


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Mot en hållbar lexikografi – En balansgång mellan återanvändning och förnyelse	
Forfatter:	Godelieve Laureys	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 13-29 Rapport från Konferens om lexikografi i Norden, Lund 24.-27. maj 2011	
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.

Mot en hållbar lexikografi

– En balansgång mellan återanvändning och förnyelse

Godelieve Laureys

1. Inledning

Begreppet *hållbar* är på frammarsch. Det används inom ett brett spektrum av aktiviteter. Det bär även en viss prägel av att vara ett modeord och uttrycker i vissa fall närmast en allmän positiv betydelse. Idag skall man göra saker och ting *på ett hållbart sätt*.¹ Inte bara skall man *odla, fiska, producera, förvalta, använda, utveckla, planera, hantera, nyttja, sköta, ta hand om, bevara, bygga, tillverka* ”*på ett hållbart sätt*”, utan man skall också *resa, jobba, flyga, investera pengar, laga mat, steka, träna, äta, gå ned i vikt, dricka, sola sig, duscha, klä sig, förbereda sig för ett barn*, eller kort sagt *leva* ”*på ett hållbart sätt*”². Den mest frekventa och mest lexikaliserade kollokationen är *hållbar utveckling*, ett begrepp som har kommit in i språket som en lånöversättning av engelskans *sustainable development*. Det kan vara intressant att notera att begreppet, som myntades av miljöforskaren och essäisten Lester Brown på det tidiga 1980-talet, vann världsvid anklång i samband med Brundtland-rapporten, *Vår felles framtid* (1987). *Hållbar utveckling* återges på danska med *bæredygtig udvikling*, på norska med *bærekraftig utvikling* och på finska med *kestävä kehitys* eller med det abstrakta substantivet *kestävyyys*. Substantivet *hållbarhet* används även på svenska ofta för att beteckna *hållbar utveckling*.

Begreppet avser i första hand en jämvikt mellan ekologiska, ekonomiska och sociala krav och vädjar om en långsiktig och medveten förvaltning och exploatering av världens naturliga resurser med tanke på kommande generationers behov och välfärd och pläderar för en vertikal solidaritet över tiden. Från den ekologiska domänen har begreppet förts vidare till de sociala och ekonomiska domänerna, där det även handlar om en långsiktig och genomtänkt hushållning

¹ Mer än en halv miljon hits på Google, maj 2011.

² Alla nämnda kollokationer är högre frekventa på Google.

med mänskliga och materiella resurser. På en ännu högre abstraktionsnivå innebär hållbarhetsidealet en strävan att åstadkomma bättre och varaktiga resultat med minskad resursåtgång och att skapa resurser som förblir tillgängliga för det framtida samhället. Då tangerar vi begrepp som *långsiktig effektivitet* och *optimering*.

Det är självfallet i den sistnämnda abstrakta bemärkelsen av *ökad effektivitet, för att åstadkomma bättre resultat med minskad resursåtgång* som begreppet *hållbarhet* används inom forskning i allmänhet och inom lexikografi i synnerhet. Nyckelbegrepp är: långsiktigt perspektiv, varaktiga investeringar i motsättning till engångsinsatser, framtidsinriktade och användarvänliga produkter, återanvändning av materiella och mänskliga resurser, överförbara resultat, robusta verktyg. Någon direkt miljöeffekt är det i vårt sammanhang nog knappast tal om, annat än att datoriseringen inom lexikografien torde bidra till ett papperslöst samhälle.

I följande danska citat används begreppet i en betydelse som i alla fall inte avses här: ”Og ordbøgerne er faktisk handy og praktiske at håndtere. Indbindingen er solid med et omslag i blødt plastic, hvilket øger *holdbarheden* i udsatte situationer. Alle tre ordbøger er udformet så de kan tåle at blive brugt i marken.” (Hovmark 2005: 15)

I mitt bidrag vill jag försöka belysa på vilket sätt hållbarhetsprincipen kan tillämpas i lexikografien och hur begreppet har utvecklats genom tiden. Jag urskiljer därvid tre steg i utvecklingen.

2. En ordbok åt gången

Ordböcker är redskap som hjälper oss att få inblick i både uttryckssidan och innehållssidan av ett språks lexikaliska element. Det finns som bekant många typer av ordböcker, som antingen fokuserar på vissa delar av ett språks lexikon eller som är avsedda för vissa funktioner och/eller som riktar sig till olika användargrupper. Trots alla möjliga specifikationer vad gäller typologi, funktion och användargrupp är ordböcker alltid i viss mån generiska verktyg, som skall kunna användas i flera olika sammanhang. Bortsett från vissa ordböcker som just vill fånga det nya och kanske det flyktiga i språket, som t.ex. nyordsordböcker eller slangordböcker, och bortsett från slutna vokabulärer som t.ex. terminologiska listor och glossarier, tenderar ordböcker överlag att beskriva språket i ett registermässigt, stilistiskt och tidsmässigt brett perspektiv. På så sätt kan ordböcker sägas ha hållbarhet som en inneboende egenskap.

Förr i tiden kom ordböcker oftast till som alster av en enskild författare. Ofta var det tal om ett mångårigt arbete eller till och med ett livsverk. De tidi-

gaste ordböckerna i Norden var latinska glossarier och översättningsordböcker från latinet. På 1600-talet, då intresset för folkspråken vann terräng, uppstod de första monolinguala ordböckerna för våra nuvarande nationalspråk. Även dessa ordböcker stod i tacksamhetsskuld till den latinska traditionen, således att många uttryck och exempel har förts vidare från ordbok till ordbok och levt sitt bästa liv i den lexikografiska traditionen. Detta belyses mycket tydligt i Anna Helga Hannesdóttirs avhandling om den enspråkiga svenska lexikografins utveckling (Hannesdóttir 1998).

