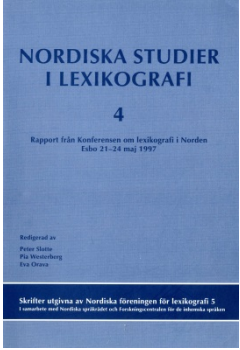


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Att göra en enspråkig definitionsordbok mer produktionsinriktad-exemplet Nationalencyklopedins ordbok	
Forfattare:	Sven-Göran Malmgren	
Kilde:	Nordiska Studier i Lexikografi 4, 1997, s. 285-290 Rapport från Konferens om lexikografi i Norden, Esbo 21.-24. maj 1997	
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.

Sven-Göran Malmgren

Att göra en enspråkig definitionsordbok mer produktionsinriktad – exemplet Nationalencyklopedins ordbok

Two kinds of 'word on the tongue'-situations are discussed. In the first case, the speaker/dictionary user is looking for a word having a special, well-defined relation to a given word (a **basic lemma**). In the second case, he knows the sense of the word he is looking for, but there is no basic lemma to start out from. In both cases, the possibilities of CD-ROM-dictionaries to lead the dictionary user to the right word are emphasized.

Enspråkiga definitionsordböcker är i allmänhet mer receptions- än produktionsinriktade. De ger i första hand svar på frågor av typen "Vad betyder ordet O?" och i mindre utsträckning svar på frågor av typen "Vilket ord har betydelsen (eller funktionen) S?" Det är den senare typen av frågor som denna lilla uppsats ska ägnas åt, och för att konkretisera tänker vi oss en språkbrukare som har ett ord "på tungan" men behöver hjälp att finna det. Vilken hjälp kan han då få i gängse enspråkiga (svenska) definitionsordböcker?

Man kan skilja mellan två olika typer av "ordet-på-tungan"-situationer. I det första fallet utgår språkbrukaren/ordboksanvändaren från ett känt ord (lemma) och trevar efter ett annat ord med en välbestämd relation till detta lemma. Exempel på sådana frågor – här uttryckta i grammatiska termer – är "vad är nomen actionis till verbet *föredra*?", "vad är nomen agentis till *agera*?", "vilket är det kausativa verb som hör ihop med adjektivet *enkel*", "vilket är det inkoativa verb som hör ihop med *hård*", "vilket verb är "stödverb" (kollokator) till *självmod*", "vilket är det förstärkande adjektivet till *ungkar*?" Också vissa typer av mer utpräglat semantiska frågor hör till denna grupp, framför allt frågor om synonymer och motsatsord. Det ord (lemma) som användaren utgår från kallas i fortsättningen **baslemmat**.¹ – Den andra typen av ordet-på-tungan-situationer kan åskådliggöras med frågor som "Vad kallas en äldre typ av enklare militärläkare?" eller "Vad kallas ett vitt, skjortliknande liturgiskt plagg?". Här har ordboksanvändaren inget välbestämt baslemma att utgå ifrån men däremot kan han i bästa fall gissa rätt på ett eller flera nyckelord i det sökta ordets definition.

Frågor av den första typen kan i princip besvaras, och har till viss del besvarats, inom ramen för tryckta ordböcker, men i praktiken blir informationsmängden så stor att CD-ROM-tekniken i varje fall ger helt nya möjligheter att tillgodose informationsbehov av detta slag. Frågor av den andra typen kan inte besvaras av konventionella definitionsordböcker, om man bortser från att användaren alltid har möjligheten att bläddra från pärm till pärm i varje enskilt fall. Också här kan CD-ROM-tekniken ge intressanta nya möjligheter, som snart ska antydas.

¹ Termen anknyter till den terminologi som utvecklats inom **kollokationsforskningen** (se t.ex. artikeln **bas** i NLO och även Hausmann 1985), men förleden *bas* används här i något allmännare betydelse genom att inte bara kollokationer utan även synonymer, motsatsord m.m. omfattas. Dvs. liksom *självmod* betraktas som bas(lemma) i förhållande till *begå* i kollokationen *begå självmord*, betraktas *hård* som bas(lemma) till *mjuk*, om en användare använder lemmat *hård* som utgångspunkt vid sökning efter ordet *mjuk*.

Skiljelinjen går, med detta sätt att se, mellan sökning efter okända ord med utgångspunkt från ett baslemma och sökning efter ord med utgångspunkt från en känd betydelse. Huvuddelen av uppsatsen ägnas åt den senare typen av ordsökning, och de flesta exemplen hämtas från CD-ROM-versionen av *Nationalencyklopedins ordbok* (jfr Malmgren 1995).

