med Språkdata under ledning av Martin Gellerstam (se vidare Gellerstam 1999). För det andra ingår ett trettiotal tvåspråkiga lexikon på stora minoritetsspråk i Sverige. *Svenska ord* har fungerat som underlag vid framställning av de tvåspråkiga delarna av ordboksserien.

*Svenska ord* har 28 500 uppslagsord. Ordurvalet är gjort med tanke på invandrare som ska lära sig svenska och få tillgång till vanliga ord samt samhällsord som är viktiga för att man ska kunna orientera sig i det svenska samhället. I förordet påpekas det dock att *Svenska ord* också kan användas som skolordbok. I övrigt innehåller ordboksartiklarna språklig information kännetecknande för en definitionsordbok (t.ex. definitioner, uttalsangivelser, böjningsformer och språkexempel), men också uppgifter av mer grammatisk karaktär (t.ex. satskonstruktion). I den tredje upplagan (1999) har samhällsorden uppdaterats och ord som har med datorer och informationsteknologi att göra har lagts till (*Svenska ord* 1999:V).

Internetversionen av *Lexin*-serien har funnits sedan 1994 och distribuerades fram till och med 2006 av Myndigheten för skolutveckling i samarbete med KTH (Kungliga Tekniska högskolan i Stockholm). Från och med den första januari 2007 tillhandahålls tjänsten i det nya Språkrådets regi. I dagsläget finns femton lexikon fritt tillgängliga via webbsidan. Bild 1 visar webbsidans utseende efter en sökning på ordet *nyckel* i *Svenska ord*.

The screenshot shows the Lexin online dictionary interface. The header includes 'Språkrådet' and 'Institutet för språk och folkminnen'. The main content area is titled 'Svenska ord' and 'Översättning av nyckel'. The entry for 'nyckel' includes its phonetic transcription [n'yk:el], a pronunciation icon, and a definition: 'nyckeln nycklar subst. redskap att öppna el. stänga lås med (bildligt "verksamt medel, lösning"; i sms även bildligt "viktig, central")'. A link to 'bildtema 10' is provided. Below the definition, there are sections for 'Exempel' and 'Sammansättningar/avledningar'. The 'Exempel' section lists 'vrida om nyckeln' and 'nyckeln till framgång'. The 'Sammansättningar/avledningar' section lists 'dörrnyckel', 'bilnyckel', 'nyckelord -et', 'nyckelroll -en', and 'nyckelperson'. The left sidebar contains navigation options like 'Svenska till svenska.', 'Rättstava okända ord.', 'Översätt!', 'Extra svenskt tangentbord:', 'Bläddra bland Lexins ord:', and 'Lista med sparade ord:'. At the bottom, there is contact information: 'Kontakt: lexin(å)sprakradet.se (obs: byt ut (å) mot @)' and 'Tekniskt stöd: webbmaster@nada.kth.se'.

Bild 1. Exempel på uppslagsordet *nyckel*, hämtat 2007-08-28.

Artikeln ger först information om uttal, som man även kan lyssna till, böjningsformer och ordklassstillhörighet. Sedan följer en betydelsebeskrivning och som synes har *nyckel* både en bokstavlig och bildlig betydelse. Det är även möjligt att se en bild på uppslagsordet via länken "bildtema 10", där man får se i vilka sammanhang en nyckel används. Vidare ges syntaktiska och morfologiska exempel samt sammansättningar och avledningar. I nedre delen av spalten till vänster om uppslagsordet finns länkar till de övriga lexikonerna i *Lexin*-serien.

### 3. Ordboksanvändning

#### 3.1. Forskning om ordboksanvändare och ordboksanvändning

Hartmann (1989:102–103) ställer upp tolv hypoteser om ordboksanvändaren. En av dem handlar om att lexikografer alltid har en viss uppfattning om vem användaren är och vilken utformning av ordboken som är mest användarvänlig. Ett annat antagande slår fast att det inte finns något ett-till-ett-förhållande mellan det sätt på vilket ordboken är avsedd att användas och det sätt som ordboken sedan faktiskt brukas på. Användarna missförstår informationen, till exempel vad en huvudkommentar motsvarande "vard." innebär (Bejoint 1989:208). De har också en övertro på ordbokens definitioner och ordurval (jfr SAOL 13:V).

