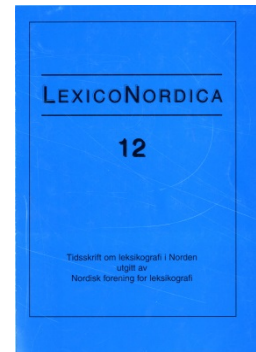


LexicoNordica

Titel: Elektroniska ordböcker i Sverige: nutid och framtid
Forfatter: Maria Toporowska Gronostaj
Kilde: LexicoNordica 12, 2005, s. 87-108
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.

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.

Maria Toporowska Gronostaj

Elektroniska ordböcker i Sverige: nutid och framtid

The article *Swedish electronic dictionaries: the present and the future* aims at a survey of the state of art concerning electronic dictionaries in Sweden. The focus is on CD-ROM dictionaries, mainly defining ones, and their functionalities of technical and conceptual character. It is assumed here that the content, structure and functionalities of the electronic dictionaries of today can make a relevant contribution to designing the truly electronic dictionaries of the future. Some proposals for extending the content of the next generation dictionaries are suggested. The question of using a more sophisticated structural and functional access to the information is also discussed.

1. Inledning

The past is print dictionaries; the present is print dictionaries with some electronic versions of the same text; the future must be print dictionaries and truly electronic dictionaries, compiled afresh for the new medium, enriched with new types of information the better to meet the needs of multifarious users. (B.T.S. Atkins 2002:1)

Lexikografin befinner sig i en process av turbulent utveckling och omvandling. Ordbokslandskapet förändras i takt med den digitala teknologins frammarsch. Det förflutnas pappersordböcker korsbefruktas med det digitala mediet och resultatet kommer att bli multimediala ordböcker som kan tillmötesgå de mest mångfasetterade sökbehov (Atkins 2002). Eftersom den digitala teknologin eliminerar pappersordbokens utrymmesbegränsningar och skapar nya kanaler för informationsöverföring, möjliggörs en radikal kvantitativ och kvalitativ modernisering av dagens lexikografiska produkter både när det gäller innehåll, presentation samt sök- och åtkomstfunktioner. I uppsatsen ges en kort översikt över det svenska utbudet av elektroniska ordböcker (i fortsättningen e-ordböcker) samt deras funktionaliteter. Förtur ges åt de funktionaliteter som bär fram det lexikaliska innehållet, eftersom det främst är dessa som stimulerar övergången från dagens digitaliserade versioner av pappersordböcker till framtidens äkta e-ordböcker.

Översikten begränsas till svenska e-ordböcker utgivna på cd-rom under de senaste tio åren och behandlar i första hand definitionsordböcker för svenska. Tvåspråkiga ordböcker och deras problematik ges något mindre utrymme. Att fokus ligger på e-ordböcker i cd-rom-

versioner beror främst på att deras form och innehåll har permanentats till skillnad från online-ordböcker vars innehåll oftare blir föremål för revision.

Artikeln disponeras på följande sätt: I avsnitt 2 uppmärksammas e-ordböckers allmänna egenskaper och i avsnitt 3 ges en översikt över utbudet av svenska e-ordböcker. Frågeställningar rörande informationsåtkomst med exempel hämtade från en rad svenska ordböcker presenteras i avsnitt 4. Övergången från digitaliserade pappersordböcker till elektroniska originalordböcker diskuteras i avsnitt 5. Där diskuteras också riktlinjer för utvecklingen av framtida e-ordböcker.

2. Allmänt om e-ordböcker

2.1. E-ordböckers typologi

Den snabba utvecklingen på IT-området utgör en viktig orsak till det växande utbudet av e-ordböcker och e-lexikon. Med termen *elektroniska ordböcker* avses konsultationsordböcker i ett digitaliserat format som används interaktivt av människor (Svensén 2004:517). Det kan noteras att denna term står i viss kontrast till termen *elektroniska lexikon* som syftar på lexikonmoduler avsedda för användning inom språkteknologin, t.ex. maskinöversättning eller informationshantering. Vid informationshantering som baseras på icke-statistiska metoder är tillgång till formaliserad lexikalisk kunskap oundgänglig, huvudsakligen för att kunna lemmatisera samt morfologiskt, syntaktiskt eller semantiskt tagga texter i syfte att utvinna information. Skillnaden mellan de två olika elektroniska produkterna ligger främst i graden av formalisering samt i olikartade förväntningar gällande gränssnittet. Det underliggande lexikala innehållet i dessa e-produkter kan dock mycket väl bygga på en gemensam kunskapsbas, vilket kan ge synergieffekter.

Utöver den ovan nämnda distinktionen finner man hos Svensén (2004:518) en ytterligare indelning av e-ordböcker, nämligen *digitaliserade pappersordböcker* och *digitaliserade originalordböcker*. Medan de förra utgör digitaliserade versioner av pappersordböcker är de senare producerade som e-ordböcker redan från början. En ordbok som tillhör den senare kategorin kan anknyta till papperstraditionen eller vara mer innovativ.

Låt oss även uppmärksamma några termer ur en typologi över e-ordböcker som är förankrad i hur informationen lagras, nämligen *offline-ordbok*, *intranät-ordbok* eller *online-ordbok* (Svensén 2004:517). En offline-ordbok finns lagrad hos den enskilde användaren på

persondator eller i form av en elektronisk fickordbok eller tal-/läspenna. En intranät-ordbok lagras på en lokal server och görs tillgänglig via organisationens eller företagets nät. En online-ordbok konsulteras vanligen via Internet och informationen kan nå antingen via en persondator eller via en mobiltelefon. Med tanke på att visualiseringen av ordböckers innehåll på mobiltelefonens, fickordbokens eller tal-/läspennans display skiljer sig väsentligt från dataskärmens fullformat spelar storleksfaktorn en viktig roll vid utarbetandet av ett lämpligt gränssnitt samt vid urvalet av lämpliga funktionaliteter för de skilda mottagningsmedierna.

2.2. Allmänna egenskaper hos e-ordböcker

Som redan antytts skiljer sig e-ordböcker i flera avseenden från pappersordböcker. Till de mest framträdande egenskaperna hör: kraftigt utökade disponibla utrymmen, en mångfald sök- och åtkomstmöjligheter till innehållet och multimedial presentation och användning (Svensén 2004:518–527).

I dagens läge råder det ingen brist på utrymme. Lagringskapaciteten på en cd-romskiva uppgår till i genomsnitt 650 megabyte och på en dvd till 17 gigabyte. Tillgången till det så gott som obegränsade utrymmet möjliggör inkludering av multimediala tekniker som ljud, bild och film som ett komplement till det egentliga ordboks innehållet. Vidare kan det digitala utrymmet på en cd-rom eller dvd utnyttjas till att samla olika e-produkter på en och samma skiva och göra dem simultant sökbara med hjälp av ett gemensamt sökprogram.

Det elektroniska mediet tillför nya sök- och åtkomstfunktioner, som tillåter användaren att:

- (i) utvinna information ur ordboksartiklar på ett selektivt sätt genom att välja någon av de av programmet fördefinierade sökfunktionerna eller formulera egna söksträngar som består av kombinationer av dessa sökfunktioner. Det kan exempelvis handla om teckensträngssökningar eller innehållsbaserade sökningar.
- (ii) direkt komma åt andra ordboksartiklar i samma ordbok med hjälp av *hypertextlänkar* (*hyperlänkar*), dvs. pekare som fungerar som länkar från ordbokens textord till respektive lemman bland uppslagsorden. (Texten med sådana länkar tillgängliga kallas för *hypertext*.) Denna hypertextfunktion är

ett komplement till de konventionella sökningarna i e-ordböcker i vilka sökordet anges i sökrutan.

- (iii) från ordbehandlingsdokument eller Internet direkt komma åt uppslagsfunktion. Uppslagsfunktionen levererar betydelseförklaring eller översättning ur den på förhand förvalda ordboken som finns installerad på datorn i ett följfönster (miniatyrfönster) eller programfönster, så fort man pekar på det sökta ordet.