Existerande ordböcker har länge utgjort den första tillgången för en lexikograf och därför innehåller ordböcker ofta ackumulerad kunskap. Det framgår till och med att vissa svagheter eller direkta fel på så sätt förs vidare från den ena ordboken till den andra. Detta är i och för sig inte så märkvärdigt eftersom all forskning bygger vidare på och förhåller sig till andras resultat. Men eftersom ordböcker kartlägger ett språks eller ett delspråks lexikaliska element på systematisk nivå, finns det bara hänvisningar till tidigare ordböcker i de fall där det är tal om en omarbetning eller en reviderad version av en redan existerande ordbok och inte när det gäller långods i enskilda belägg. Strävan efter att kartlägga det kanoniska språkbruket, vilket metaforiskt uttryckt faller under "ällemansrätten", gör ordbokstexter till ett slags inventarier, som inte avspeglar en individuell författares eller talares idiosynkratiska språkbruk, utan som representerar det kollektiva språkbruket. Ordböcker innehåller m.a.o. en förteckning över ett språksamhälles gemensamma arvegods, som självfallet är *open source*. Att lexikografer använder existerande ordböcker som källa kan även leda till vissa frågor angående den intellektuella egendomsrätten. Det som är upphovsrättsligt skyddat är själva urvalet både på mikro- och på makronivå. Den unika selektionen av språkligt material, baserad på vissa redaktionella principer, är *intellectual property*³. Det är det som utgör resultatet av en noggrann samlings- och systematiseringsprocess, de enstaka uppslagsorden är däremot *share ware*. När det gäller metaspråkliga uppgifter, såsom betydelsesuméer, definitioner och grammatisk annotation möter vi samma problem. Ingen redaktör kan sägas vara ägaren av pluralisformen till substantivet *flicka*, men när det t.ex. gäller orddefinitioner finns det tydliga upphovsrättigheter.

Frågan gäller även exempelaterialet och autentiska korpusexempel. Vissa korpusar är *open source*, andra är licensierade och får inte användas utan tillstånd. De korpusar som har kommit till stånd med offentliga medel är överlag tillgängliga för forskarsamhället och ofta även för en bred publik. Frågan är om

³ Vissa ordboksförfattare har därför tagit upp s.k. fantomord eller virusord, dvs fiktiva ord som kan spåras för att påvisa plagiat.

dess material också får användas av privata företag i kommersiellt syfte. Många ordboksredaktörer kommer på så sätt att arbeta inom en gråzon.

3. Från ordbok till ordboksserie

På 1900-talet kom det fart i ordboksproduktionen. Ett antal förlag började specialisera sig på ordböcker. Det ledde till olika ordboksserier, som skulle förstärka förlagets varumärke. Seriebildningen var ett kraftigt medel i marknadsföringsstrategin. Serierna blev också alltmer specialiserade. Särskilt gäller detta för tvåspråksordböcker, som delades upp i skolordböcker, rese- eller mini-ordböcker, medelstora ordböcker och för vissa språk även stora professionella ordböcker.

Dessa ordboksserier har inte alltid representerat en enhetlig redaktionspolitik. Visserligen har ett antal ordböcker kommit till i förlagets regi, dvs. de blev utarbetade på själva förlaget av interna redaktörer, men ett ännu större antal ordböcker har utarbetats på beställning av externa redaktörer. I vissa fall köptes helt enkelt färdiga manuskript eller kom det först till en överenskommelse, när manuskriptet var färdigt i stora delar. Det som faktiskt mest höll ihop en serie var pärmens utformning, vilket är helt i överensstämmelse med dåvarande marknadsstrategi.

På 1990-talet, då den första stora fusionsvågen inom förlagsvärlden ägde rum, började de stora förlagen att lansera tanken på en egen förlagspolitik vad gäller ordböcker. Kvaliteten i de många nya uppköpta produkterna skulle granskas och en viss profilering på innehållssidan trängde sig på. Detta sammanföll med att datoriseringen gjorde sitt inträde i branschen. Tanken på att samköra olika databaser och göra dem inbördes kompatibla växte fram. Drömmen om en masterdatabas eller ett s.k. moder-bestånd, som ett antal olika ordböcker kan härledas ur, skymtade vid horisonten. Men ännu var det en lång väg att gå. Det fanns inga standardiserade softwarenormer. Datorer var mestadels stand alone-maskiner eller var kopplade till lokala nätverk, det manuella arbetet i post-redigeringsfasen sparades sällan elektroniskt och mycket gick förlorat på vägen. Det är den period då förfarandena kan betecknas som datorstödda, inte som datorbaserade eller datorstyrda. Datorerna var trogna, noggranna och flitiga medarbetare, som i bästa fall utförde det som redaktörerna hade bestämt, men som varken påverkade redaktionsprinciperna eller ordbokens utformning.

Som exempel kan nämnas Gyldendals *røde ordbøger* och Norstedts *randiga serie*. Det danska förlaget Gyldendal har utgett ordböcker alltsedan 1798. Den första *røde ordbog* utkom 1902 som ett svar på Nordisk Förlags blå ordboksserie.

Ordbøgerne var indbundet i rødt lærred i modsætning til den uindbundne udgave fra Nordisk Forlag – og så var de endda en krone billigere. De kostede den nette sum af

1,50 kr. Den første røde ordbog var Dansk-Norsk-Engelsk Ordbog. I 1902 udkom desuden Dansk-Tysk ordbog – derefter blev ordbogsserien udbygget med alle de vigtige sprog i løbet af de følgende år. I dag tæller serien mere end 30 titler fordelt på 24 sprog, fra svensk til persisk, polsk og tyrkisk. De røde ordbøger blev en øjeblikkelig succes og blev hurtigt en af Gyldendals vigtigste udgivelseslinjer. Det samlede salg ligger på ca. 14 millioner eksemplarer, og serien er i dag et af de bedst kendte danske varemærker.⁴



© Gyldendal

I Sverige har Norstedts profilerat sig som ledande ordboks-förlag. Varumärket används också som en kvalitetsgaranti, som det framgår av följande citat:

”Ordboken är utvecklad av Norstedts, som har en lång tradition av ordboksutgivning. Norstedts randiga ordböcker är kända av de flesta, och ordböckerna borgar för kvaliteten.”⁵



4. Återanvändning av ordboksmaterial och generiska redaktionsprinciper

Sedan millennieskiftet har den traditionella grafiska ordboken så småningom varit på väg att bli en nischprodukt. De första elektroniska ordböckerna, som distribuerades på CD-rom på 1990-talet, avvek dock inte stort från sina grafiska modeller, förutom att layouten var mer översiktlig för att man inte längre skulle ta hänsyn till utrymme. Själva mediet hade blivit annorlunda, det gick snabbare att slå upp ett ord eller ett uttryck, man kunde bläddra med en browser, men sökningsprocedurerna var och är desamma som i grafiska ordböcker. Man kan exempelvis inte selektera vissa informationskategorier, utan man får presenterat

⁴ Citat från Gyldendals Forlags hemsida <http://ordbog.gyldendal.dk/> maj 2011.

⁵ Citat från pärmen till den Svensk-Nederländska ordboken. Se även (Laureys & De Groot 1996).

allt på skärmen precis som det står i den grafiska ordboken. Men uppkomsten av elektroniska ordböcker utgör ändå ett första exempel på återanvändning: ur samma underliggande intellektuella material härleds två olika produkter.

Vi fick emellertid vänta på de nätbaserade lexikaliska databaserna och ordböckerna för att verkligen kunna ta det tredje steget fullt ut. Här rymms helt andra möjligheter för att lagra, presentera, distribuera och uppdatera lexikalisk information. Webbordböcker är inte lineära och rymmer därför mycket större möjligheter att på ett flexibelt sätt komma åt just den information man är intresserad av. Från användarens synvinkel bjuder dessa ordböcker på mycket större flexibilitet. Man kan selektera vissa typer eller delmängder av information som man är intresserad av och skapa sig en ordbok med en skraddarsydd profil. Informationskategorier som man inte har användning för gömmer man undan, men vid behov kan användaren ändå klicka sig fram till den bortsorterade informationen. Allt oftare lagras ordboken numera under arbetets gång den önskade profilen⁶ och förvandlas på så sätt till ett dynamiskt verktyg.

Samtidigt kan användaren också på ett associativt sätt hoppa runt i ordboken genom att klicka sig fram från ett uppslagsord till ett annat. De nyaste webbordböckerna är multimediala och till texten kopplas bild, ljud och video. I samma applikation finns också tillgång till elektroniska korpusar och inbyggda verktyg (konkordanser, förteckningar över frekventa kollokationer o.d.). Den kvalificerade användaren kan uppehålla sig länge på webben och navigera sig fram genom interna länkar mellan olika ordboksartiklar och genom externa länkar föras direkt till hypertexter och/eller hypermedier. Vi närmar oss begreppet "hyperlexikon", såsom det beskrivs i "Handbok i lexikografi" (Svensén 2004: 525) och såsom det på ett visionärt sätt lanserades av Sue Atkins redan på 1990-talet (Atkins 1994 och 1996).

Sedan digitaliseringen och särskilt sedan uppkomsten av nätbaserade ordböcker har även hela reviderings- och uppdateringsprocessen fått en helt annan karaktär. Det lexikografiska materialet kan uppdateras löpande och revisionen är en permanent process. En ordbok är inte längre någon färdig och statisk produkt, som ligger oförändrad till förfogande tills en ny reviderad upplaga utkommer. Tvärtom, så kan ändringsförslag samlas in och exempelvis läggas ut på nätet med några få dagars eller kanske veckors mellanrum. Dessutom kan det tas hänsyn till användarnas kommentarer och förslag, så att produkten ständigt utvecklas och görs mera anpassad till användarnas krav och behov. Detta dyna-

⁶ Fuzzy SF (Simultaneous Feedback) är t.ex. ett system, utarbetat av Gilles-Maurice De Schrijver, som automatiskt strukturerar informationen i en ordbok på grundval av användarens sökstrategier. På så sätt bildas dynamiska "ultimate customised dictionaries" för varje individuell användare (De Schrijver & Prinsloo 2001).

miska tillvägagångssätt innebär en viss interaktion mellan redaktörerna och användarna. Det finns utrymme för dialog och *user based content* på chattforum eller genom andra kanaler. Dessutom kan redaktörer också erhålla nya insikter genom att studera användarnas faktiska beteende och sökstrategier, som kan rekonstrueras ur logfilerna. Sistnämnda heuristiska förfarande skapar nya möjligheter att fokusera på det som användarna verkligen söker efter (vilka ord? vilka informationstyper?). Ur logfilerna kan man också härleda hur användarna går tillväga, när de inte genast hittar det de letar efter. Analysen av denna sökprocess kan ligga till grund för en optimering av lemmasektionen och för etableringen av nya länkar och datorstyrda instruktioner t.ex. i form av pop ups.

Inom språkteknologin har det på senare tid gjorts ihärdiga försök att utveckla standardiseringsnormer som underlättar återanvändning av vissa verktyg och material⁷. Förr hindrades återanvändning av att resurserna hade utvecklats inom ramen för ett bestämt projekt i ett idiosynkratiskt format och med bestämda skraddarsydd datastrukturer. I det följande beskriver jag två konkreta exempel på återanvändning: det nordiska Lexin-projektet och mina egna erfarenheter från verkstadsgolvet i samband med framtagningen av olika bilinguala ordböcker: nederländska-svenska/danska/finska.