Från ett baslemma till relaterade ord

I de flesta större enspråkiga svenska definitionsordböcker ges, som en del av den semantiska och grammatiska beskrivningen, synonymer, motsatsord, språkprov m.m., och det kan synas som om informationsbehoven av den första typen vore ganska väl tillgodosedda. Många av språkproven i t.ex. *Nationalencyklopedins ordbok* är eller innehåller t.ex. **kollokationer** i Hausmanns (1985) mening. Specialordböcker (konstruktions- eller kombinationsordböcker) som t.ex. *Svensk handordbok* har givetvis en högre täckning av detta slag. Problemet är emellertid – och det gäller även de speciella konstruktionsordböckerna – dels att någon systematisk genomgång i syfte att täcka alla viktiga lexikala relationer vid alla viktiga ord inte har gjorts, dels att den information som ändå finns inte är disponerad på grundval av (typer av) lexikala relationer, dvs. på ett sådant sätt att en användare som är intresserad av en viss typ av relation direkt kan hitta motsvarande språkliga uttryck. Ibland finns informationen inte ens under baslemmat utan under det uppslagsord som användaren inte kan komma på. Under alla omständigheter måste användaren behärska svenska på nära nog modersmålsnivå. (Jfr Hausmann 1985 för delvis likartad kritik av tyska ordböcker.)

Det finns en prototyp till en ordbok (för ett annat språk) som uppfyller dessa krav, men en ordbok med god täckningsgrad av ordförrådet, baserad på denna prototyp, kommer knappast att färdigställas inom överskådlig tid. Det arbete som avses är Mel'chuk/Zholkovskys (1984) fragment av en kombinatorisk ordbok för ryska, en ordbok som innehåller – i stort sett – fullständiga beskrivningar av drygt 100 centrala ryska ord. En uppsättning (ett drygt femtiotal) lexikala relationer, däribland de som exemplifierades i inledningen av denna uppsats, har etablerats och ”checkats av” mot varje uppslagsord (baslemma). Informationen ges i en sorts pseudomatematisk notation. Överfört till några av exemplen i inledningen ser det ut så här; S_0 , S_1 , $Oper_1$, $Magn$ och de övriga ca 50 semantiska och grammatiska relationerna kallas **lexikala funktioner**, en term som – något informellt – kommer att användas i det följande:

Nomen actionis:	S_0 (föredra) = föredragande, föredragning
Nomen agentis:	S_1 (agera) = aktör
”Aktivt” stödverb:	$Oper_1$ (självmod) = begå
Förstärkande adjektiv:	$Magn$ (ungkarl) = inbiten

(Indexen anger vilken ”aktant” som aktualiseras; en etta betyder i princip att det handlar om (djup)subjektet eller i kasusgrammatiska termer agenten.)

För engelskans del finns en känd ordbok som – utan att det explicit omtalas – bygger på idén med lexikala funktioner: *The BBI combinatory dictionary of English*. Den täcker visserligen bara något tiotal av Mel'chuk/Zholkovskys lexikala funktioner och omtolkar dem ibland på ett ganska osofistikerat sätt, men å andra sidan behandlar den i princip hela det centrala engelska ordförrådet. Vidare är artiklarna i BBI-ordboken disponerade efter lexikala funktioner, så att användaren snabbt kan hitta det språkliga uttryck som står i en viss relation till det givna baslemmat (se dock

Gellerstam, Malmgren & Toporowska-Gronostaj 1992 där vissa brister i detta avseende påvisas). Också en L2-användare utan avancerade kunskaper i engelska kan därför ha glädje av ordboken.

Den i särklass största nordiska kombinationsordboken är Jón Hilmar Jónssons *Orðastaður*. I viss mån tycks den vara inspirerad av BBI-ordboken, men framför allt bygger den på ett eget, ytterst genomtänkt system för disposition av artiklarna efter semantiska kriterier och för att leda användaren (som inte behöver kunna isländska på modersmålsnivå) rätt.