(Mer om sök- och åtkomstfinesser i avsnitten 4 och 5.)

Tillgång till kombinationen av multimediala tekniker med hypertext, som benämns *hypermedia*, öppnar helt nya presentationsmöjligheter och därmed användningsområden. Medan användargränssnittet för presentationen av sökträffar är på förhand bestämt och dess utformning växlar beroende på söktypen, kan vissa moment i användargränssnittet, t.ex. fontstorlek, samgruppering av e-ordböcker, fördefiniering av sökslingor i dessa m.m., individuellt ställas in av ordboksanvändaren. Användningsområdena är mycket flexibla och kan efter sökbehov växla mellan reception, produktion och edutainment. I moderna e-ordböcker skapas förutsättningar för diverse övningsmöjligheter. I t.ex. LDOCE, MED och CALD kan man träna såväl uttal som lexikala färdigheter. I *Norstedts första engelska ordbok* finns spel som lär ut och aktiverar språkliga färdigheter.

Allt detta förstärker den interna dynamik och potential som finns inneboende i e-ordböcker.

3. Om utbudet av e-ordböcker i Sverige

Detta avsnitt ägnas åt en översikt över olika typer av e-ordböcker som finns tillgängliga på den svenska marknaden. Tyngdpunkten kommer att ligga på cd-romordböcker som numera ofta samsas med det snabbt växande utbudet av fritt tillgängliga online-ordböcker. För att ge en uppfattning om det totala utbudet presenterar jag först en översikt över några on-lineordböcker för att sedan gå över till cd-romordböcker. Särskilt uppmärksammas de svenska fritt tillgängliga online-ordböckerna med förlagor i det tryckta bokformatet. Dessa digitaliserade ordböcker har en viss ”kvalitetscertifiering” av det lexikaliska innehållet i kontrast till produkter som skapas on-line av Internets användare. Som tecken på den ständigt växande efterfrågan på fritt tillgängliga online-ordböcker kan ses dels det ständigt växande antalet nya ordboksprojekt som bedrivs på ideell basis, dels den statistik som

dokumenteras i loggböcker över antalet sökningar per ordbok. Den senare finns tillgänglig för bl.a. ordböcker i LEXIN:s serie. Viljan att dela sina material med andra gör att Internetpublicering snabbt ökar i antal, vilket bidrar till kunskapens spridning och demokratisering.

Ordböcker ur LEXIN-serien (se Gellerstam 1999:3–18) hör till on-lineresurser som Myndigheten för skolutveckling i samarbete med Kungliga tekniska högskolan (KTH Nada) fritt distribuerar via Skoldatanätet. Bland dessa finner man den enspråkiga inlärningsordboken *Svenska ord* och nio tvåspråkiga ordböcker från svenska till något av följande språk: engelska, finska, ryska, bosniska, kroatiska, turkiska, albanska, grekiska och spanska. Kännetecknande för de nät-baserade LEXIN-ordböckerna är att man kontinuerligt tillför nya funktionaliteter. Som exempel på tekniska finesser som finns att tillgå i on-lineversionen av *Svenska ord* kan nämnas: (i) möjlighet att söka efter uppslagsord på både grundformen och de i ordboken specificerade böjningsformerna, (ii) tillgång till inspelat uttal för uppslagsorden i *Svenska ord* och (iii) möjlighet till en inverterad textsökning i de tvåspråkiga ordböckerna som delvis kompenserar avsaknaden av tvåspråkiga lexikon som översätter från L2 till L1. Med andra ord kan sökningen utgå från L2 och returnera artiklar ur L1 om sökordet påträffas i L1.

Som exempel på några andra fritt tillgängliga online-ordböcker med förlagan i en pappersordbok kan nämnas:

- *Svenska Akademiens ordbok* (SAOB) (<http://www.saob.se>) en historisk ordbok som beskriver svenskt skriftspråk från 1521 till våra dagar och för närvarande täcker ord med begynnelsebokstäverna till T.
- *Bonniers synonymordbok* (<http://www.synonymer.se>) utarbetad av G. Walter, består av ca 30.000 uppslagsord med ca 190.000 synonymer och fraser i den svenska databasen. Dessutom innehåller den ca 1000 antonymer.
- *Norstedts skandinaviska ordbok* (<http://www.nordskol.org/ordbog>), utarbetad av Birgitta Lindgren, Skirne Helg Bruland, Allan Karker och Ståle Løland, med 3000 ord på svenska, norska och danska som kan vålla bekymmer vid kontakter mellan skandinaviska länderna.

Mellan de fritt tillgängliga online-ordböckerna och dem för vilka en prenumeration krävs finns ett antal Internetpublicerade ordböcker som intar en mellanställning beträffande fri åtkomst till materialet. Dit hör

de ordböcker som kan nås gratis från lokala nätuppkopplingar, t.ex. via bibliotek. Som exempel på en sådan resurs tillgänglig för alla Göteborgs universitetsbiblioteks boklånstagare kan nämnas *Nationalencyklopedins ordbok* (NEO) och *Engelska ordboken* med ingångar svenska resp. engelska vilka finns tillgängliga från *Nationalencyklopedins* (NE) hemsida, där även encyklopedin görs sökbar. För övriga användare ingår dessa två ordböcker i den prenumeration som tecknas för NE.

Bland de ca 50 på cd-rom utgivna e-ordböckerna hittar vi både allmänna enspråkiga och tvåspråkiga ordböcker och ett antal en- och tvåspråkiga fackordböcker. Nedan listas några av dem med information om utgivningsåren för deras första och senaste elektroniska upplaga, om deras förlaga och storleksordning.

- *Nationalencyklopedins ordbok* (NEO) ((1997), 1999) utgår från pappersversionen utarbetad av Språkdata och utgiven av Bra Böcker. NEO är en definitionsordbok över det nutida språket med ca 62.000 uppslagsord och ca 137.000 språkliga enheter definierade i den första upplagan. Den andra upplagan utökades med nya ord. Cd-rom:ens överlägsenhet består främst i möjligheten att utföra avancerade innehållsbaserade sökningar, i jämförelse med NEO:s pappersversion och även andra senare tillkomna e-ordböcker (se avsnitt 4.4 för beskrivningen av dessa). Cd-romversionen innehåller utöver all den information som finns i bokverket även information om de ämnesområden uppslagsorden hör till samt inspelade uttalsangivelser för ca 12.000 ord. (För recension av NEO:s cd-version se Martola 1998).
- *Norstedts svenska ordbok* (SO) (1997 i *Skribent*, 2004 i *Stora Skribentordboken*). Den senaste cd-upplagan är utgiven av Norstedts Akademiska Förlag i samarbete med WordFinder. Till grund för alla upplagor ligger *Svensk ordbok*, 3:e upplagan, utarbetad av Språkdata. Det är en definitionsordbok med ca 100.000 ord och fraser. Den senaste e-upplagan erbjuder en mängd nya sök- och åtkomstmöjligheter. Några av dem beskrivs i avsnitt 4.
- *Svenska ordboken* (2001) utgår från en pappersversion med samma titel författad av Bengt Oreström och utgiven av Gustavas förlag. Den innehåller över 40 000 uppslagsord, uttryck och fraser och täcker svenskans centrala och allmänna ordförråd. I cd-romversionen finns utöver traditionella strängebaserade sökmöjligheter också unika möjligheter till

ljusbildssökning med felstavningskompensering. Ett av målen med ordboken har varit att mjukvaran också skulle passa dyslektiker.