4.1 LEXIN

Ett bra exempel på återanvändning är det svenska LEXIN-projektet, som redan i ett tidigt stadium på 1980-talet koncipierade en serie monodirektionella bilinguala ordböcker för invandrare. Alla utarbetades med samma svenska källspråksdatabas som utgångspunkt. Modellen övertogs senare i det utvidgade nordiska LEXIN-projektet. Tanken bakom denna form för återanvändning var i första hand inspirerad av eller rent av framtvungen av den begränsade marknaden för en stor del av invandrarordböckerna. Därtill kom den ideologiska principen att ordböckerna skulle vara lättillgängliga och billiga. I det nordiska Lexin-projektet har återanvändningsprincipen tillämpats på två plan. Dels tillämpar man även för den danska och den norska versionen samma princip med en gemensam källspråksdatabas och dels har den svenska ordbasen i stort legat till grund för den danska resp. norska makro- och mikrostrukturen.

Naturligtvis är det i och för sig ingen nyhet att ett ordboksprojekt i ett nordiskt land utnyttjar arbete som tidigare utförts inom ordboksprojekt i andra nordiska länder. Men möjligen innebär arbetet på nordiska invandrarordböcker det hittills tydligaste exemplet

⁷ En del standarder har utvecklats inom ramen för europeiska samarbetsprogram som EAGLES och ISLE. Andra viktiga instanser i detta samband är ISO/T37, W3C samt LISA (Localising Industry Standards Association).

på möjligheterna till sådant samarbete, i syfte att så långt som möjligt undvika dubbelarbete. (Bergenholtz & Malmgren 1999: 1)

4.2 Hub and Spoke-modellen tillämpad inom bilingual lexikografi

Under de senaste femton åren har jag som huvudredaktör successivt arbetat med tre bilinguala ordböcker, som har nederländska som källspråk och resp. svenska, danska och finska som målspråk. (Laureys & De Groot, 1996; Laureys et al 2004; Laureys & Moisisio 2012.) På så sätt har jag kunnat fokusera på de metodiska och organisatoriska aspekterna av ett ordboksprojekt och jag har fattat ett särskilt intresse för de metalexikografiska perspektiven. Ovannämnda tre ordböcker uppvisar ett antal likheter och skillnader. Likheter kommer till synes i de färdiga produkterna, eftersom ordböckerna återspeglar de gemensamma målsättningar och lexikografiska principer som ligger till grund för dem. Likheter har sin logiska förklaring i att alla tre ordboksparen skall uppfylla samma funktionskrav, att de är avsedda för parallella målgrupper, har identisk ambitionsnivå vad angår mikro- och makrostruktur, är bifokala samt är utarbetade enligt i stort samma redaktionsprinciper. Alla tre ordböcker omfattar ca 45 000 uppslagsord och ca 75 000 exempel, de är korpusbaserade och behandlar det moderna ordförrådet både i tal och skrift.

Utgångspunkten för den nederländska makro- och mikrostrukturen i dessa ordböcker utgör det s.k. RBN (*Referentiebestand Nederlands*), en generisk korpusbaserad relationell ordbas, utarbetad på uppdrag av *Nederlandse Taalunie*⁸ med det specifika syftet att bilda underlag för ett antal bilinguala ordböcker med nederländska som källspråk. I detta syfte har även redigeringsprogrammet OMBI utvecklats. OMBI är skraddarsytt till RBN-strukturen och är utstyrt med en avancerad vändfunktion på LU-nivå. LU definieras som "lexikalisk enhet", dvs. en viss betydelsevariant av ett visst uppslagsord.⁹ Varje lexikalisk enhet (LU) i källmaterialet länkas i databasen till en eller fler översättningsekvivalenter på målspråket och på detta sätt etableras en kontrastiv databas på LU-nivå. Programmet tillåter redaktören att när som helst byta källspråk och målspråk. RBN omfattar inte bara lexikaliska data (uppslagsord och exempel) utan också olika informationskategorier och annotationer. Informationen för varje uppslagsord omfattar förutom den semantiska profilen, uppgifter om stavning, morfologi, syntaktiska egenskaper och mönster, pragmatik, stilistiska särdrag

⁸ "The Dutch Language Union", det gemensamma Nederländska Språkrådet för Nederländerna och Flandern.

⁹ Vid monosemiska uppslagsord sammanfaller alltså formenheten (FU) med den lexikaliska enheten (LU).

samt geografisk tillhörighet. Dessa annotationer överstiger den information som tas upp i grafiska eller elektroniska ordböcker, men är till stor hjälp under redaktionsprocessen och berikar även databasen för framtida lexikologisk forskning. För att kunna använda RBN till en konkret ordbok måste redaktionen därför först fastställa selektionsprinciper med avseende på makro- och mikrostrukturen. Med tanke på ordbokens funktion(er) och på målgruppens behov måste ett antal uppslagsord strykas och andra läggas till och likaså måste ett relevant urval göras ur det stora antalet exempel i RBN.

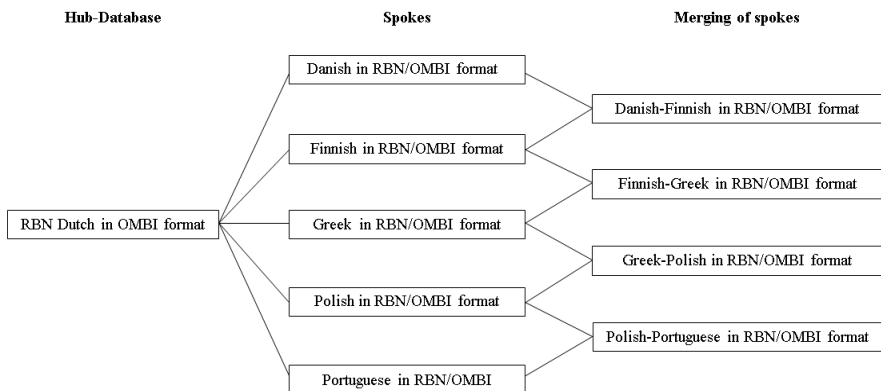
I samband med det nederländsk-danska ordboksprojektet lades ned mycket tid och pengar på en adekvat selektion av uppslagsord och exempel för att åstadkomma en relevant makro- och mikrostruktur för den tilltänkta målgruppen. Denna investering kan ses som produktutveckling av RBN-materialet. Det var just denna tanke på produktutveckling som gjorde att jag bestämde mig för att återanvända materialet och dra nytta av denna investering i ett nytt ordboksprojekt. I projektbeskrivningen för den nederländsk-finska ordboken framsattes förslaget att återanvända den variant av RBN, som vi hade utarbetat i samband med det danska projektet¹⁰. Till grund för detta beslut låg dock inte bara ekonomiska övervägningar utan också en djärvare syn på återanvändning. Från början har tanken varit att komma underfund med vilka möjligheter som ligger i att framställa råmaterial till en bilingual ordbok för ett nytt språkpar på grundval av en multilingual databas. Jag hänvisar i detta samband till *Hub and Spoke*-modellen, lanserad av Willy Martin (2001) (2003) och (2004). Jag ville pröva denna modell i praktiken genom att senare försöka länka ihop den danska och den finska målspråksdatabasen och på semi-automatisk väg framställa en experimentell version av en finsk-dansk ordbok.