Det råder dessutom en okunskap om det varierande utbud av ordböcker som finns och deras olika syften (Bejoint 1989:208). Ordboksanvändaren kanske till exempel letar efter definitioner i en synonymordbok. Beträffande e-ordböcker tillkommer svårigheten att välja rätt sökalternativ bland den mängd av avancerade sökmöjligheter som finns.

Vad gäller användningen av e-ordböcker befinner sig forskningen i en initial fas och än så länge är det svårt att veta vad utvecklingen av e-ordböcker har betytt för ordboksanvändningen. Enligt Svensén (2004:553f) har användarforskningen tidigare i huvudsak inriktat sig på inlärningsordböcker och tvåspråkiga ordböcker. Däremot vet man lite om ordboksanvändning av svenska, enspråkiga allmänordböcker och de elektroniska varianter som finns. De flesta empiriska undersökningar har dessutom genomförts med enbart studenter som informanter. Hur olika yrkesgrupper, framför allt de som yrkesmässigt arbetar med språk, eller människor i allmänhet använder ordböcker har man mindre kunskap om. Genom att studera loggfiler, där användaren är anonym, kommer man åt (de verkliga) användarna utan att veta deras ålder, kön eller sysselsättning. Det finns dock goda möjligheter att få reda på mer information om användarna om de registrerar vilka de är.

### 3.2. Loggfilsstudier

Tidigare studier av loggfiler visar att de är ett utomordentligt hjälpmedel i arbetet med att göra ordböcker bättre. Genom att analysera loggfiler kan den typen av hypoteser som Hartmann ställer upp prövas eftersom man får reda på vad användarna faktiskt efterfrågar. Dessutom kan antaganden om den typiska användaren undersökas genom att man "bjuder in" brukarna till att kommentera och svara på frågor on-line om sitt ordboksanvändande.

Bergenholtz & Johnsen har undersökt loggfilerna från *Den Danske Netordbog*, <[www.netordbog.dk](http://www.netordbog.dk)>, en enspråkig ordbok med 109 000 lemman, och konstaterar inledningsvis följande:

Analyses of log files reveal exactly which lemmas and which types of information have been requested, and, perhaps more significantly, which lemmas and which types of information have been requested but were not found in the dictionary. (Bergenholtz & Johnsen 2005:117)

Med hjälp av loggfilerna kunde författarna ta reda vilka typer av problem användarna stötte på när de sökte på ett ord och inte fick träff. Loggfilerna visade till exempel att användarna sökte på passiv- och imperativformer av verb utan att få träff. Det fanns alltså ett behov av att kunna utföra den typen av sökningar och därför bör man fundera över om inte dessa former ska vara sökbara.

Studien visade också att nästan alla uppslagningar i topp 100-listan av sökningar som misslyckades hade sin förklaring i felstavningar. Det finns olika orsaker till att man stavar fel och de vanligaste anledningarna torde vara slarv, okunskap och

dyslexi. En stor del av felstavningarna bland de undersökta loggfilerna berodde enligt författarna på en uttalsmässig stavning. En annan orsak till att man inte fick träff var att man inte visste om ett uttryck skrevs som ett eller två ord. Icke-träffarna bestod också av ord som inte finns, vilket exempelvis kunde bero på att ett ord skrevs med suffixet *-ning* (*beskrivning*) i stället för det i danskan korrekta *-else* (*beskrivelse*). Användarna verkade också vara osäkra på när de skulle använda *foge-s*.

Ett av de största användningsområdena för loggfilerna är dock upptäckten av lemmaluckor genom att man ser vilka sökningar som inte fått träff. Många av de ord som användarna inte fann i ordboken var datarelaterade, finansiella, juridiska och medicinska termer och begrepp. De vanligaste begreppen inom respektive område, de som redan används frekvent i allmänspråket, borde enligt författarna ingå i ordboken (Bergenholtz & Johnsen 2005:137–138).

Vidare har de Schryver & Joffe (2004) undersökt loggfiler i en tvåspråkig Internetordbok mellan ett sydafrikanskt bantuspråk, Sesotho sa Leboa, och engelska. Även här bestod en överraskande hög andel av sökningarna av felstavade ord. Författarna redogör också för hur loggfilerna regelbundet utnyttjades under projektets gång. Ett sätt var att frekventa icke-träffar, som visade sig vara lemmaluckor i ordboken, lades till när ordboken uppdaterades.