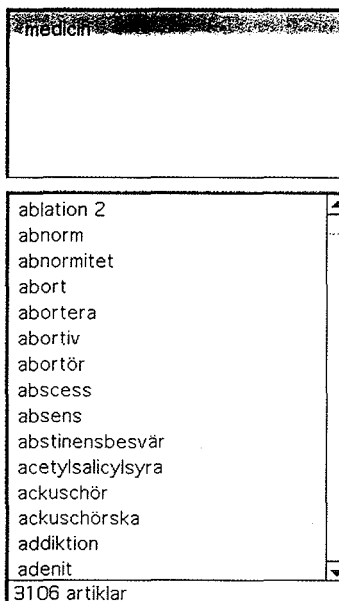
För andra upplagan av Nationalencyklopedins ordbok (NEO) genomgås hela ordbokens lemma-uppsättning systematiskt med hänsyn till ett urval viktigare lexikala funktioner. Redan i NEO:s första upplaga finns systematiska uppgifter om nomen actionis till alla verb; dessa uppgifter kompletteras nu med systematiska uppgifter om nomen agentis, om nomen patientis ("objektet" för verbhandlingen: *undervisa – elev*), om kausativa och inkoativa verb till (huvudsakligen) givna adjektiv (*enkel – förenkla* resp. *hård – hårdna*), om klassificerande adjektiv till givna substantiv (*anorexi – anorektisk*), om förstärkande morfem/ord/fraser till adjektiv, substantiv och verb (*högtaktuell, svårt sjuk, gammal som gatan, en inbiten unkarl, dricka kopiöst*), om olika typer av stödverb (*fatta beslut, begå självmord, föra krig*) etc. Det är dock inte säkert att en disposition av denna nya information på det sätt som antytts (efter lexikala funktioner) är helt förenlig med den tryckta ordbokens struktur; förmodligen passar åtminstone en del av denna information bättre på senare CD-ROM-versioner, gärna kopplade till ordbehandlingsprogram.

Sökning efter ord utan tillgång till välbestämt baslemma

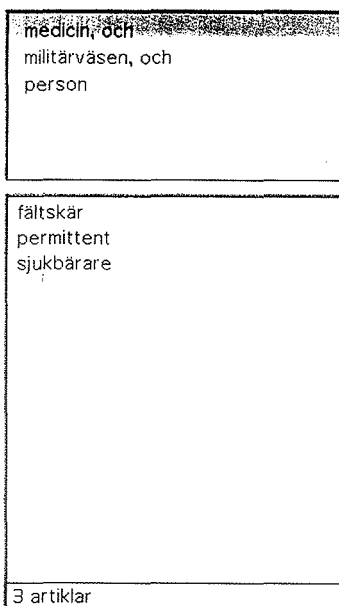
För att få svar på frågor av den andra typen (välbestämt baslemma saknas) kunde en ordboks-användare tidigare använda sig av två olika typer av hjälpmedel, en tesaurs eller en bildordbok. Utan att på något sätt ha gjort dessa hjälpmedel passé ger CD-ROM-tekniken ordboks-användaren åtminstone två nya möjligheter att hitta "ordet på tungan". Den ena – och kanske mindre seriösa – möjligheten är välkänd för många korsordslösare: wild card-sökning. Språkbrukaren kommer kanske ihåg det sökta ordets begynnelsebokstav och dess slut, och sökning efter alla ord som innehåller vissa givna bokstäver är en standardfunktion hos de flesta CD-ROM-ordböcker. Den andra möjligheten innebär en sorts förfining av tesaurstanken och bygger i grunden på att språkbrukaren kan gissa vissa nyckelord (helst fler än ett) i det sökta ordets definition, eller på att han kan ange vilket ämnesområde det sökta ordet tillhör.

I CD-ROM-versionen av Nationalencyklopedins ordbok går det att göra alla dessa typer av sökningar. I motsats till vad som är fallet i ordbokens tryckta version är alla lemman (och alla underbetydelser) här försedda med uppgifter om relevanta ämnesområdestillhörigheter. Också ord som inte är facktermer men som faller inom ett visst ämnesområde markeras; t.ex. anges att *snuva* hör till ämnesområdet "medicin". Ett knappt hundratal ämnesområden har urskilts. De minsta (t.ex. "diplomati") omfattar några tiotal ord, de största (t.ex. "medicin") flera tusen ord i NEO:s CD-romversion. Utöver ämnesområden har kategorier som "ord som betecknar person" och "ord som betecknar färger" upprättats och är sökbara på samma sätt som egentliga ämnesområden.

Vi exemplifierar först med det gynnsamma fallet att det av språkbrukaren efterfrågade ordet tillhör flera olika ämnesområden. För att återvända till det ena av exemplen i inledningen, bör det ord som betecknar en äldre typ av enklare militärläkare vara markerat både "medicin" och "militärväsen" och dessutom "person". Det sökta ordet bör således finnas i skärningsmängden av motsvarande tre ordmängder. Den inledande sökningen på "medicin" ger träff på 3106 artiklar:



Fortsatt sökning med "och"-funktionen på artiklar markerade "militärväsen" reducerar antalet träffar till 15, och det ytterligare villkoret "person" ger till sist följande tre återstående träffar:



Användaren bör nu kunna se att det efterfrågade ordet är *fältskär*, ev. efter konsultation av de tre kvarstående artiklarna i ordboken.

