

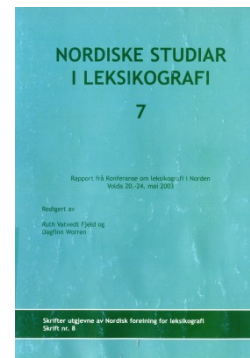
NORDISKE STUDIER I LEKSIKOGRAFI

Titel: Nord-Lexin

Forfatter: Tove Bjørneset og Ásta Svavarsdóttir

Kilde: Nordiske Studiar i Leksikografi 7, 2005, s. 52-66
Rapport frá Konferanse om leksikografi i Norden, Volda 20.-24. maj 2003

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.

Nord-Lexin

The original LEXIN-series, which consists of bilingual dictionaries between Swedish and various minority languages, was produced to meet the needs of immigrants in Sweden. In the 1990s, Swedish authorities offered their neighbouring countries to use the material as a basis for the development of similar dictionaries for other Nordic languages. The idea was that by replacing the Swedish entry words, examples, idioms, etc., by equivalents in another language, an extensive use could be made of the translations, already done between Swedish and various target languages, by linking them semi-automatically to a new source language. Pilot projects were launched in Norway, Denmark and Iceland. Consequently the complete dictionary base was adapted to Norwegian, and the project has already resulted in several on-line LEXIN-dictionaries in Norway. Work on the Icelandic project was taken up in 2001, in collaboration with the Norwegian project group. As both projects use the same database, an interesting by-product has emerged, i.e. a trilingual corpus of Nordic equivalents: Swedish–Norwegian–Icelandic.

In the paper, the authors discuss the experiences from the Icelandic and Norwegian LEXIN-projects of transferring a dictionary base from one source language to another. Furthermore, a semi-automatic linking procedure between Norwegian, as a newly established source language, and English, as a target language, already on hand in the Swedish base, is discussed and compared to a manual translation to Tamil, a new target language in the Norwegian LEXIN. Finally, there are some speculations about the possibilities presented by the tripartite database, e.g. with respect to the further development of the material to serve new needs and new functions.

Innleiing

I Sverige er det sidan 1980-talet utgjeve ei mengd ordbøker for innvandrarar under namnet *LEXIN – Språklexikon för invandrare*. Dei fleste av desse er tospråklege, frå svensk til ulike minoritetsspråk, og dei er alle baserte på eit felles svensk ordbokunderlag. Dette underlaget inneheld eit utval lemma med forklaringar eller definisjonar, døme i form av frasar, samansetjingar, avleiingar, kollokasjonar og idiom med forklaringar, og for nokre ord også informasjon om mellom anna stilnivå, syntaks og grammatikk, noko som tener til å forklare orda sine tydingar, bruk og grammatiske eigenskapar.¹ Sjølve ordbokunderlaget er utgjeve som ei einspråkleg svensk ordbok, *Svenska ord – med uttal och förklaringar* (1994).

På 1990-talet oppstod den tanken at LEXIN burde kunne eksporterast, ikkje berre som ein modell for liknande ordbokprosjekt i andre land, men også sjølve materialet. Ein gjekk ut frå at dersom ein kunne finne ekvivalentar til delar av LEXIN-materialet i andre språk, så burde eit tilsvarende omfang av det opphavlege kjeldespråket kunne «byttast ut» med nye kjeldespråk. Deretter kunne ein vonleg med utgangspunkt i det nye kjeldespråket etablere elektronisk tilgang til og bruke om att omsetjingar frå svensk til rundt 20 språk. I teorien ville det altså berre vere naudsynt å omsetje manuelt frå det nye kjeldespråket til målspråka

i tilfelle med manglande ekvivalens eller dersom det av andre grunnar var gjort endringar i ordbokbasen med dei nye kjeldespråka.

Frå svensk side fekk dei nordiske landa så tilbod om å bruke det svenske Lexin-materialet som grunnlag for egne LEXIN-ordbøker. Som følge av den nære slektskapen mellom dei nordiske språka (med unntak av finsk) og det at språksamfunna også har mange andre likskapstrekk, var det ikkje urealistisk å gå ut frå at ein vesentleg del av grunnlaget burde kunne «resirkulerast». Vidare var det klart at dersom ideen kunne realiserast, ville ein kunne spare både tid og pengar. Dette er særleg viktig for at ein i det heile teke skal kunne lage ordbøker for så små målgrupper som det her ofte er snakk om. Danmark, Noreg og Island takka difor ja til tilbodet, og i desse landa vart det gjennomført pilotprosjekt på slutten av 1990-talet (Holmboe 1999; Bjørneset 1999, 2000, 2001, 2002, 2003; Svavarsdóttir 1999).

Det norske LEXIN-prosjektet har sidan halde fram, og frå 2002 og utover er det utvikla fleire nettbaserte ordbøker: ei einspråkleg norsk ordbok og ei norsk–engelsk ordbok, kvar med 34 000 ord (ordtalet er rekna ut på grunnlag av lemma, samansetjingar og avleiingar), og ei mindre norsk–tamilsk ordbok. Rundt 2 700 av orda i basen er illustrerte. Illustrasjonane har form av eit vedlegg med 33 tematisk ordna sider. Her kan ein søkje båvegar mellom alle språka i det norske prosjektet per i dag: tamil, engelsk, bokmål og nynorsk.² Arbeidet med to nye ordbøker, norsk–kurmandji og norsk–sorani, tek til sommaren 2003.