- *Svensk synonymordbok* (1997 i *Skribent*, 2004 i *Stora skribentordboken*). Det är fjärde upplagan av *Ord för ord* som står för innehållet. Den senaste versionen är sökbar med WordFinders programvara.
- *Norstedts stora engelska ordbok* (1997 i *Skribent*, 2004 i *Language Suite*) har utkommit i åtskilliga cd-upplagor som är digitaliserade versioner av ordbokens pappersversion vilken innehåller mer än 270.000 ord och fraser. Den senaste versionen är sökbar med WordFinders programvara, som används också som sökmotor i *Stora skribentordboken*.
- *Engelsk ordbok*, utgiven av förlaget Gustava, finns tillgänglig i både cd-rom- och nätversion. Nätversionen kan nå från Nationalencyklopedins hemsida (www.ne.se). E-ordboken innehåller en engelsk-svensk och en svensk-engelsk ordbok och är baserad på den 7:e pappersupplagan. Boken innehåller cirka 95.000 ord, uttryck och fraser. Den har ett inbyggt program som korrigerar felstavningar.
- *Norstedts första engelska ordbok* (2004) är en multimedial tvåspråkig ordbok avsedd för yngre elever, utarbetad av Vincent & Kerstin Petti. Den är baserad på *Norstedts första engelska ordbok* i bokform. Målgruppen är nybörjare, dvs. barn i åldern 9–13 år. Den innehåller utöver den lexikala informationen bilder, animationer och inspelat uttal för alla engelska uppslagsord. Det finns totalt 33.000 uppslagsord och fraser. För vissa ord ges utöver den konventionella informationen om uttalet, ekvivalenter och exempel även förteckningar över ord som är semantiskt besläktade, ofta i kombination med grafisk presentation (se figur 4.3.1). Programmet har ett gränssnitt som ger stöd även för användare med funktionshinder.

Bland övriga digitaliserade tvåspråkiga ordböcker finner man tio språk representerade: engelska, tyska, franska, spanska, italienska, holländska, norska, danska, finska och svenska. Alla dessa är kopplade till WordFinders programvara, vilket bl.a. innebär att man kan bygga grupper av lexikon med fördefinierade sökslingor. Dessutom finns det med samma programvara ca 20 fackordböcker i cd-romversion för olika

områden, t.ex. teknik, affärsliv, data och medicin, som är digitaliserade versioner av tidigare pappersupplagor.

4. Sökfunktioner och informationsåtkomst

Informationsåtkomst förutsätter tillgång till effektiva teckenbaserade och innehållsbaserade sökfunktioner å ena sidan och ett välstrukturerat användargränssnitt å andra sidan. Med exempel hämtade främst ur svenska e-ordböcker belyses nedan de allra viktigaste sökfunktionerna med fokus dels på teckensträngsorienterade, dels på innehållsbaserade sökfunktioner. Avsnittet avslutas med en översikt över övriga åtkomstfinesser.

Det övergripande syftet med denna genomgång är att dokumentera de ur åtkomstsynpunkt centrala sökfinesserna för att värna om progression i framtida e-ordböcker.

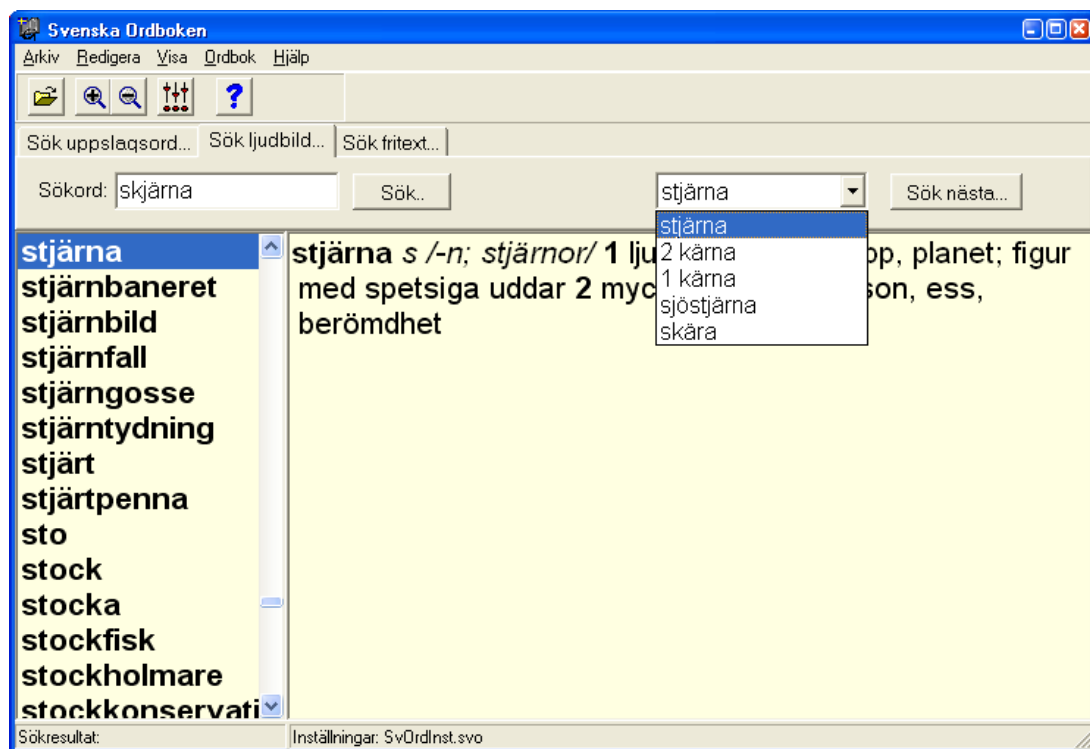
4.1. Teckensträngsbaserade sökningar

Bland de teckenbaserade sökfunktionerna kan man urskilja två huvudtyper med avseende på sökningens omfång, dels de som är förlagda till enskilda fält i mikrostrukturen, t.ex. uppslagsordsfältet, dels de som tar hela ordbokstexten som sökningsfält, fritextsökning. Dessutom kan sökumfånget utvidgas till att omfatta ordboksgrupper. Vid dessa sökningar kan sökordet eller sökfrasen vara antingen en fullständig eller delvis specificerad söksträng, dvs. utan eller med något ersättningstecken. Som stöd för direkt lokalisering av söksträngen i enskilda artiklar vid fritextsökningen markeras den med avvikande färg. I vissa e-ordböcker (NEO, SO: Skribent) ges också numerisk information om antalet funna belägg.

Såväl teckensträngs- som fritextssökfunktioner tillhör numera det basutbud som erbjuds i alla nyproducerade e-ordböcker. Det intressanta med möjligheten att göra fritextsökningar i ordbokstexten är att ordboken utöver sin vanliga funktion även kan återanvändas som korpus. Detta gynnar ordboksanvändaren genom att han/hon dels kan söka bekräftelse på den information som ges i artikeln, dels få den ytterligare belyst och exemplifierad i ordbokskorpusen. Lexikografen å sin sida får möjlighet till en extra kontroll av innehållet i ordboken. För att göra åtkomsten av den typen av information ännu mer

användarvänlig kan man komplettera den med en möjlighet att visualisera resultaten i KWIC-konkordansformat. KWIC-visualisering finns som tillval i t.ex. *CD-Perussanakirja* (Haarala & Lehtinen 1998) och *Kielitoimiston sanakirja* (Romppanen 2005) för fritextsökningar.

En vidare pejling av utbudet av strängrelaterade sökmöjligheter i svenska e-ordböcker resulterar i ett annat intressant fynd skapat främst med tanke på de användare som är osäkra på sökordets stavning. Genom att välja optionen ”ljudbilder” i Gustavas *Svenska ordboken* får man, när man har skrivit ett fåtal bokstäver i ett ord, förslag till potentiella uppslagsord (figur 4.1.1), vars definitioner kan undersökas snabbt. I och med att ljudbildsökningen fyller en nyckelfunktion för att komma åt innehållet i en ordbok skulle den vara en värdefull komponent i alla e-ordböcker oavsett målgrupp.