I överensstämmelse med ambitionen om att optimera redaktionsprocessen har vi dessutom försökt att återanvända existerande finskt material. Vi erhöll en licens på den elektroniska underliggande databasen till *Suomi-saksa opiskelusanakirja*, som ägs av WSOY-förlag. Materialet har konverterats så att det kunde laddas ned direkt i OMBI. Ovannämnda ordbok, som omfattar 51 000 uppslagsord, är en typisk inlärningsordbok för finstalande som skall lära sig tyska. Fördelen med detta förfarande var att vår databas fylldes från starten med 51 000 finska uppslagsord och ett stort antal exempel. Förutom detta material

¹⁰ Det skall betonas att vi inte blint reproducerade den danska varianten av RBN-databasen, då vi satte igång med den nederländsk-finska ordboken. Redaktionen fick självfallet ta hänsyn till den nya målgruppen och till den kontrastiva relevansen för språkparet nederländska-finska. Men även efter denna bearbetning kan man räkna med att ca 80% av det nederländska materialet är gemensamt för den nederländsk-danska och den nederländsk-finska ordboken.

blev även en annan viktig resurs i förväg laddad ned i databasen. Från KOTUS¹¹ erhöll vi en licens på den annoterade makrostrukturen av den stora monolinguala finska ordboken *Suomen kielen perussanakirja*, som omfattar 93 000 uppslagsord försedda med grammatisk information. I denna ordbok är alla substantiv och verb försedda med en kod, som hänvisar till flexionstabellerna. Dessa grammatiska upplysningar fogades automatiskt till de finska uppslagsorden i vår databas.

I och med att den danska resp. den finska ordbokstexten i stort kan betraktas som en översättning av identiska nederländska uppslagsord och exempel, torde det i princip vara möjligt att länka ord och strängar från båda målspråken till varandra. Utgångspunkten för experimentet med en finsk-dansk ordbok utgör alltså de elektroniska kontrastiva lexikaliska databaser för språkparen nederländska/danska och nederländska/finska, som utarbetades i samband med redaktionen av ordböckerna för motsvarande språkpar. Under hänvisning till *Hub and Spoke*-modellen har jag experimenterat med att foga ihop de två kontrastiva databaserna i en integrerad trespråkig databas. I den integrerade databasen kan det på ett entydigt sätt etableras en länk mellan en dansk och en finsk ordsträng på grund av att de danska resp. de finska uppslagsorden och exemplen är markerade som översättningar av bestämda nederländska uppslagsord och exempel. Jag har med andra ord betraktat materialet i den danska och den finska databasen som två *spokes* länkade till samma *hub*.



Visualisering av Hub and Spoke-Modell

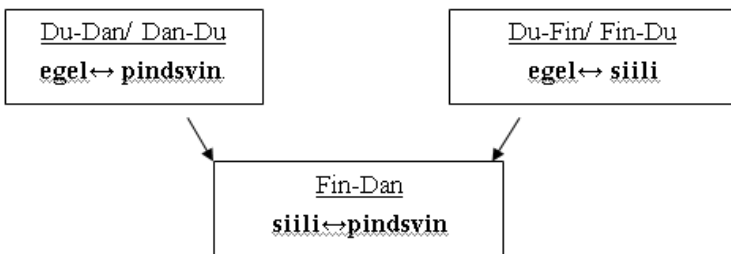
Jag påvisar här den stora fördelen med att både den danska och den finska databasen blev utarbetade på grundlag av samma RBN-version, vilket medför att de

¹¹ Kotimaisten kielen keskus (Institutet för de inhemska språken), Helsingfors.

båda lexikaliska databaserna er maximalt parallella. Därmed är det viktigaste kravet för en meningsfull och framgångsrik *merging* uppfyllt: strukturen i databasen er identisk och de semantiska och grammatiska specifikationerna följer samma mall. Därutöver förekommer ofta identiska exempel. Det innebär att den danska och den finska delmängden i databasen utgör en ideell utgångspunkt för en experimentell länkning, som siktar mot en finsk-dansk kontrastiv databas, som i sin tur kan utgöra en rå version av en finsk-dansk ordbok. Den bakomliggande tanken är hypotesen att, om en finsk LU och en dansk LU är länkade till en och samma LU i den nederländska databasen, så finns det en rimlig sannolikhet att den danska och finska LU även inbördes är adekvata översättningsekvivalenter. Spåret är den gemensamma länk som den finska och den danska LU har till en och samma nederländsk LU. Om A är lika med B och C är lika med A, så råder det även likhet mellan A och C.

Jag har försökt att på semi-automatisk väg härleda ett antal artiklar för en tilltänkt finsk-dansk ordbok utgående ifrån de två föreliggande bilinguala databaserna. Med hjälp av några exempel vill jag illustrera tillvägagångssättet och de uppnådda resultaten. Jag vill visa möjligheterna och begränsningarna med denna procedur och i synnerhet utpeka vilka zoner i ordbokstexten som ägnar sig bäst till semi-automatisk länkning.