I studien redogör de Schryver & Joffe även mer detaljerat för hur de kunde följa varje specifik användares sökningar och därmed undersöka dennes sökstrategier. De lade därtill upp ett formulär on-line där användarna kunde ge feedback på ordboken. Författarna påpekar till sist hur kontinuerlig analys av loggfiler tillsammans med användarnas feedback kan utnyttjas till att anpassa innehåll och utformning efter brukarnas behov (jfr Bergenholtz & Johnsen 2007).

#### 4. Material

Den studie som presenteras här grundar sig alltså på en analys av loggfilerna från Lexin *Svenska ord* från den första oktober 2006, ett slumpmässigt valt datum, och utgörs av 16 903 sökningar.<sup>1</sup> Av dessa sökningar är 41 % (6 941) felaktiga uppslagningar, vilka jag har valt att fokusera min studie på. Felaktiga uppslagningar är de sökningar som inte leder till träff i ordboken, alltså icke-träffar. Av tabell 1 framgår hur siffrorna ser ut i *Svenska ord* under delar av 2006 och 2007.

1 Tack till Viggo Kann, professor i datalogi vid KTH i Stockholm, som möjliggjort denna studie genom att tillhandahålla loggfilerna och svara på frågor om loggfiler och statistik.

Månad	År 2006			År 2007		
	Totalt antal uppslagningar	Felaktiga uppslagningar	Procent felaktiga ord	Totalt antal uppslagningar	Felaktiga uppslagningar	Procent felaktiga ord
Maj	410 154	173 978	42	627 593	247 945	40
Juni	221 865	89 891	40	300 567	116 328	39
Juli	200 048	69 670	34	231 654	91 496	39

Tabell 1. Antal uppslagningar i *Svenska ord* under maj-juli 2006 respektive maj-juli 2007.

I de två tidigare loggfilsstudierna jag refererar till, (se avsnitt 3.2), är både antalet sökningar och andelen icke-träffar lägre. I den danska nätordboken utfördes i snitt drygt 2 200 per dag mellan januari 2003 och augusti 2004 och andelen misslyckade sökningar var 19,5 % under den period som sökningarna loggfördes. I den afrikansk-engelska ordboken var snittet per dag drygt 3 500. Studien genomfördes från det att Internetordboken blev tillgänglig, april 2003, och sex månader framåt. Två uppdateringar genomfördes under de undersökta månaderna och efter den andra uppdateringen sänktes felprocenten från 33 % till 25 %. I båda ordböckerna ökade antal besökare snabbt med tiden.

Antalet uppslagningar är alltså mycket högre i *Svenska ord* än i de två ovan nämnda ordböckerna. Det beror troligtvis på att *Lexin*-seriens ordböcker har funnits tillgängliga via Internet uppemot tio år längre. De har dessutom distribuerats av Myndigheten för skolutveckling med fokus på andraspråksinläring och som skolordbok. Den har därmed haft en tydligt avgränsad målgrupp och myndigheten har på olika sätt aktivt spritt kunskap om e-ordbokstjänsten.

Efter en första genomgång av de uppemot 7 000 icke-träffarna visade det sig att uppemot 45 % av dem bestod av felstavningar. Det som dock fängslade mig var, för det första vilka övriga typer av sökningar som resulterade i icke-träffar, och för det andra användarnas benägenhet att upprepa sina sökningar. Jag valde således att inte fokusera på felstavningarna vilka, tillsammans med en grupp oidentifierbara uppslagningar, endast kort berörs i avsnitt 5.2. Det material som redovisas i det följande utgör ungefär en tredjedel av loggfilerna.

## 5. Resultat

Resultaten presenteras under tre avsnitt där jag först tar upp vilka typer av sökningar som resulterar i icke-träff. Sedan beskrivs hur ordboken är programmerad att svara när sökordet inte finns i ordbokens lemmauppsättning. Till sist redogör jag för de sökstrategier som användarna tar till då de inte lyckas i sin första sökning.