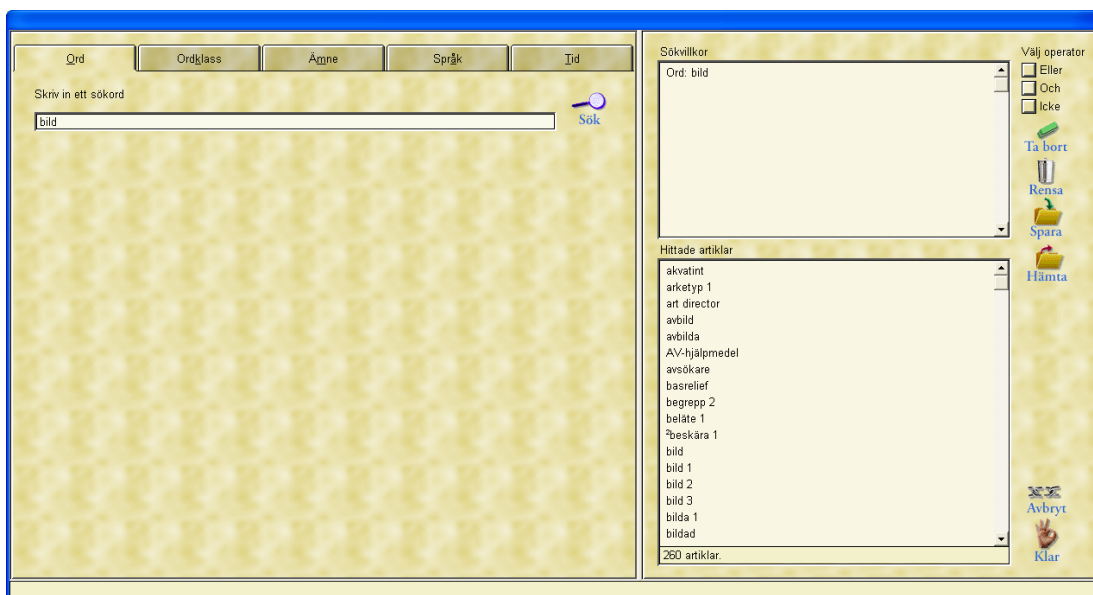
Figur 4.1.1. *Svenska ordboken* med ljudbildsökningen aktiverad

Det kan noteras att en sökfunktion som påminner om ljudbilder finns med i *Macmillan English Dictionary* (MED), där uttalsnotationen kan användas som ett alternativ för att komma åt det sökta ordet om man inte kan stava det korrekt. Denna åtkomstfacilitet är utan tvekan värdefull inte minst i samband med inläringen av ett främmande språk vars stavning inte är helt fonetisk, vilket är fallet med t.ex. engelska och svenska.

4.2. Innehållsbaserade sökningar

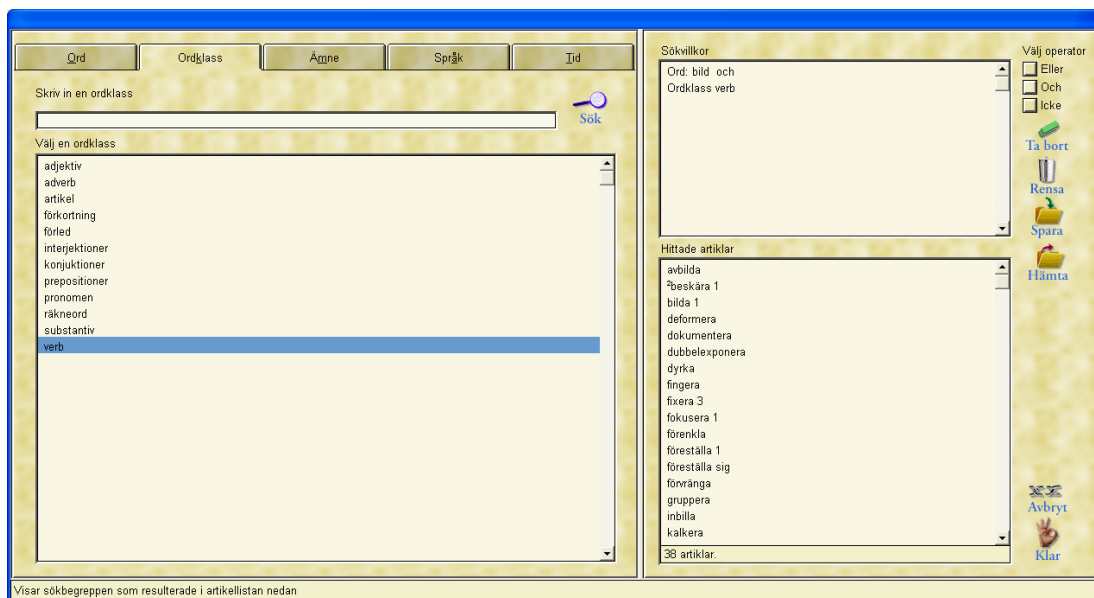
Vid innehållsbaserade sökningar söker man ord som man vet finns (och kanske har på tungan), men inte kan komma på. I NEO:s cd-romversion kan en sådan sökning snabbt genomföras med stöd av en uppsättning fördefinierade typer av sökkriterier (Malmgren 1999). Till de sökkriterier som ställs till användarens förfogande hör bl.a. ord som ger möjlighet att avgränsa det sökta ordets semantiska fält genom att man anger dess hyperonym, närsynonym eller andra besläktade ord, ordklass som det sökta ordet hör till och dess ämnesområde valt från någon eller några av de 90 listade domänspecifikationerna. Ordboksanvändaren kan med dessa sökkriterier samt Booleska operatorer ELLER, OCH och INTE beskriva det sökta ordet och sökprogrammet levererar det ordet i många fall eller en begränsad lista av möjliga kandidater.

Med hjälp av fyra skärmdumpningar illustreras nedan de successiva stegen i sökningen och den selektionsprocess som sökprogrammet utför. Det aktuella ordet som söks är verbet *retuschera*. De ca 62.000 uppslagsord i NEO:s cd-romversion granskas och gallras bort utifrån följande söksteg. Först anges begreppets semantiska fält eller dess ontologiska association som i det här fallet är "bild". Sökningen resulterar i 260 förslag till uppslagsord vilka listas i den högra kolumnen (se bilden nedan).

