

Hyperlexikon – finns de?

Lars Törnqvist

Vid lexikografikonferensen i Tórshavn 2001 presenterade Sven Lange och jag en vision av framtidens ordbok (Lange & Törnqvist 2003). Vi föreställde oss ett hyperlexikon bestående av inbördes sammanlänkade artiklar med definitioner, grammatisk information, uttalsuppgifter, översättningsekvivalenter och faktainformation, alltihop uppbyggt kring en semantisk stomme i form av en thesaurus. Idén var att göra ordsökningen effektivare genom att kombinera onomasiologiskt och semasiologiskt söksätt och att samla alla former av lexikal information i ett och samma söksystem.

Nu har vi kommit några år in i vad som då var framtiden. Har vår vision blivit verklighet?

Vad är ett hyperlexikon?

Med *hyperlexikon* menar vi ett elektroniskt lexikon där olika typer av strukturerad information är ömsesidigt förbundna med hyperlänkar. Informationen kan avse både språk och fakta, vilket innebär att både ordböcker och encyklopedier kan ingå.

Det finns många sätt att utnyttja hyperlänkning i ordböcker. Svensén (2004: 523–526) delar in länkarna i ordboksinterna länkar och ordboksexterna länkar. Som exempel på ordboksintern länkning nämner han bland annat länkar från en ordboksartikel till en annan, från diagram till ordboksartikel, från uppslagsord till auditiv uttalsangivelse och länkar till översikter över semantiskt och ontologiskt besläktade ord. Exempel på ordboksextern länkning är bland annat länkar från ordbok till encyklopedi, fackordbok, thesaurus eller textkorpus.

Hyperlexikon kan innehålla både interna och externa länkar. Vad som tillhör internt respektive externt material spelar egentligen ingen större roll för användaren så länge som komponenterna är integrerade i ett system. Vilka komponenter som ingår i ett hyperlexikon kan naturligtvis

variera beroende på syftet med lexikonet och tillgången på lexikala data för projektet. Hyperlexikon kan därför vara uppbyggda på mycket olika sätt och ha mycket varierande komplexitet. Vår vision visar alltså inte det enda tänkbara hyperlexikonet, utan ett hyperlexikon av många möjliga.

Lexikon som inte är hyperlexikon

Dagens elektroniska ordböcker är i allmänhet mer tekniskt avancerade än traditionella tryckta ordböcker. Framförallt är sökmöjligheterna mer utvecklade, och artiklarna innehåller ofta fler typer av information. Sådant är dock inte tillräckligt för att göra dem till hyperlexikon.

I elektroniska lexikon är sökningen inte begränsad till alfabetiska listor över uppslagsord, utan det finns möjlighet till sökning på delar av ord – början, slutet eller en teckensträng mitt i ordet. Man kan också söka på annat än uppslagsord, till exempel ord eller kombinationer av ord i artikeltexterna, vilket utökar sökmöjligheterna avsevärt. Det är dock fortfarande bara fråga om sökning av teckensträngar utan semantiska kopplingar.

Sedan länge finns det stora databaser som innehåller flera ordböcker av samma slag. Den vanligaste tillämpningen är termbanker, men på senare år har idén kommit att användas även för andra ordbokstyper, till exempel ordbokshotell för dialektordlistor (Tvedt, Lien & Eide 2006). Inte heller dessa databaser är hyperlexikon.

Gemensamt sökgränssnitt för flera lexikala databaser är ytterligare en utvecklingslinje. Ofta rör det sig om ordböcker och uppslagsverk från samma utgivare. Flera av de stora engelskspråkiga ordboksförlagen har sådana tjänster på Internet, ofta med enklare funktioner fritt tillgängliga och mer avancerade funktioner tillgängliga för betalande abonnenter. *Merriam-Webster Online* innehåller fyra lexikon: definitionsordbok, thesaurus, spansk-engelsk ordbok och medicinskt lexikon. Vid sökning anger man vilket av lexikonen man vill söka i. Det finns vissa funktioner för vidarelänkning i systemet, men lexikonen är inte sammanlänkade med varandra. *Cambridge Dictionaries Online* är en liknande söktjänst som innehåller sex lexikon, både enspråkiga och tvåspråkiga. Ett svenskt exempel är *Nationalencyklopedins Internettjänst*, där man söker samtidigt i *Nationalencyklopedin*, *Nationalencyklopedins ordbok* och *Nationalencyklopedins juniorlexikon*.