### 5.1. Sökningar som resulterar i icke-träff

De uppslagningar som inte ger träff i *Svenska ord*, efter ovan nämnda avgränsning, är av olika slag och delas in i olika typer så som tabell 2 visar.



Typ av sökning	Antal	Procent
Enkla ord, avledningar och sammansättningar	800	58
Utländska ord (90% engelska)	160	12
Encyklopediska och Google-liknande sökningar	140	10
Fasta konstruktioner: fraser, idiom, kollokationer	90	6,5
Tillfälliga konstruktioner	65	5
Årtal, sifferkombinationer	50	3,5
Förkortningar	30	2,5
Felaktigt särskrivna uttryck	25	2
En/ett+substantiv och att+verb	10	<1
Totalt	1370	100

Tabell 2. Typer av sökningar som resulterar i icke-träff ordnade efter frekvens.

Det är i en del fall svårt att avgöra i vilken grupp ett ord ska placeras, t.ex. om en fras är en fast eller mer tillfällig konstruktion eller om en encyklopedisk sökning i stället tillhör gruppen med enkla ord, avledningar och sammansättningar. En del ord och uttryck har därför placerats i två grupper. Siffrorna och procenttalen har avrundats men torde ändå vara representativa för hur vanliga de olika typerna av sökningarna är i förhållande till varandra.

#### 5.1.1. Sökning på enkla och sammansatta ord, avledningar samt utländska ord

Mest frekventa är sökningarna som utgörs av enkla ord, avledningar och sammansatta ord. Användarna söker på såväl grundform som pluralformer och bestämda former. Nedan anförs exempel på de tre kategorierna:

- Enkla ord: *egal, mejla, hölster*
- Avledningar: *integration, finansiering, modernisering*
- Sammansättningar: *friskvård, ansiktsmimik, fängelsevistelse, hissfobi*

Det kan först konstateras att användarna har höga förväntningar på ordboken vad gäller dess urval av uppslagsord. Därför är det inte enbart negativt att många av de undersökta icke-träffarna utgörs av söksträngar som är rättstavade befintliga ord i svenska språket. Att man inte får träff beror dels på det begränsade lemmaurvalet (28 500 lemman) i *Svenska ord*. Urvalet är gjort med tanke på invandrarens behov och *Lexin* verkar attrahera en mycket mer heterogen grupp än så.

Med tanke på den stora mängd olika ord användarna söker på verkar de inte veta vilken typ av ordbok *Lexin* är och vilken målgrupp den har. Precis som med tryckta ordböcker, där förord och inledning ofta inte läses, är det säkert så att webbsidans besökare inte tar reda på vilken typ av ordbok de har framför sig på skärmen. Eller så vet de det, men hittar inget annat alternativ och prövar med *Lexin*. Orden i denna kategori utgör alltså lemmaluckor i ordboken och visar hur loggfiler kan utnyttjas till att uppdatera och utöka lemmauppsättningen. Det skulle tillfredsställa en större och bredare grupp av användare utan att för den skull missgynna den primära målgruppen, andraspråksinlärare.



Näst efter den nyss nämnda gruppen kommer sökningar på utländska ord, ca 160 stycken, varav närapå 90 % engelska, till exempel *education*, *throw* och *influential*. Det är något förvånande att utländska ord återfinns bland sökningarna i *Svenska ord* av det skälet att man via *Lexins* startsida aktivt måste välja vilken av ordböckerna man vill använda.

### 5.1.2. Flerordssökning

Tvåordslemman som partikelverb och mer eller mindre fasta konstruktioner som t.ex. idiom och kollokationer samt tillfälliga konstruktioner är också en av de vanligaste typerna av sökning som leder till icke-träff:

- Reflexiva verb: *föresätta sig*, *anstränga sig*
- Partikelverb: *känna igen*, *fälla ihop*
- Idiom: *kalla fötter*, *lägga lök på laxen*
- Kollokationer: *psykisk ohälsa*, *civil olydnad*
- Tillfälliga konstruktioner: *hierarkiskt samhälle*, *säljande image*

Med sökfunktionen fritextsökning och generösa hänvisningar skulle användaren kunna hitta fram till idiom, kollokationer, vanliga sammansättningar samt annan information som ligger gömd i ordboksartiklarna. I nuvarande skick får användaren inte träff på ord som faktiskt finns i ordboken, t.ex. det reflexiva verbet *föresätta sig* eller partikel verbet *känna igen*, p.g.a. att verbets uppslagsform står i presens (*föresätter sig* respektive *känner igen*). Användaren får således informationen att orden inte finns i ordboken vilket är rent felaktigt. Ett enkelt verb rättas dock automatiskt till uppslagsformen presens i de fall man söker på en annan tempusform.