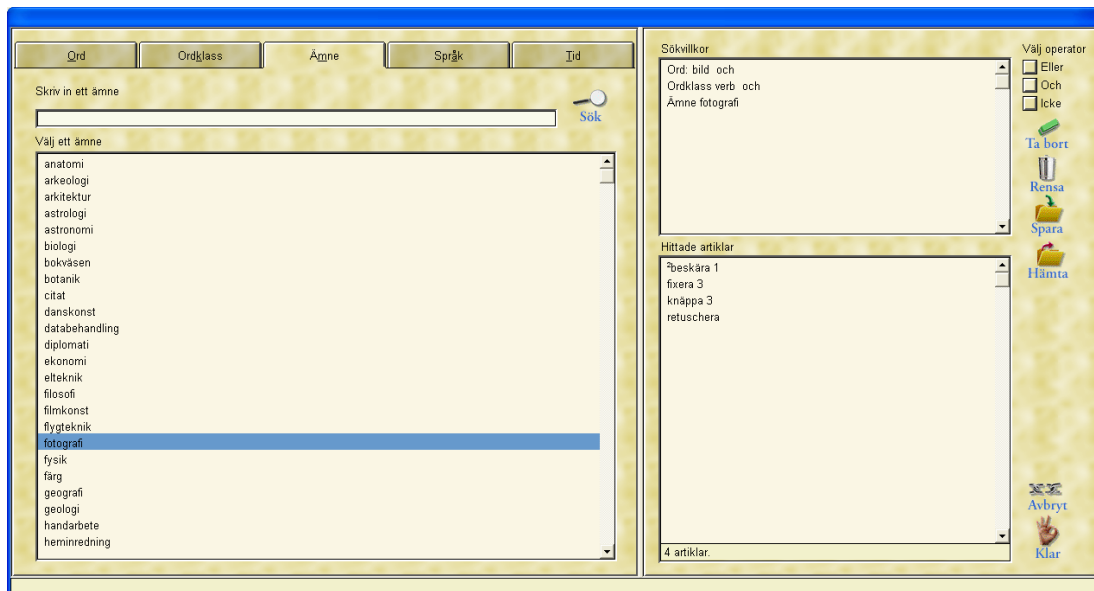


Figur 4.2.1. NEO: sökning efter *bild*-relaterade uppslagsord

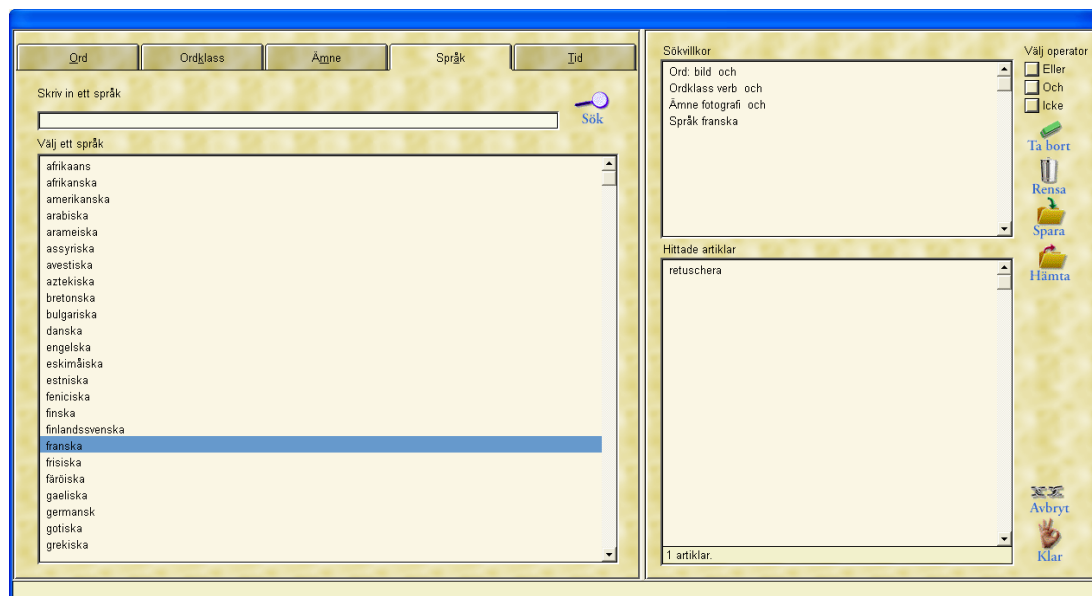
Denna lista reduceras vidare genom att man i kategori ordklass väljer verb, varefter 38 av 260 förslag blir kvar (Figur 4.2.2). Dessa sökningar samt de två efterföljande utförs med operatoren OCH. Genom att avgränsa sökordets ämnesområde till "fotografi" får man en lista på 4 uppslagsord av vilka bara ett är ett franskt lånord, vilket ger det lemma man hade på tungan vid söktillfället (Figur 4.2.3 & Figur 4.2.4).



Figur 4.2.2. NEO: avgränsning av sökfältet till verb



Figur 4.2.3. NEO: avgränsning av sökfältet till ämnet fotografi



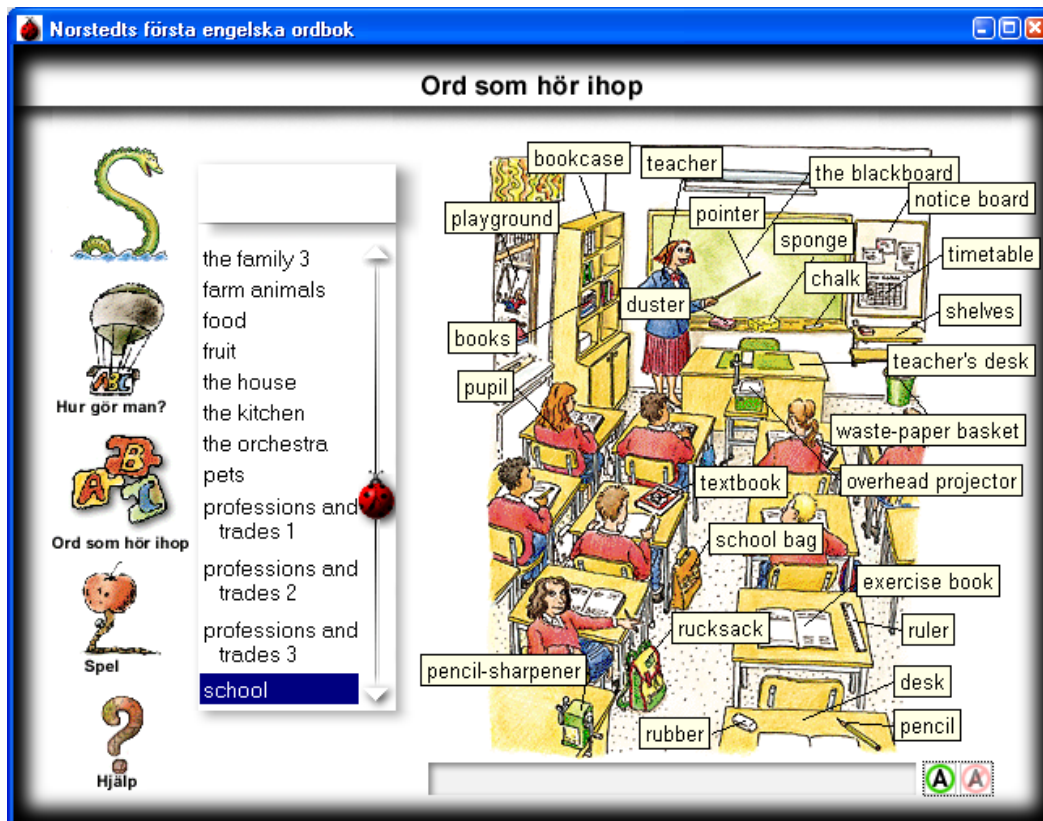
Figur 4.2.4. NEO: avgränsning av sökfältet till lånspråket franska

Denna bildserie visar inte bara hur en möjlig avgränsning av begreppen går till, utan illustrerar även en tesauristisk åtkomst till information som kan genereras utifrån den information som finns i NEO, särskilt då ord- och ämnesområdesökningen anropas.

Smarta sökningar av denna typ skapar också förutsättningar för att aktivera användarens passiva lexikaliska kunskap och på ett effektivt sätt stimulera språkinläringen genom att låta användaren upptäcka nya namn på kända begrepp.

4.3. Andra åtkomstfinesser: uttal, bilder, animationer, hypertext och interfacing

Den multimediala tekniken blir allt mer viktigare i moderna engelska och svenska e-ordböcker. Inspelat uttal av uppslagsord, talsyntes för uppläsning av artikeltexter, bilder, animationer samt ljud- och film-sekvenser kompletterar allt oftare det egentliga lexikaliska innehållet. I *Norstedts första engelska ordbok* ges inspelat uttal för alla engelska uppslagsord och ett flertal av dessa illustreras med bilder eller animationer. Bilder med tematiskt grupperade föremål är försedda med talande pop-upetiketter, vilket förbättrar såväl informationsåtkomsten som edutainmentfaktorn.



Figur 4.3.1. *Norstedts första engelska ordbok*

Det kan noteras att redan cd-romversionen av NEO (1997, 1999) innehöll inspelat uttal för ca 12 000 ord som ansågs kunna vålla problem, bland dessa många lånord. I moderna engelska inlärningsordböcker, som t.ex. LDOCE och CALD, får ordboksanvändaren inte bara inspelat uttal för alla ord utan även en möjlighet att träna sitt uttal. Dessutom läses orden upp då ordboken används i dess uppslagsfunktion vid läsningen av externa texter. Uttalsangivelser finns också för både engelska och franska ord i Collins-Roberts tvåspråkiga e-ordbok.