Det enklaste fallet är ett monosemiskt nederländskt ord som matchar ett monosemiskt danskt och ett monosemiskt finskt ord. I normalfallet torde det danska och det finska ordet också vara översättningsekvivalenter. I exemplet fungerar det nederländska LU *egel* som en *hub* och den finska och danska översättningen som *spokes*.



Mera invecklat blir det när vi har att göra med polysemiska uppslagsord på danska och finska. För det danska ordet *bølge* finns fem olika betydelser (LU) i databasen, som emellertid svarar mot två olika – i sin tur – polysemiska uppslagsord i finskan, *aalto* och *laine*.

aalto [101.] [I] <NOUN>

- [veden pinnan kohouma] *bolge a. (500) aallot b. (500) korkea aalto c. korkeat aallot høje bolger d. (500) raivoavat aallot e. keinua aalloilla ride på bølgerne f. vaipua aaltoihin forsvinde i bølgerne*
- [värähdysliike] (fys) *bolge a. pitkät aallot langbolge b. lyhyet aallot kortbolge c. lähettää lyhyillä aalloilla NIHIL*
- [kuv: aallon kaltainen ilmiö] *bolge a. uusi aalto NIHIL b. vihreä aalto grøn bolge c. kuumat aallot hedetur*

laine [48.] <NOUN>

- [aalto] *bolge a. valaa öljyä laineille gyde olie på vandene*
- [aaltoa muistuttava kuvio] *bolge a. (500) jonkun hiukset ovat laineilla b. (500) kam-mata laineita hiuksiin c. muotoilla hiukset laineille lave bolger i håret*

Det mest komplicerede fallet är då de olika LU tillhörande ett polysemiskt finskt uppslagsord är länkade till LU som tillhör olika (polysemiska) danska uppslagsord. I dessa fall är distansen mellan de två berörda språkens lexikaliska struktur som störst¹².

kello [101.] <NOUN>

- [aikaa mittaava laite] *ur a. katsoa kelloa se på uret b. kello lyö yksitoista klokken slår elleve c. kello edistää/on edellä uret er foran d. kello jätetään/on jäljessä uret er bagud e. kelloni jättää/edistää kaksi minuuttia mit ur er to minutter bagefter/foran f. kello seisoo uret er gået i stå g. (500) kello on pysähtynyt h. vetää kello trække uret op i. siirtää kelloa eteenpäin NIHIL j. asettaa kello oikeaan aikaan stille uret k. (500) asettaa kellot samaan aikaan l. (500) kellon koneisto*
- [ajankohta] *time a. kysyä kelloa NIHIL b. paljonko/mitä kello on? Hvad er klokken? c. (500) voitteko sanoa, paljonko kello on? d. (500) kello on yksi e. kello on viisi klokken er fem f. kello viideltä klokken fem g. tasan kello kaksitoista NIHIL h. nukkua kellon ympäri sove i tolv timer*
- [soittolaite] *klokke a. soittaa (kirkon)kelloja ringe med (kirke)klokkerne b. kellot soivat NIHIL c. soittaa (ovi)kelloa ringe på d. (500) soittaa kelloa stemme dorklokker som leg e. kello soi det ringer på døren f. (500) hänelle tuli toinen ääni kelloon g. (500) sitten hänelle tuleekin toinen ääni kelloon h. nyt on toinen ääni kellossa nu får piben en anden lyd i. ei vahinko tule kello kaulassa en ulykke kan ske hurtigt, det kan pludseligt gå galt*

Ovanstående exempel, som är genererade på semi-automatiskt vis från den integrerade trespråkiga databasen visar att en hel del översättningar faller på plats

¹² Ännu mera extremt är naturligtvis de fall där det inte föreligger någon översättningsekivalent alls på grund av att ett visst betydelseinnehåll inte är lexikaliserat i ett av de båda språken. I så fall finns det helt enkelt ingen länk, utan uppslagsordets betydelse på språk A återges med en parafraiserande beskrivning på språk B.

med en gång och kan återanvändas i en finsk-dansk ordbok. Detta gäller både *lexikaliska enheter* (LU) och *exempel* (EU). De fall där systemet inte genererar någon översättningsekivalent alls markeras med *NIHIL*. Det implicerar att den nederländska motsvarigheten i systemet inte fungerar som en *hub*, i och med att den endast är länkad till det finska materialet och inte till det danska. Uppslagsordet eller exempelmeningen härstammar i sådana fall alltså inte från RBN. Detta kan bero på att det är en nederländsk översättning av ett finskt uppslagsord eller en finsk exempelmening som antingen har tillfogats av de finska redaktörerna eller som har genomgått drastiska förändringar. De finska exempel som har försetts med koden <500> däremot har sällats ut som uppslagsord eller exempelmeningar i den finsk-nederländska delen och måste följaktligen avlägnas från den tilltänkta finsk-danska ordboken. Det är finskt material som befinner sig i databasen enbart som översättning av nederländska uppslagsord och exempel, men som inte kvalificerar sig som källspråkmaterial beroende på t.ex. låg frekvens, udda lexikalisering eller brist på kulturell specificitet. Länken är i så fall unidirektionell¹³. De finska ord(strängarna) visas här bara för att illustrera proceduren.

På grundval av ett antal experimentella körningar kan vi dra slutsatsen att det är möjligt att på ett semi-automatiskt sätt härleda en rå version av en finsk-dansk ordbok ur en trespråkig databas, där nederländska som *hub* är länkad till danska och till finska som *spokes*. På den utarbetade finsk-danska databasen kan självfallet senare vändfunktionen i OMBI tillämpas, så att vi även får en rå version av en dansk-finsk ordbok. Den stora vinsten är hur som helst att all relevant grammatisk information och alla stilistiska och pragmatiska specifikationer redan är befintliga för danska och finska uppslagsord. Den semi-automatiska länkningen genererar alltså även de befintliga annotationerna till uppslagsorden, de lexikaliska enheterna samt exemplen. Självfallet finns det en hel del efterredigering att utföra, men arbetsinsatsen för en ny ordbok reduceras ansevärt och återanvändningen av de befintliga resurserna ger en god utdelning.