### 5.1.3. Encyklopediska och Google-liknande sökningar

En beaktansvärd andel uppslagningar består av vad jag kallar encyklopediska och Google-liknande sökningar. I den förstnämnda gruppen ingår bland annat sökning efter geografiska platser, namn på skådespelare och medicin- och historierelaterade begrepp:

- Geografiska platser: *Asien*, *Tanzania*, *Brunei*
- Skådespelare o.d.: *Aretha Franklin*, *Bengt Andersson* (skådespelare)
- Medicin: *alzheimer*, *leukemi*, *rohypnol*
- Historia: *Konstantin I*, *hellenism*, *Versaillesfreden*, *postmodernism*
- Övriga: *SAOL*, *Dawsons 's Creek*, *Vattuman*, *Burma* (kattras)

De Google-liknande sökningarna utgörs av flerordssökningar som liknar informationssökning, t.ex. *VM i innebandy 2006* och *kategori: orter i Bayern*, där man kan anta att användaren är ute efter information om något snarare än en definition av den sökta strängen. Andra exempel är *de fyra grundlagarna*, *borgerliga partier*, *bolibompa* och *Karolinska institutet*. Dessa typer av sökningar visar att användarna är vana att söka på det sätt som Google och andra sökmotorer tillåter.

Det verifierar också mitt tidigare påstående att användarna inte verkar vara medvetna om vilken typ av ordbok *Lexin* är och vilken målgrupp den har.

#### 5.1.4. Övriga sökningar

Andra typer av uppslagningar som är värda att kommentera är för det första en grupp övriga sökningar bestående av årtal, nummer, förkortningar och egennamn:

- Årtal och nummer: *1170-talet, 1099, 1395, 1449*
- Förkortningar: *mg/l, dna, bnp, sms, ACSE, JWST, EAPR*
- Egennamn: *Daniel, Jas*

De i sammanhanget frekventa fyrsiffriga talen framstår som en aning kryptiska. Är de årtal, koder eller något annat, och i så fall vad? Förkortningarna är av två slag, igenkännbara som *mg/l* och *dna* samt vid första anblicken icke igenkännbara bokstavskombinationer. Exempelen ovan har senare identifierats via Google.<sup>2</sup> Ett telefonnummer och tre datum återfanns också i materialet.

För det andra finns det en samling felaktigt särskrivna uttryck bland loggfilerna, men dessa är färre än vad jag förväntade mig. I de flesta fall har användaren utan framgång först sökt på uttrycket i sin hopskrivna, korrekta form, och sedan valt att särskriva det. Detta har jag tolkat som en sökstrategi och det är en av de sju strategier som jag beskriver i avsnitt 5.3.

#### 5.2. Ordbokens programmerade svar vid icke-träff

När man söker på ett ord som inte finns i ordboken, och rutan "Rättstava okända ord" är ikryssad, får man i många fall förslag på ord som man kan ha menat. Dessa förslag varierar i antal mellan ett och upp till ett fyrtiotal. En genomgång av en tredjedel av materialet ger vid handen att en inte obetydlig mängd av icke-träffarna, ca 130 stycken, får över tio ordförslag. Nedan följer exempel från de olika typer av sökningar som resulterar i höga förslagsantal:

- Existerande ord: *blogg, lakej, ninja*
- Oidentifierbara och icke-existerande ord: *frola, aaaa*
- Felstavade ord: *betuta* (jfr *betuttad* 'förtjust i')
- Engelska ord: *event, exam*
- Övrigt: *Daniel, bla*

Exempelorden i sammanställningen får ett förslagsantal mellan 30 och 40. Att ge så många förslag på vad användaren kan ha menat är näppeligen användarvänligt. Hur det ser ut om man söker på ordet *blogg* framgår av bild 2:

<sup>2</sup> ACSE står för Association Control Service Element, JWST för James Webb Space Telescope och EAPR för Euroatlantiska Partnerskapsrådet.