Navigering med hjälp av hypertextlänkar är numera en standard-facilitet och omfattas normalt hela ordbokstexten inklusive ord i böjda och oböjda former. För att sådana länkar ska fungera måste texten antingen på förhand vara lemmatiserad eller ha tillgång till en lemmatiserare. För flertalet heterografa ord fungerar hypertextlänkar på ett tillfredställande sätt. Problem blir det däremot med homografer vars länkar kan peka på fel uppslagsord och dessutom på olika sådana i olika e-ordböcker, vilket kan bero på att olika lemmatiseringsprogram har använts. Det länkade uppslagsordet och dess tillhörande text kan antingen visas i huvudfönstret eller i ett kompletterande programfönster i ordböcker med WordFinders sökprogram (se figur 5.4 för ett gränssnitt med programfönstret). Den senare lösningen med simultan

visning av båda artiklarna tillämpas i ett flertal nyproducerade en- och tvåspråkiga e-ordböcker.

Hypertextlänkar ligger också bakom inverterade sökfunktioner i tvåspråkiga ordböcker. Simultan tillgång till inverterade sökningar möjliggör jämförelse av information i produktions- och receptionsordböcker, vilket ger ett fylligare urval av exempel och fraser.

Uppslagsfunktionen med interfacing är en annan funktion som tillhandahålls i e-ordböcker och som automatiskt slår upp ordet ur en dokumenttext i den förvalda e-ordboken lagrad på datorn eller server. Faciliteten som sådan är användbar, men den kräver en bättre anpassning av många aktuella e-ordböcker. Som ett första steg i denna anpassning kunde listningen av homografa strängar göras frekvensbaserad, vilket skulle hjälpa till att undvika olika fallgropar, t.ex. när det som översättning av det engelska hjälpverb *must* i första hand föreslås substantivet 'must', 'drumust'.

5. Övergången från digitaliserade pappersordböcker till elektroniska originalordböcker

Digitaliserade pappersordböcker kan ses som ett förstadium till elektroniska originalordböcker med hypertext och multimediala finesser som bas. Det mesta av den information som finns i digitaliserade ordböcker kommer att finnas i framtida originalordböcker fast i en utvidgad och reviderad skepnad. En snabb genomgång av artikelns mikrostruktur i en definitionsordbok visar att en mer effektiv och heltäckande åtkomst till lexikalisk information är önskvärd i framtida e-ordböcker. Här är några förslag till kvantitativa och kvalitativa utvidgningar och kompletteringar av innehållet i de svenska en- och tvåspråkiga e-ordböcker.

Uttalsangivelser. Uttalsangivelser för uppslagsord bör ingå i basutbudet. Tillgång till inlästa syntaktiska exempel och fraser där uppslagsordet förekommer i kontext är en lockande möjlighet. Som ett extra tillval kan finnas en uppläsning av hela artikeltexten med tal-syntes, en nyttig facilitet för användare med funktionshinder.

Morfologi. Avstavningsinformation, kompletta böjningsparadigm för varje uppslagsord, förekommande avledningstyper samt information om ords förmåga att bilda olika semantiska sammansättningstyper bör finnas tillgängliga.

Syntax. Valensmönster med direkta länkar till berörda exempel, utförlig listning av partikelverb samt systematisk presentation av kollokationer och idiom hör till de viktigare önskvärda utvidgningarna av

innehållet. Projicering av valensmönster på de medföljande exemplen vore en användarvänlig facilitet sett ut åtkomstsynpunkt. Den skulle underlätta återanvändning av materialet vid datorstödd undervisning.

Semantik. Information om uppslagsordets plats i ett semantiskt fält, dess tillhörighet till ämnesområde och information om dess tematiska associationer skulle ge en djupare insikt i ordets betydelse. Ju fler sökkategorier det finns, desto mer nyanserade sökningar kan formuleras, vilket leder till bättre åtkomst till materialet. Införandet av betydelseindikatorer för polysema ord samt bättre utnyttjande av metaspråklig information skulle förkorta processen och leda användaren rätt till det relevanta betydelsemomentet.

Pragmatik. Den elektroniska ordboken kunde också vara mer explicit beträffande språkets sociokulturella aspekter och därmed utöver pragmatiska uppgifter om lämpligt språkbruk inkludera även notiser och hänvisningar som belyser kulturella och sociala traditioner och organisationer i Sverige. De senare dyker ibland upp i syntaktiska exempel i dagens ordböcker. Upplysningar om ords frekvens i tal- och skriftsspråk skulle också vara värdefulla.

Sådana informationskategorier i framtida e-ordböcker skulle innebära radikala förbättringar av det befintliga användargränssnittet i svenska e-ordböcker som än så länge alltför troget följer bokformatet. Det nya användargränssnittet i kombination med hypertext kan förväntas framhäva e-ordbokens struktur och göra den lätt navigerbar. Det finns redan ett antal intressanta förlagor och inspirationskällor till hur det nya gränssnittet kan utformas, t.ex. i LDOCE, CALD, MED samt Collins-Roberts ordbok för franska och engelska.

Låt oss ta en titt på LDOCE:s gränssnitt (se figur 5.1).

Valet mellan modulerna Dictionary, Activator och Exercises framgår tydligt. I fältet strax över uppslagsordet visas ett antal sökmöjligheter. En klick på "Menu" resulterar i ett programfönster som listar indikatorer för de tio betydelsemoment för ordet *cup* som finns beskrivna där.