Det finns också fristående söktjänster som utför parallell sökning i flera fritt tillgängliga ordboksdata-baser. De flesta finns av naturliga skäl

inom det engelska språkområdet, där det finns ett mycket stort utbud av lexikala databaser. Den mest omfattande av dessa tjänster är *OneLook Dictionary Search*, som söker i mer än 900 ordböcker av olika slag, både allmänspråkliga och fackspråkliga. Ett svenskt exempel med något blygsammare omfattning är *NätLex* från Trilobite Software, som söker i flera engelsk-svenska och svensk-engelska ordböcker.

Ingen av de här nämnda söktjänsterna kan räknas som hyperlexikon eftersom de olika ordböckerna i systemen inte är inbördes länkade.

Ordboksintern länkning

Ett första steg till ett hyperlexikon är klickbara hyperlänkar i en vanlig ordboksartikel, en funktion som egentligen är en elektronisk variant av de tryckta ordböckernas och uppslagsverkens hänvisningar mellan olika artiklar. Terminologiska ordlistor brukar vara särskilt rikligt försedda med interna hänvisningar direkt från ord i definitionerna. På så vis undviker man redundans i de systematiskt uppbyggda ordlistorna, men en nackdel är att läsaren måste bläddra mycket för att få hela sammanhanget klart för sig.

Många tidiga termbanksprojekt avstod från att utnyttja denna länkning av tekniska skäl. Nyare ordböcker, som har utformats direkt för webbpublicering, använder dock länkning i större eller mindre omfattning. Ett projekt där länkningen utnyttjas mycket effektivt är Wikipedia, som presenteras närmare nedan.

Ordboksextern länkning

Länkar från ordboksartiklar till externa källor har blivit vanligare på senare år. Framförallt utnyttjas detta i encyklopediska uppslagsverk, där många artiklar har en rad länkar till andra webbplatser med mer omfattande fakta. Extern länkning förekommer även i en del rena ordböcker, där det kan finnas länkar till illustrationer eller artiklar i ett uppslagsverk, vanligen Wikipedia. Inte heller detta gör ordböckerna till hyperlexikon, men vi närmar oss.

Multifunktionell ordbok

Den franska TV-kanalen TV5 Monde har ett lexikon på sin webbplats: *le dictionnaire multifonctions*. Som namnet antyder är det en ordbok med flera funktioner.

I ordbokens sökfönster finns en sökruta och sex flikar för olika typer av information: definitioner, synonymer, böjning, fraser, fransk-engelsk ordbok och engelsk-fransk ordbok. Man skriver in ett ord och markerar önskad flik, och strax kommer en artikel upp i fönstret. Orden i artiklarna är klickbara, vilket innebär att man kan hoppa direkt från en definition till en annan, eller direkt från en synonym till en annan. Man kan också markera en annan flik för att gå från definitionen för ett ord till synonymer, böjning eller översättning av samma ord. Det går att hoppa i alla riktningar, vilket ger stora möjligheter för ordsökning.

Nu är vi framme vid målet – detta är ett hyperlexikon. Definitionerna är dock mycket enkla, och det saknas andra semantiska relationer än ren synonymi.

Wikipedia och Wiktionary

Ett annat hyperlexikon med många funktioner bildas av Wikipedia och dess olika systerprojekt. Wikipedia är ett encyklopediskt uppslagsverk som skrivs av sina användare och som finns i många språkversioner med varierande omfattning. Vid sidan av encyklopedin finns ett antal självständiga parallellprojekt med koppling till Wikipedia, av vilka ordboksprojektet Wiktionary är mest intressant i det här sammanhanget. En grundläggande princip för alla dessa projekt är att vem som helst får skriva och bearbeta artiklar i Wikipedia, oavsett formell status. Detta medför att kvaliteten på innehållet kan variera mellan högsta vetenskapliga klass och rent dravel. De större och äldre språkversionerna, exempelvis den engelska, håller i allmänhet en ganska hög klass, inom vissa fackområden fullt jämförbar med professionellt utarbetade uppslagsverk.