Bild 2. Sökning på ordet *blogg*, hämtat 2007-08-28.

Alla förslag som ges ryms inte på bilden, men scollar man neråt finner man att en sökning på *blogg* ger 40 ordförslag. Många av förslagen är dessutom, både i detta och i de flesta andra fall, ologiska. Det beror sannolikt på att rättstavningsverktyget endast är baserat på editeringsavstånd. Det innebär omkastning av två intilliggande bokstäver eller insättning, borttagning eller utbyte av en bokstav (Viggo Kann, personlig kommunikation).

### 5.3. Användarnas strategier vid icke-träff

Användarna tar till olika strategier när de inte får träff i sitt första försök. Några lämnar ordet och går inte vidare med sin sökning, och andra fortsätter att söka. Gemensamt för dem som fortsätter att söka är att de tenderar att upprepa sina sökningar, men de gör det på olika sätt. Jag har funnit sju sökstrategier, nämligen följande:

- 1) Upprepad sökning med identisk stavning
- 2) Sökning på del av ord
- 3) Sökning med annan stavning
- 4) Borttagning eller tillägg av böjningsformer/ändelser
- 5) Särskrivning
- 6) Utbyggd sökning
- 7) Kombination av flera strategier

De sju strategierna presenteras i det följande i nummerordning.

Strategi 1 är helt enkelt att upprepa sökningen med samma stavning:

- *tvärhand* → *tvärhand* → *tvärhand* → *tvärhand* → *tvärhand*
- *inskriftion* → *inskriftion* → *inskriftion*

Användaren söker i det första exemplet efter ordet *tvärhand*, ett korrekt stavat svenskt ord som dock inte ingår i *Svenska ord*. Sökningen upprepas fyra gånger. Såväl i detta fall som i exemplet *inskriftion* får användarna beskedet att ordet inte finns i ordboken men de väljer ändå att upprepa sin sökning.

Strategi 2 är att upprepa sökningen och söka på en del av ett ord:

- *enhetssträvandena* → *strävandena*
- *biosfär* → *bio* → *sfär* → *biosfär*

I det första exemplet söker användaren på ordet *enhetssträvandena*, utan resultat, och sedan på *strävandena*. Det ger en träff på ordet *strävan*. Vad gäller det andra exemplet får användaren träff på både *bio* och *sfär*, men inte på *biosfär*. Att användaren återgår till den första söksträngen, som han eller hon gör i det här fallet, är också vanligt.

Strategi 3 är att pröva en annan stavning. Först har vi de som söker på rätt stavning i sin första sökning och inte får träff. Ordet finns alltså inte i *Svenska ord*. Då prövar de en annan stavning och de börjar följaktligen stava fel:

- *desorienterad* → *desorigenterad*
- *habituell* → *hatibuell* → *habutiell*
- *klassificering* → *kalsificering* → *klassificering* → *klassificiering* → *klaceficering* → *klassifisering* → *klassifisera*

I första exemplet verkar användaren pröva en mer uttalsenlig stavning, i det andra kastas bokstäverna *b*, *t*, *i* och *u* om. Det tredje exemplet är ett av många belägg på den envishet som användarna visar prov på. Användaren får här till slut träff på ordet *klassificerar* men borde, enligt mig, kunna få träff tidigare med ett vassare rättstavningsverktyg.

I det andra fallet stavar användaren fel i sin första sökning och börjar sedan, som ovan, att pröva med att stava ordet på ett annat sätt:

- *geslong* → *geslång* → *jeslong* → *cheslomg* → *sheslong*
- *kyvär* → *kuvär* → *kyver* → *kuver* → *kuvär* → *kuvett*
- *shinettzel* → *shinettzel* → *skinettzel* → *skinetsel* → *schinettzel* → *schinetsel* → *schitsel*

Gemensamt för huvuddelen av sökningarna i den här gruppen är att orden användarna efterfrågar är svårstavade, som till exempel *schäslong*, som inte finns med i *Lexin*, och *kuvert*, som finns med och som användaren till slut får träff på. Det tredje fallet är ytterligare ett exempel på användarnas envisa beteende.