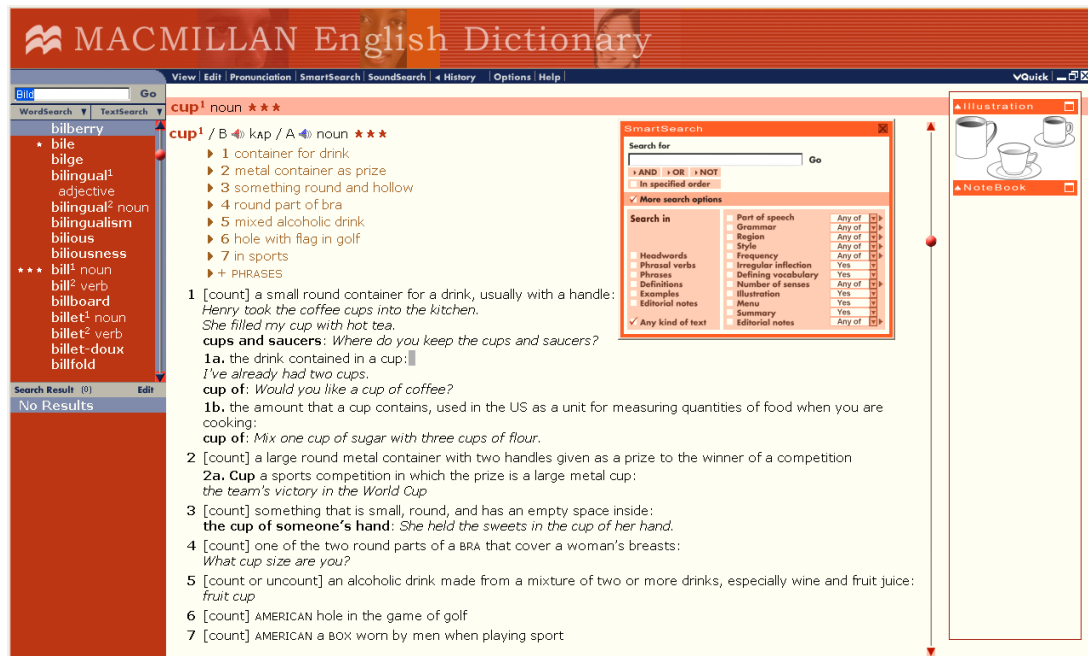


Figur 5.1. LDOCE och länk till Activator längst till vänster

Huvuddelen av fönstret upptas av betydelsebeskrivningar som inleds med betydelseindikatorer och följs av definitioner vilka illustreras med exempel. Betydelseindikatorerna fångar på ett koncist sätt det centrala i betydelsen och strukturerar samtidigt artikeln. Beläggen i frasbanken har länkar till relevanta betydelsemoment, vilket skapar snabba genvägar till artikeln. Från kollokationerna finns det länkar till exempelbanken, där det första relevanta exemplet markeras. Som ett alternativ till löpande exempeltext i exempelbanken kan KWIC-formatet väljas. Allt detta framhäver kärnan i artikeln, som efter behov kan kompletteras med andra tillgängliga typer av information som uttal, etymologiska uppgifter, relevanta tematiska områden samt frekvensplacering bland de 10.000 mest frekventa orden i tal- och skriftspråk. Vidare är LDOCE:s betydelsemoment länkade till relevanta delar av begreppsbeskrivningar i ACTIVATOR, vilket ytterligare förankrar begreppet i dess konceptuella nät. Dock saknas det i LDOCE en möjlighet att utföra kombinerade selektiva innehållsbaserade sökningar i stil med dem som finns i t.ex. NEO och MED.

MED ger möjlighet för användaren att formulera selektiva innehållsbaserade sökningar som utgår från artikelns mikrostruktur.

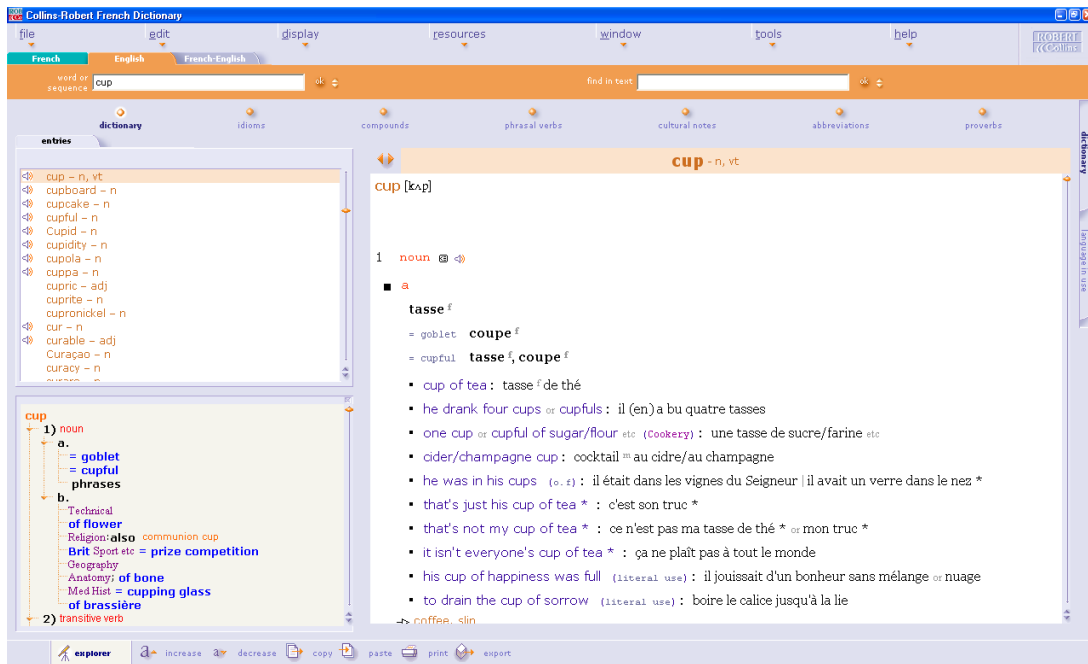
Informationen i respektive fält finns sökbar var för sig eller i meningsfulla sökkombinationer som kan bestå av ett sökuttryck (t.ex. *musical instrument*) och en uppsättning sökkriterier som formuleras med Booleska operatörer.



Figur 5.2. MED med tillval av SmartSearch

Bland sökkriterierna i SmartSearch finner man följande: uppslagsord, partikelverb, fraser, definitioner, exempel, redaktionella kommentarer, ordklass, grammatik, engelskans geografiska område, stil, frekvens, oregelbundna böjningsmönster, definitionsvokabulär, antal betydelser, illustrationer, menusökning och sökning bland brukskommentarer. För många av dessa kategorier finns det på förhand listade särdrag eller subklasser med uppgift som t.ex. i att leda användarna till de mest adekvata sökträffarna.

Vikten av strukturering framträder tydligt i såväl en- och tvåspråkiga e-ordböcker. I tvåspråkiga e-ordböcker ställs man bl.a. inför frågan av en lämplig presentation av berörda språkmoduler. I cd-versionen av *Collins-Roberts French Dictionary* har färgen använts på ett mycket skickligt sätt för just detta och även andra ändamål (se figuren nedan, som tyvärr i svart-vitt tryck missar en del av dessa finesser).