¹³ The editorial team opted for an integrated contrastive bi-directional database, from which both the Dutch-Finnish and Finnish-Dutch volumes are derived. To this end all translation links in the database have been marked according to their directionality status: *bi-directional*, if the translation is univocal, adequate and relevant in both directions. In that case both the Dutch and the Finnish LUs and EUs are represented in both volumes either as a source or as a target item; *unidirectional Du→Fi*, if the Finnish item only figures as a translation unit, but does not qualify as a LU or a EU in the Finnish-Dutch volume; *unidirectional Fi→Du*, if the Dutch item is represented only as a translation unit in the Finnish-Dutch volume and must be prohibited from appearing in the Dutch-Finnish direction. (Laureys 2007: 301.)

Ingen eller endast marginell efterredigering krävs när det gäller monosemiska ord, lexikaliserade flerordsfraser och idiom samt enkla icke-kontextuella exempelfraser. Tillsammans utgör dessa fall i själva verket drygt 70% av ordbokstexten. Betydligt mera efterredigering behövs när det gäller polysemiska substantiv och verb, de allra flesta adjektiv och överlag när det gäller korpusexempel. Det ska härvid beaktas att översatta exempel inte är genuint korpusbaserade och att det föreligger en viss risk för bias mot hub-språket. Dessutom ska den kontrastiva relevansen prövas för varje nytt språkpar. När det gäller funktionsord, högfrekventa substantiv och verb, som oftast förevisar en hög grad av polysemi, duger däremot resultaten av den semi-automatiska länkningen knappast till en direkt bearbetning.

5. Integrerad lexikalisk databas eller ”hyperlexikon”

Eftersom det både är dyrt och tidskrävande att sammanställa lexikaliska databaser, är det önskvärt att försöka reducera kostnaderna och att försöka förlänga ”livstiden”, dvs. *the usable life time* (Banski & Przepiorkowski 2011:34 och Simons & Bird 2008:16) för dessa resurser. “To be sustainable, the results of our work must transcend computer environments, communities of practice, domains of application, and especially the passage of time.” (Bird & Simons, 2008:4). Bird & Simons (2004:4) anger sex nödvändiga och tillräckliga villkor för en hållbar användning av språkliga resurser: ”the resource must be extant, discoverable, available, interpretable, portable and relevant”. Det läggs stora krav dels på tillgängligheten och den formella aktualiseringen av språkliga resurser och dels på den innehållsmässiga relevansen, vilket kräver en ständig ajourföring av resurserna.

Avslutningsvis fokuserar jag på de nya utmaningar som ligger i strävan efter hållbarhet inom lexikografin. En integrerad lexikalisk databas är den moderna lexikografins svar på hållbarhetskravet. En integrerad databas förutsätter att man optimerar arbetet och utför det så vitt möjligt ”en gång för alla”. För att åstadkomma en integrerad databas får man tänka långsiktigt. Databasen skall kunna ligga till grund för många olika ordböcker och lexikografiska verktyg. Den skall vara robust och återanvändbar. Därtill skall designen vara öppen och dynamisk och samtidigt skall den konformera sig till fasta och kontrollerbara standarder, som beskrivs explicit i metadokument och som ständigt kan valideras. Man använder sig av kalibrerade verktyg enligt internationella normer. Etiketterna är entydiga och standardiserade, så att de alltid svarar mot en viss infor-

mationstyp och tvärtom¹⁴. Databasstrukturen är enkel och entydig. Layoutmässigt har den integrerade basen en flat struktur, så att layout och formgivning kan bestämmas per produkt eller produktserie. Input i strukturen är medieneutral. Alla produkter framställs med hjälp av en transparent redaktionsmanual, så att produktionsprocessen snabbas upp och görs mera effektiv.

Databasen företer en modulär mjukvarudesign, vilket innebär att den rymmer möjligheten att en framtida redaktör skall kunna selektera relevanta kategorier på alla nivåer inom ramen för ett konkret projekt, utan att hierarkin eller den överordnade strukturen förstörs. Databasen består m.a.o. av olika deldatabaser, som kan definieras och härledas på beställning. Samtidigt finns det länkar till andra databaser och deldatabaserna emellan. Denna s.k. interoperabilitet torde vara en markant fördel med en integrerad databas.

En ordbok utvecklas inte längre som ett isolerat projekt, utan som en del av ett samlat övergripande projekt. En enskild ordbok blir i så fall ett derivat ur en annoterad lexikalisk databas. Det innebär att varje redaktionsgrupp slipper göra en hel del arbete, som redan är gjort av andra. Feedback från användarna, kan, om det är relevant, generaliseras och förbättringar eller uppdateringar sker på en gång för alla derivat.

Särskild uppmärksamhet kräver procedurerna för förvaltningen av materialet. För att databaserna skall kunna användas inom ramen för nya projekt måste de inte bara vara lättillgängliga, utan även beskrivas på ett transparent sätt i relevanta metadokument.¹⁵ Kontinuerlig uppdatering av databaserna och säkring av tillgängligheten av data med nya versioner av relevant software är en förutsättning för en permanent och stabil tillgång i snabbt förändrande omgivningar.

Stora insatser med avseende på infrastrukturella resurser och databasarkitektur har gjorts inom ramen för det europeiska projektet CLARIN¹⁶ och inom

¹⁴ Ett viktigt begrepp i detta samband är *persistent identifier*. Med *persistent identifier* avses en permanent, unik och entydig etikett som identifierar en digital resurs (text, audio, video, databas etc) helt oberoende av dess fysiska befintlighet eller placering. En *persistent identifier* garanterar att resursen alltid kan hittas och konsulteras via nätet.