Strategi 4 innebär att man tar bort eller lägger till böjningsformer respektive

ändelser. I det andra exemplet nedan söker användaren till exempel först på particip- och preteritumformer och till sist på verbets grundform:

- *transferering* → *transfereringar*
- *internaliserade* → *internaliserad* → *internalisera*

Även här består de sökningar som inledningsvis felstavas av svårstavad ord, men den här gruppen innehåller få belägg.

De två näst sista strategierna, 5 och 6, innehåller också få exempel. För det första har vi användare som söker på ett sammansatt ord utan resultat och som då väljer att särskrivna ordet:

- *språkkunskap* → *språk kunskap*
- *innanhav* → *innan hav*

För det andra finns det de som bygger ut sin sökning:

- *tillkorta* → *tillkortakommanden*
- *kongruens* → *kongruensböjning*

Den sista strategin är att kombinera två eller flera av ovan nämnda taktiker. Det kan vara att man i sin andra sökning ändrar stavning och vid tredje försöket lägger till en pluraländelse.

I genomsnitt upprepades varje sökning tre till fyra gånger. I tabell 3 ställs strategierna upp ordnade efter frekvens och även här har procentsiffrorna avrundats. Strategi 3, 4 och 7 är uppdelade i två grupper – sökningar som är rättstavade respektive felstavade vid första uppslagningen – så som de presenterats ovan.

Strategi	Antal	Procent
Sökning med annan stavning – felstavat vid första sökningen	52	19
Borttagning/tillägg av ändelse – rättstavat vid första sökningen	50	18,5
Kombination av strategier – felstavat vid första sökningen	38	14
Sökning på del av ord	33	12
Sökning med annan stavning – rättstavat vid första sökningen	30	11,5
Kombination av strategier – rättstavat vid första sökningen	28	10
Identisk stavning	13	5
Borttagning/tillägg av ändelse – felstavat vid första sökningen	9	>3
Särskrivning	9	>3
Utbyggd sökning	9	>3
Totalt	271	100

Tabell 3. De sju strategierna ordnade efter frekvens.

De två vanligaste strategierna är för det första att pröva annan stavning i de fall man stavar fel i första sökningen och inte får träff, och för det andra att ta bort eller lägga till en ändelse i de fall man stavar rätt i första sökningen men inte får träff.

## 6. Sammanfattning och diskussion

I artikeln redogör jag för en analys av de icke framgångsrika sökningarna gjorda i Internetordboken *Lexin on-line Svenska ord*. De icke-träffar som jag ytterligare avgränsar min studie till utgör ca en tredjedel av materialet.

För de första delar jag upp icke-träffarna i olika typer av sökningar. Det visar sig att en stor mängd icke-träffar utgörs av rättstavade svenska ord, men de ingår inte i *Svenska ords* lemmauppsättning. Vidare är sökningar på fraser av olika slag och encyklopediska sökningar frekventa. Dessa resultat visar inte bara att loggfiler kan utnyttjas till att upptäcka lemmaluckor utan också ett behov av att kunna utföra fritextsökningar så att man hittar fram till sammansättningar, idiom m.m. Det visar sig också att många utländska ord återfinns i *Svenska ords* loggfiler. Det finns redan en lista på länkar till andra webblexikon via *Lexin* men dessa bör lyftas fram mer och kompletteras med fler. Det bör också finnas en lättillgänglig länk till Google och andra sökmotorer. Detta för att hjälpa användaren att hitta fram till rätt källa.

Det andra avsnittet behandlar det höga antal ordförslag som många gånger presenteras för användaren vid en icke framgångsrik sökning. Jag efterlyser här ett mer avancerat verktyg, baserat på såväl editeringsavstånd som uttal, för att kunna uppnå ett mer logiskt urval av förslag på ord. Det skulle resultera i färre antal icke-träffar och därmed fler nöjda användare.

Slutligen redovisas de sju sökstrategier användarna tar till när deras sökning misslyckas. Användarna tenderar att upprepa sina sökningar på olika sätt. De två vanligaste strategierna är att pröva annan stavning eller att använda en kombination av flera taktiker. Benägenheten att upprepa sina sökningar skulle, i värsta fall, kunna leda till att användarna börjar ifrågasätta sina språkkunskaper. Det vore mycket olyckligt i de fall brukarna faktiskt stavar rätt i sin första sökning och utgör ytterligare ett skäl till att ha ett kraftfullare rättstavningsverktyg.