Det som gör Wikipedia till ett hyperlexikon är de högt utvecklade länkningsfunktionerna och användningen av kategorier. Genom en enkel markering i texten kan den som skriver en artikel skapa en länk direkt till en annan artikel i uppslagsverket. Man kan också länka till motsvarande artikel i en annan språkversion. Genom länkning till kategorier kan man bygga upp förteckningar över uppslagsord inom ett visst område. Kategorierna kan ordnas systematiskt genom att en kategori kan innehålla flera underkategorier, och på så sätt har ett thesaurusliknande system bildats. Denna klassificering präglas dock av ad hoc-lösningar,

inte av någon genomtänkt helhetssyn. Utöver kategorierna finns också artiklar som består av listor. Sådana listor innehåller ofta begrepp som inte har egna artiklar i uppslagsverket.

Systerprojektet Wiktionary är en elektronisk ordbok som bygger på samma programvara som Wikipedia. Den fungerar både som definitionsordbok och översättningsordbok. Definitionsartiklarna innehåller även morfologisk information och i många fall uppgifter om uttal och synonymer. Till skillnad från traditionella ordböcker anges alla böjningsformer i fullständig form, vilket är naturligt när man är fri från den tryckta ordbokens utrymmesbrist. Från definitionsartiklarna finns länkar till översättningsartiklar där motsvarigheter på ett eller flera andra språk anges. Från översättningsartiklarna i den svenska språkversionen finns länkar till motsvarande översättningsartiklar i de andra språken, och från dessa finns i sin tur länkar till definitionsartiklar på samma språk. Det finns möjlighet att kategorimärka uppslagsorden på samma sätt som i Wikipedia, men detta har ännu inte gjorts i någon större omfattning i den svenska versionen.

Det finns möjlighet att länka mellan artiklar i Wikipedia och Wiktionary. På så sätt kan man gå direkt från en ordboksartikel till en faktaartikel eller omvänt.

Länkningen mellan olika språkversioner gör att uppslagsverket Wikipedia även kan användas som översättningsordbok, åtminstone så länge det handlar om facktermer och egennamn. Artiklarna på olika språk kan också i viss mån användas som parallellkorpus, med reservation för bristfälliga översättningar och bristande överensstämmelse mellan texterna i de olika språkversionerna.

Korpusbaserade hyperlexikon

Ett mera forskningsinriktat sätt att konstruera hyperlexikon är att utgå från stora textkorpusar. Exempel på sådana projekt är *Das Digitale Wörterbuch der deutschen Sprache des 20. Jh.* (DWDS), som har utvecklats av Berlin-Brandenburgische Akademie der Wissenschaften, och *Wortschatz Universität Leipzig*.

DWDS bygger på en ordbok och en textkorpus. Ordboken är *Das Wörterbuch der deutschen Gegenwartssprache* (WDG), en östtysk ordbok som utarbetades mellan 1952 och 1977. Utifrån ordbokens definitioner och exempel har man med hjälp av datorprogram genererat semantiska relationer till synonymer, hyperonymer och hyponymer. De

mest frekventa kollokationerna för varje uppslagsord har beräknats på motsvarande sätt utifrån korpusen.

I webbgränssnittet presenteras informationen i fyra fönster: ordboksartikel, semantiska relationer, textexempel ur korpusen och kollokationsdiagram (se illustration). Ordboksartiklarna innehåller länkar till sammansättningar, men själva artikeltexten är inte klickbar. Däremot är alla ord i relationsfönstret klickbara, så att man direkt kan hoppa till ordboksartikeln för en synonym eller något annat relaterat ord. Även kollokationsdiagrammen är klickbara på motsvarande sätt.

Textexemplen i korpusen presenteras som konkordans. Även böjda former av uppslagsordet visas i konkordansen. En sökning på *essen* ger således exempel med verbformer som *aß*, *ißt* och *gegessen*.

The screenshot shows the DWDS website in a Windows Internet Explorer browser. The search results for 'essen' are displayed in four main sections:

- Wörterbuchartikel aus dem WDG:** Provides the dictionary entry for 'essen', including its definition, grammatical information, and various compound words like 'Abonnementsessen', 'Betriebsessen', etc.
- Automatisch berechnete semantische Relationen aus dem WDG:** Lists synonyms (e.g., Festmahl, Bankett) and hierarchical concepts (e.g., Mahlzeit, Festessen).
- Textbeispiele aus dem DWDS Kerncorpus:** Shows a concordance of text examples where 'essen' and its forms (isst, esse, essen, gegessen, aß) are used in context.
- Automatisch berechnete Kollokationen aus dem DWDS Kerncorpus:** A network diagram with 'essen' at the center, connected to related terms like 'Mittag', 'Brot', 'trinken', 'satt', and 'mehr'.