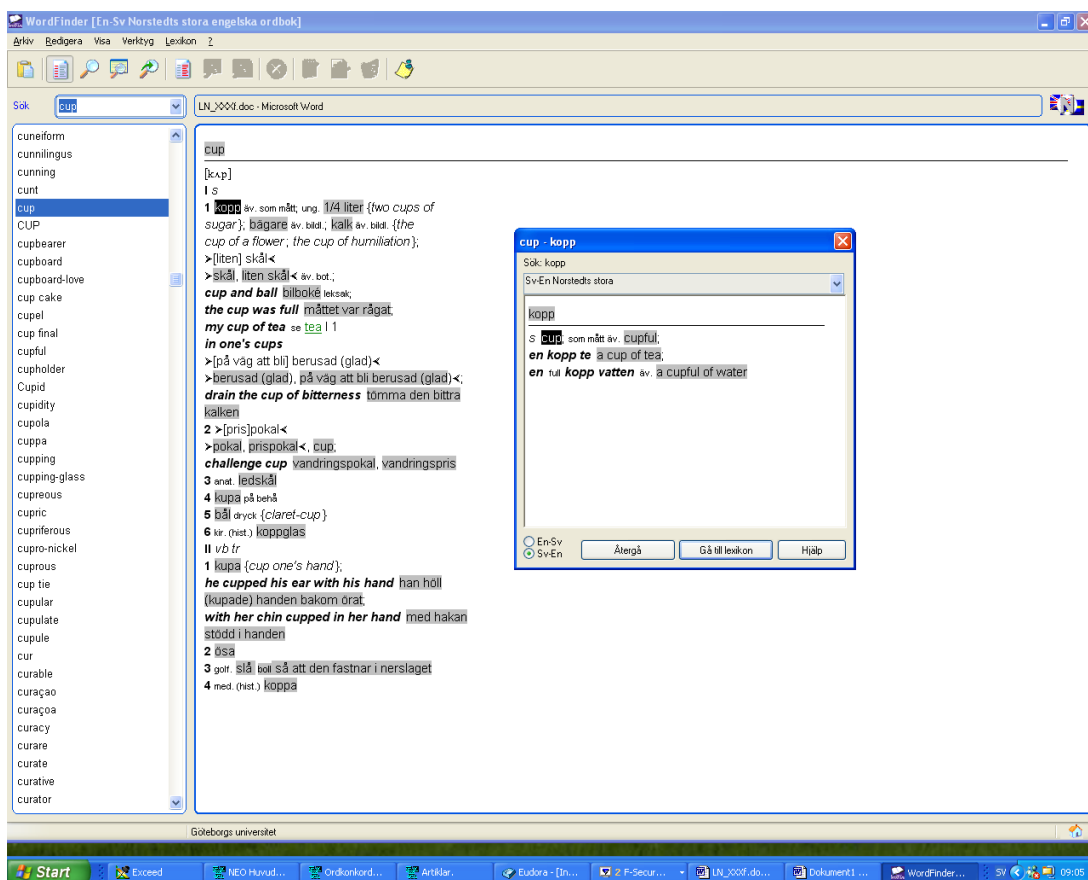


Figur 5.3. *Collins-Roberts French Dictionary*

Färgen fungerar som markör för de enskilda språkmodulerna samt för den metaspråkliga informationen i artikeltexter. Ordbokens användargränssnitt framhäver tydligt dess struktur med följande fält: ordbok, idiom, sammansättningar, kulturella notiser, förkortningar, ordspråk samt en modul med pragmatiska rekommendationer om språkbruk i konkreta situationer. Dessutom visualiseras med sökfunktionen ”Explorer” (som visas i den vänstra nedersta rutan) en uppsättning ledtrådar som ger en överblick över artikelns struktur. Med hjälp av hypertextlänkar förflyttas användaren från noder i denna struktur till respektive delar av en artikel. Denna facilitet i kombination med utnyttjande av den metaspråkliga information som redan finns i artiklar i tvåspråkiga ordböcker är ett annat exempel på en kreativ återanvändning av information för att förbättra åtkomsten. Det finns flera andra smarta lösningar i Collins-Robert som platsar i framtida e-ordböcker med anspråk på multimedial karaktär. Dit hör bl.a. inspelade uttalsangivelser för 95.000 uppslagsord samt notiser med encyklopediska uppgifter om t.ex. viktiga kulturella och sociala institutioner i respektive länder.

Collins-Roberts gränssnitt med dess tydliga strukturering framhäver informationsrikedom i kontrast till WordFinder-versioner av Norstedts ordböcker som väljer att lägga fokus på den praktiska samverkan mellan ordboken och textdokument, inte minst i fråga om kopieringsfaciliteter. Skuggade partier i artikeltexten är potentiella

kopieringskandidater. Till basutbudet av sökfunktioner i WordFinders tvåspråkiga ordböcker hör sökningar bland uppslagsord, fritextsökning, sökning på ekvivalentsidan samt möjlighet att föra egna anteckningar och ordlistor. Alla dessa sökningar kan göras i ordboksgrupper med samma ingångsalfabet. Till skillnad från *Skribent*-versionen visas inte alla sökresultaten samtidigt i *WordFinder Stora skribentordboken*, vilket delvis missar huvudpoängen med det underliggande jämförande syftet vid gruppering. Delvis kompenseras detta av tillgången till en simultan visning av artiklar ur L1-L2 och L2-L1 ordböcker med hjälp av programfönstret, vilket illustreras nedan med ett exempel ur *Norstedts stora engelska ordbok*.



Figur 5.4. *WordFinder Language Suite: Norstedts stora engelska ordboken*

Av de anförda exemplen ur olika e-ordböcker framgår tydligt att det finns potential att vidare förädla både informationsåtkomsten och användargränssnittet i svenska e-ordböcker. Förädling av informationsåtkomst innebär dels uppdatering av lexikala databaser med nya typer av information, dels bättre anpassning av innehållet och gräns-

snittet i e-ordböcker till ordboksanvändarens sökbehov. Många sökningar finner sina svar med hjälp av interna ordbokslänkar under förutsättning att behövlig information finns i den berörda databasen. För andra sökningar behöver man utöka sökandet till externa källor, vilket aktualiserar frågan om integrering av e-ordböcker eller deras databaser (Atkins 2002, Lange & Törnquist 2003: 188–190, Toporowska Gronostaj 1994). Ju mer integrerade dessa är desto mer adekvata blir sökträffarna. Etablering av sådana länkar på betydelsenivå är inte ett helt trivialt problem men det måste hanteras i framtida e-ordböcker, förhoppningsvis semiautomatiskt. Mer enhetliga betydelsestrukturer och homogent metaspråk är en förutsättning för ett lyckat resultat i samband med informationssökningar och åtkomster. Dessutom skulle en större grad av formalisering av lexikaliska data gynna deras återanvändning i olika språkteknologiska applikationer, t.ex. maskinöversättning, informationsutvinning, datorstödd undervisning och automatisk textanalys.

Litteratur

Ordböcker

- Bonniers synonymordbok*. 3 uppl. 2000. Göran Walter. Stockholm.
- CALD = *Cambridge Advanced Learner's Dictionary*. CD-ROM. 2003. Version 1.1.
- Collins-Robert = The Unabridged Collin-Roberts French Dictionary*. CD-ROM. 2003. Version 1.0.
- Engelska ordboken*. <http://www.ne.se>
- LDOCE = *Longman Dictionary of Contemporary English*. CD-ROM. 2003: London: Longman
- LEXIN = *Lexikon för invandrare, Svenska ord*.
<http://lexin.nada.kth.se/sve-sve.shtml>
- LEXIN = *Lexikon för invandrare, Svensk-engelsk*.
<http://lexin.nada.kth.se/sve-eng.shtml>
- MED = *Macmillan English Dictionary for Advanced Learners*. CD-ROM. 2002. Version 1.2.
- NEO = *Nationalencyklopedins ordbok*. CD-ROM. 1999. Höganäs
- Norstedts första Engelska ordbok*. CD-ROM. 1999. Stockholm.
- Norstedts skandinaviska ordbok*, <http://www.nordskol.org/ordbok>
- Norstedts stora engelska ordbok*. CD-ROM. 2004.
- SAOB = *Svenska Akademiens ordbok*, <http://www.saob.se>
- Svenska ordboken*, CD-ROM/on-line. 2001. version 1.0.

Svensk synonymordbok utarbetad av G. Walter.

<http://www.synonymer.se>

WordFinder Language Suite. CD-ROM. 2004. *Norstedts stora engelska ordbok*, svensk-engelsk och engelsk-svensk, samt *Grammatik 2*, ett textkontrollprogram för engelska.

WordFinder Stora skribentordboken. CD-ROM. 2004. *Svenska Akademiens ordlista över svenska språket, Ord för ord, Svensk ordbok, Förkortningsordbok, Svenska skrivregler*.

Övrig litteratur

Atkins, B.T.S. (1996) 2002: *Bilingual Dictionaries: Past, Present and Future*. I: Marie-Hélène Corréard (utg.): *Lexicography and Natural Language Processing. A festschrift in Honour of B.T.S. Atkins*. Euralex 2002, 1–29.

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

Haarala, R. & Lehtinen, M. 1997. En ny typ av produkt — en recensent med gammal typ av ammunition. I: *LexicoNordica* 4:205–221.

Lange, S. & Törnquist, L. 2003. Thesaurus Lex: En ny svensk elektroniskt länkad thesaurus. I: *Nordiske studier i leksikografi* 6, 181–190.

Malmgren, S.-G. 1999: Att göra en enspråkig definitionsordbok mer produktionsinriktad – exemplet Nationalencyklopedins ordbok. I: Slotte, P. et al. (red.): *Nordiska studier i leksikografi* 4. (Rapport från Konferensen om leksikografi i Norden. Esbo 21–24 maj 1997.) Helsingfors, 285–290.

Martola, N. 1998: Nationalencyklopedins ordbok på cd-rom. I: *LexicoNordica* 5, 187–204.

Romppanen, B. 2005: Den nya finska elektroniska ordboken *Kielitoimiston sanakirja*. I: *LexicoNordica* 12, 285–291.

Svensén, B. 2004: *Handbok i leksikografi. Ordböcker och ordboksarbete i teori och praktik*. Norstedts Akademiska Förlag: Stockholm.

Toporowska Gronostaj, M. 1994: Mot en mångdimensionell tvåspråkig ordbok. I: Garde, A. & P. Jarvad (red.): *Nordiske studier i leksikografi II*. (Rapport fra Conference om Leksikografi i Norden 11.–14. maj 1993.) Köpenhamn, 96–122.

Institutionen för svenska språket
Box 200
SE-405 30 Göteborg
maria.gronostaj@svenska.gu.se