¹⁵ Det läggs mycket vikt på kodning, långsiktig digital arkivering och på tillgänglighetskrav. Viktiga initiativ har tagits inom ramen av ELSNET (European Network of Excellence in Speech and Language Technology).

¹⁶ "CLARIN (Common Language Resources and Technologies Infrastructure) is committed to establishing an integrated and interoperable research infrastructure and technology for language resources and its technology. It aims at removing the current fragmentation, to offer a stable, persistent, accessible and extendable infrastructure and therefore enabling e-Humanities". (Broeder et al. 2008: 1457.)

DAM-LRprojektet¹⁷ och mer nyligen även inom OLAC (Open Language Archives Community)¹⁸.

Vi är på väg mot en mer integrerad, en mera användarvänlig, en mera funktionell, effektiv, flexibel och hållbar lexikografi. Perspektivet är långsiktigt, vägen är utstakad, nödvändigheten påvislig, men det finns mycket arbete kvar. Lexikografen i Norden står vid en skiljeväg. Med den uppbyggda erfarenheten och expertisen, med viljan till ett öppet samarbete har vi alla möjligheter att bidra på ett positivt sätt till denna utveckling. Av flera sektionsföredrag under konferensen har det framgått att utvecklingen av webbordböcker och hyperlexikon är i full gång i alla de nordiska länderna.

LITTERATUR

- Atkins, Sue & Fillmore, Charles J. et al., 1994 : The Dictionary of the Future: a Hypertext Database. Presentation and on-line demonstration at the Xerox-Acquilex Symposium on the Dictionary of the Future. Uriage.
- Atkins, Sue, 1996: Bilingual Dictionaries: Past, Present and Future. I: Euralex'96 Proceedings, (eds.) Gellerstam, M. , J. Järborg, S.-G. Malmgren, K. Norén, L. Rogström and C. R. Pappmehl. Gothenburg, S. 515-590.
- Bergenholtz, Henning & Malmgren, Sven-Göran (red), 1999: Förord. I: Lexico Nordica 6, S.1-3.
- Broeder, Daan, Nathan, David, Strömquist, Sven & van Veenendaal, Remco, 2008: Building a federation of Language Resource Repositories; the DAM-LR project and its continuation within CLARIN. www.lrec-conf.org/proceedings/lrec2008/pdf/370_paper.pdf, S.1457-1461.
- Brundtland, Gro Harlem, 1987: Vår felles framtid. Oslo (även som: Our Common Future: The World Commission on Environment and Development. Oxford.
- De Schryver, Gilles-Maurice & Prinsloo, D. J., 2001 : Fuzzy SF: Towards the ultimate customised dictionary. I: Studies in Lexicography 11.1, S. 97–111.
- Hannesdóttir, Anna Helga, 1998: Lexikografihistorisk spegel. Den enspråkiga svenska lexikografins utveckling ur den tvåspråkiga. Meijerbergs arkiv för svensk ordforskning 23. Göteborg.
- Hovmark, Henrik, 2005: Tre franske miniordbøger til rejsebrug og daglig tale. I: Lexico Nordica 12, S.13-33.

¹⁷ The DAM-LR project aimed at investigating the needs and laying the basis for establishing a federation of language archives that can offer their users a shared virtual space of resource location facilities and resource access. (Broeder et al. 2008: 1457.)

¹⁸ OLAC står för Open Language Archives Community. "OLAC is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by: (i) developing consensus on best current practice for the digital archiving of language resources, and (ii) developing a network of interoperating repositories and services for housing and accessing such resources." (Simons & Bird 2008:8.)

- Kielitoimiston sanakirja, 2006. (Uusittu ja laajennettu versio Suomen kielen perussanakirjasta). Kotimaisten kielten tutkimuskeskuksen julkaisuja 140. Helsinki.
- Laureys, Godelieve, 2007: Optimizing Procedures for the Making of Bilingual Dictionaries and the Concept of Linking Contrastive Lexical Databases. I: *International Journal of Lexicography* 20. 3, S. 295-311.
- Laureys, Godelieve & De Groot, Hans, 1996: *Van Dale Handwoordenboek Zweeds-Nederland /Nederlands-Zweeds*. Utrecht (även som: *Norstedts nederländsk-svensk/svensk-nederländska ordbok*. Stockholm).
- Laureys, Godelieve et al., 2004: *Gyldendals Hollandsk-Dansk ordbog*. Köpenhamn.
- Laureys, Godelieve & Moisis, Maritta, 2012: *Prisma Groot Woordenboek Nederlands-Fins en Fins-Nederlands/ Hollanti-Suomi ja Suomi-Hollanti Suursankirja*. Houten – Antwerpen.
- Martin, Willy, 2001: Linkability, the Hub-and-Spoke Model and the (semi-)automatic derivation of a German-French Dictionary. I: H. E. Wiegand (ed.). *Studien zur zweisprachigen Lexikographie VI, Germanistische Linguistik* 16. Hildesheim: Olms, S. 67-91.
- Martin, Willy, 2003: Lexicography, Lexicology, Linking and the Hub-and-Spoke Model. I: W. Botha (ed.). *Festschrift D. Van Schalkwijk*. Stellenbosch, S. 237-249.
- Martin, Willy, 2004: SIMuLLDA, the Hub-and-Spoke Model and Frames or How to Make the Best of Three Worlds? I: *International Journal of Lexicography*, 17 (2): 175-187.
- Nathan, David, 2006: Proficient, permanent, or pertinent: aiming for sustainability. I: Barwick, L. and Thieberger, N., (eds.). *Sustainable Data from Digital Fieldwork*. The University of Sydney, S. 57-68.
- Simons, Garry F., Bird, Steven, 2008: Toward a global infrastructure for the sustainability of language resources. *Proceedings of the 22nd Pacific Asia Conference on Language, Information and Computation. PACLIC 22*. S.87-100.
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm.

Godelieve Laureys

Institutionen för nordiska språk vid universitetet Gent
 godelieve.laureys@ugent.be