Det andra korpusbaserade hyperlexikonet, *Wortschatz Universität Leipzig*, är annorlunda uppbyggt. Som i de andra lexikonerna skriver man

in sitt sökord i en ruta, men här får man svaret som en lång lista med underrubriker för bland annat kategoritillhörighet, grammatiska egenskaper, sammansättningar och fraser, avledningar, semantiskt relaterade ord, klass i Dornseiff-thesaurusen och statistiskt framräknade kollokationer. I listorna finns också diagram över ord som förekommer med signifikant hög frekvens tillsammans med sökordet. Alla relaterade ord som anges i listorna är klickbara.

Redovisningen av semantiska relationer är ovanligt detaljerad. Här finns exempelvis hänvisningar till synonymer med positiv och negativ laddning. Eftersom en klassbaserad thesaurus ingår i systemet finns också länkar till ord med mer avlägsen relation än synonymi och hyponymi, till exempel från *Kuh* till *Esel*, *Gans* och *Schaf*.

All information om kollokationer är automatiskt genererad utifrån en korpus. Vilka ord som finns i kollokationsavsnitten beror därför i hög grad på vilka texter som har råkat ingå i korpusen, särskilt när det gäller egennamn.

I motsats till de andra hyperlexikonen saknar *Wortschatz* definitioner, utöver korta förklaringar av polysemer. Liksom DWDS är det också helt inriktat på studium av tyska språket, och innehåller därför ingen information om ekvivalenter på andra språk.

Konklusion

Har vår vision från 2001 förverkligats? Ja, delvis. Det finns i dag flera hyperlexikon fritt åtkomliga på Internet. De flesta av de ordboksmoduler vi tänkte oss – definitionsordbok, thesaurus, översättningsordbok, encyklopedi och några till – finns i åtminstone något existerande hyperlexikon, men inget av lexikonen innehåller alla modulerna. TV5 Monde innehåller definitioner och översättningar, men saknar semantiska relationer utöver synonymer. Wikipedia/Wiktionary innehåller definitioner, encyklopediska uppgifter och översättningar, alltihop effektivt sammanlänkat, men saknar heltäckande thesaurus och håller ojämn kvalitet. DWDS har en bra definitionsmodul och användbara semantiska relationer, men saknar översättningar. *Wortschatz* har bra thesaurusfunktioner och bra redovisning av kollokationer, men innehåller varken definitioner eller översättningar.

Men med tanke på vad som redan har åstadkommit dröjer det förmodligen inte så många år innan vår vision har blivit verklighet någonstans i världen.

Litteratur

- Cambridge Dictionaries Online*, <<http://dictionary.cambridge.org/>>.
- DWDS = *Das Digitale Wörterbuch der deutschen Sprache des 20. Jh.*
Berlin-Brandenburgische Akademie der Wissenschaften,
<<http://www.dwds.de/>>.
- Lange, Sven & Lars Törnqvist 2003. Thesaurus Lex: En ny svensk elektroniskt länkad thesaurus. I: Hansen, Zakaris Svabo & Arnfinnur Johansen (red.) *Nordiske studier i leksikografi* 6. Rapport fra Konference om leksikografi i Norden Tórshavn 21.–25 august 2001. Tórshavn: Nordisk forening for leksikografi. S. 181–190.
- Merriam-Webster Online*, <<http://www.m-w.com/>>.
- Nationalencyklopedins Internettjänst*, <<http://www.ne.se/>>.
- OneLook Dictionary Search*, <<http://www.onelook.com/>>.
- NätLex*. Trilobite Software,
<<http://www.trilobitesoftware.com/products/natlex/>>.
- Svensén, Bo 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- TV5 Monde: le dictionnaire multifonctions*,
<<http://dictionnaire.tv5.org/dictionnaires.asp>>.
- Tvedt, Lars Jørgen, Elisabeth Lien & Øyvind Eide 2006. *Ordbokshotellet – varig lagring og formidling av norske ordsamlinger*,
<<http://www.edd.uio.no/artiklar/leksikografi/Ordbokshotellet.pdf>>.
- Wikipedia*, <<http://sv.wikipedia.org/wiki/Portal:Huvudsida>>
- Wiktionary*, <<http://sv.wiktionary.org/wiki/Wiktionary:Huvudsida>>
- Wortschatz Universität Leipzig*, <<http://wortschatz.uni-leipzig.de/>>.
- (Samtliga webbadresser gäller 2007-09-10.)