Ordet i det andra exemplet i inledningen (vad kallas ett vitt, skjortliknande liturgiskt plagg?) erhålls tillsammans med 11 andra ord som resultat av sökning på ord som samtidigt tillhör ämnesområdet "religion" och ämnesområdet "kläder". (Ordet är *röcklin*.)

Ofta räcker det dock inte med sökning på ämnesområde(n). Den möjlighet som då står till buds är sökning på (förmodade) nyckelord i det sökta ordets definition, eventuellt kombinerad med sökning på ämnesområde. En användare söker kanske efter växtnamnet *brudsporre*, och det han vet är att denna växt bär violetta blommor. Sökning på ordet "violett" och ämnesområdet "botanik" (skärningsmängden) ger 17 träffar, däribland det sökta ordet. Vid så begränsade träffmängder torde användaren i de flesta fall lätt identifiera det ord han trevade efter.

Vi avslutar med ytterligare några exempel: 1) Användaren kan inte komma på ordet *kulturrevolution*. Om han slår på ämnesområdet "politik" och på ordet "Kina" får han det sökta ordet tillsammans med två andra ord. 2) Användaren trevar efter ordet *marinad*. Sökning på ämnesområdet "kokkonst" och på ordet "smaksättning" ger 12 träffar, däribland det efterfrågade ordet. 3) Användaren har ordet *jalusi* "på tungan". Sökning på ämnesområdet "heminredning" kombinerat med sökning på ordet "skydd" ger detta ord tillsammans med 11 andra.

Självfallet måste dessa möjligheter testas av användarna, innan man kan avgöra hur stor hjälp de innebär i ordet-på-tungan-situationer. När man sitter och laborerar med NEO-CD:n är det också lätt att se olika sätt att förbättra sökmöjligheterna. En klar förbättring vore möjlighet till lemmatiserad sökning, dvs. sökning i ett sammanhang efter samtliga böjningsformer av ett ord (t.ex. både *hand* och *händer*) – detta är t.ex. möjligt i NEO-CD:ns finska motsvarighet *CD-Perussanakirja*. Vidare skulle säkert möjligheterna att hitta det rätta ordet öka, om det gick att söka i ett sammanhang på ett ord och alla dess synonymer och kanske också hypo- och hyperonymer. En användare som trevar efter ordet *glänta* kan t.ex. inte veta om huvudordet i dess definition är "plats" eller "ställe". Man kan tänka sig ett system där det räcker att söka på "ställe" för att få samtliga förekomster i definitionerna av *ställe* och alla dess böjningsformer och dessutom av *plats* och alla dess böjningsformer.

Litteratur

- Benson, Morton, Benson, Evelyn & Ilson, Robert 1985: *The BBI Combinatory Dictionary of English*. New York.
- CD-Perussanakirja. Uusin tieto nykysuomen sanoista*. 1997. Helsingfors.
- Gellerstam, Martin, Malmgren, Sven-Göran & Toporowska Gronostaj, Maria 1992: Inlärningslexikon. I: Ruth Vatvedt Fjeld (utg.), *Nordiske studier i leksikografi. Rapport fra Konferanse om leksikografi i Norden 28.-31. mai 1991*, 347-357. Oslo.
- Hausmann, Franz Josef 1985: Kollokationen im deutschen Wörterbuch: Ein Beitrag zur Theorie des lexikographischen Beispiels. I: Henning Bergenholtz & Joachim Mugdan (red.), *Lexikographie und Grammatik: Akten des Essener Kolloquiums zur Grammatik im Wörterbuch 28.-30.6 1984*, 118-129. Tübingen.
- Jón Hilmar Jónsson 1994: *Orðastaður*. Reykjavík.
- Malmgren, Sven-Göran 1996: Nationalencyklopedins ordbok – några karakteristiska typer av information. I: Ásta Svavarsdóttir, Guðrún Kvaran & Jón Hilmar Jónsson (utg.), *Nordiske studier i leksikografi 3. Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995*, 275-282. Reykjavík.
- Mel'chuk, Igor & Zholkovsky, Alexander K. 1984: *Explanatory combinatorial Dictionary of modern Russian. Semantico-Syntactic Studies of Russian Vocabulary*. [Äv. rysk titel.] Wien.

Nationalencyklopedins ordbok 1–3. 1995–96. Höganäs.

Nationalencyklopedins ordbok, CD-ROM-version. 1997. Höganäs.

NLO=Henning Bergenholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson & Bo Svensén,
Nordisk leksikografisk ordbok. 1997. Oslo.

Svensk Handordbok. Konstruktioner och fraseologi. 1966. Stockholm.