Materialet består av sökningar från en dag och får anses vara relativt litet, trots att de uppgår till nästan 7 000. Jag har heller inte haft möjligheten att identifiera varje specifik användare och kan således inte med säkerhet knyta de upprepade sökningarna till en och samma användare. Men det är uppenbart att användarna har höga förväntningar på ordbokens innehåll och att de är envisa med tanke på hur de tenderar att upprepa sina sökningar. Det faktum att många sökningar på vanliga, rättstavade ord inte ger träff kan tolkas på ett positivt sätt, nämligen att användarna har en mycket stor tilltro till ordboken vad beträffar dess innehåll. Att *Lexin* fyller en viktig funktion för många användare står också klart i och med de jämförelsevis mycket höga antal uppslagningar per månad i förhållande till den danska respektive afrikansk-engelska ordboken (se kapitel 4 ovan). Det är dock ett problem att användarna så ofta inte lyckas nå den önskade informationen genom sina sökningar. Det bör dock här poängteras att det inte har varit min mening att kritisera *Lexin*. Ordboken lever upp till de mål som låg inom ramen för *Lexin*-projektets uppdrag.



Det står klart att loggfiler är ett mycket användbart redskap för att upptäcka lemmaluckor. Jag har i denna artikel visat att det finns anledning att vidare undersöka loggfilernas funktion som ett effektivt verktyg inom lexikografi och efterlyser fler studier av och med hjälp av loggfiler. Man kan diskutera hur stort inflytande användarna ska ha i det lexikografiska arbetet, men jag menar att man inte ska vare sig underskatta eller överskatta deras möjliga insatser. Användarna bör stimuleras till att ge feedback i form av nyordsförslag, definitioner m.m. och även uppmantras att kommentera hur olika sökfunktioner uppfattas och om gränssnittet är tillfredsställande. Det vore även intressant att förhöra sig om i vilken situation behovet av en ordbok uppstår och inte minst vad användarna gör med den information de får genom sina sökningar. På så sätt skulle man kunna kartlägga användarnas sökbeteende mer ingående och identifiera den verkliga målgruppen. Allt i syfte att förbättra såväl elektroniska ordböcker som pappersordböcker.

## Litteraturförteckning

### Ordböcker

Den Danske Netordbog: <<http://www.netordbog.asb.dk>>

LEXIN = *Lexikon för invandrare, Svenska ord* <<http://lexin.nada.kth.se/svenskaord>>

SAOL 13 = *Svenska Akademiens ordlista* 2006. (13:e uppl.). Stockholm: Norstedts Akademiska Förlag.

*Svenska ord – med uttal och förklaringar* 1999 (3 uppl.) Stockholm: Norstedts.

### Övrig litteratur

Béjoint, Henri 1989: The Teaching of Dictionary Use: Present State and Future Tasks.

I: *Wörterbücher, Dictionaries, Dictionnaires. First Volume*. Berlin/New York: Walter de Gruyter, 208–215.

Bergenholtz, Henning & Mia Johnsen 2005: Log files as a Tool for Improving Internet Dictionaries. I: *Hermes – Journal of Linguistics* 34, 117–141.

Bergenholtz, Henning & Mia Johnsen 2007: Log Files Can and Should Be Prepared for a Functionalistic Approach. I: *Lexikos* 17, 1–21.

Gellerstam, Martin 1999: LEXIN – lexikon för invandrare. I: *LexicoNordica* 6, 3 – 18.

Hartmann, Reinhard R. K. 1989: Sociology of the Dictionary User: Hypotheses and Empirical Studies. I: *Wörterbücher, Dictionaries, Dictionnaires. First Volume*. Berlin/New York: Walter de Gruyter, 102–111.

de Schryver, Gilles & 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.

Svensén, Bo 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. (2 uppl.) Stockholm: Norstedts Akademiska Förlag.



Anki Hult  
Fil. mag., f. 1977  
Institutionen för svenska språket  
Box 200  
SE-405 30 GÖTEBORG  
ann-kristin.hult@svenska.gu.se