Det islandske prosjektet vart sett i gong att i slutten av 2001. Målsetjinga er å lage ei islandsk–engelsk nettbasert ordbok, og den fyrste delen vil etter planen verte publisert i 2005. Det har i heile perioden vore eit nært samarbeid mellom LEXIN-prosjekta i Noreg og på Island.³ Mellom anna arbeider den islandske prosjektgruppa direkte i den norske databaseapplikasjonen *Termbase* via Internett.

I den fyrste delen av denne artikkelen gjer forfattarane greie for røynslene dei har frå arbeidet med å overføre og adaptere det svenske ordbokgrunnlaget til norsk og islandsk. I den andre delen vert det gjeve ein kort omtale og samanlikning av dei to fyrste norske målspråksprosjekta med ei norsk–engelsk og ei norsk–tamilsk ordbok, nærare bestemt om skilnaden mellom «ombruk» av omsetjingar og manuell omsetjing. Den tredje delen av artikkelen handlar om den skreddarsydd databaseapplikasjonen *Termbase*, med vekt på den vidareutviklinga som har funne stad i høve til det svenske prosjektet. I artikkelen sin siste del kjem forfattarane inn på det nye NORD-LEXIN-materialet som har oppstått ved at den norske og den islandske prosjektgruppa arbeider i ein felles databaseapplikasjon; ei ekvivalent samanstilling av leksikografisk informasjon på svensk, norsk og islandsk.

Tilpassing av ein svensk ordbokbase til norsk og islandsk

Tilpassinga av det svenske kjeldespråksmaterialet til norsk og islandsk gjekk i fyrste rekkje ut på å ekvivalere alt leksikalsk materiale frå den svenske basen med dei to respektive språka. Resultatet av ekvivaleringa avgjer i kva grad målspråka i det svenske LEXIN-materialet seinare vil kunne koplast med dei nye kjeldespråka. For at den overførte ordbokbasen skal vere representativ for dei nye kjeldespråka, må ein eventuell modifikasjon av ordutval, døme og anna i det svenske materialet vurderast. Til sist inneber tilpassinga byte av semantisk, syntaktisk og grammatisk informasjon om dei svenske orda og uttrykka

med tilsvarende informasjon på dei nye kjeldespråka. Denne fasen vert ikkje handsama her, då han ikkje har direkte samanheng med det hovudemnet vi har valt for denne artikkelen: våre røynsler med gjenbruk av eit ordbokmateriale ved å overføre eit kjeldespråk til to grannespråk.

I Noreg starta arbeidet ved årsskiftet 1996/1997 med å overføre det svenske kjeldespråkmaterialet og på grunnlag av det etablere eit tilsvarende norsk materiale. Det vart årleg lagt ned rundt to årsverk til gjennomføring og kontroll av dette arbeidet, og ved utgangen av 1999 var prosessen fullført. Meir enn 80 prosent av det norske kjeldespråkmaterialet var då overført manuelt frå den svenske LEXIN-databasen gjennom etablering av ekvivalensar. Det norske materialet vart deretter supplert med mellom anna nye ordbokpostar, døme, ordspråk og idiom. Det var ei klar målsetjing å leggje inn døme i så mange former som råd i ordbokpostane, ut frå den tanken at det ligg eit læringspotensial i det at målgruppa får sjå oppslagsorda i fleire språklege kontekstar. I nettbaserte ordbøker er ikkje plassøkonomisering ei problemstilling, noko som er positivt for brukarane.

Den norske ordbasen er i dag noko større enn den svenske: medan den svenske inneheld rundt 30 000 lemma og døme i form av samansetjingar og avleiingar, er det tilsvarende talet for den norske ordbasen rundt 34 000. Den norske basen inneheld rundt 15 000 døme av ulike typar og rundt 2 500 idiom med forklaringar. Tilsvarende tal finst enno ikkje for islandsk, då overføringa av basen til islandsk ikkje er fullført. Generelt er det fleire ulikskapar mellom islandsk og svensk enn mellom norsk og svensk, og på bakgrunn av dette kan ein vente at ein lågare prosentdel av materialet kan ekvivalerast i det islandske materialet enn i det norske.

I redigeringsgrensesnittet er den svenske ordbokinformatjonen og dei norske ekvivalentane presentert samla i ei fast rekkjefølgje og med engelske omsetjingar. Den islandske prosjektgruppa har dermed to kjeldespråk og eit målspråk å ta utgangspunkt i ved overføringa, noko som er nyttig ved kartlegging og val av islandske ekvivalentar. Nye norske lemma kan også ekvivalerast til islandsk, dersom dei skulle vere aktuelle også i islandsk samanheng.

Fullekvalens mellom to språk kan sjeldan ventast på andre område enn fagområde med stringent terminologi. Ein kan vere etter måten trygg på at det er fullekvalens mellom det svenske oppslagsordet *dermatologisk* og det norske likelydande oppslagsordet, som bår refererer til læra om hud og hudsjukdomar. Tilsvarende er det fullekvalens mellom dei svenske og norske oppslagsorda *galax* og *galakse* og *immigration* og *immigrasjon*. Desse orda har dessutan nær identisk form på svensk og norsk. Det gjeld derimot ikkje islandsk, der terminologien i hovudsak byggjer på heimlege ordstammar, i samsvar med den islandske puristiske tradisjonen.

Ulikskapar på uttrykkssida skal likevel ikkje i seg sjølv hindre kongruens på innhaldsida. Mellom anna er det fullekvalens mellom svensk og norsk *geologi* og den tilsvarende islandske termen *jarðfræði*. Det viser seg likevel ofte at det er manglande ekvivalens, sjølv i fagtermar. Det ville til dømes ikkje ha vore noko problem med substantivet *dermatolog* (ikkje med i den svenske basen), som har *húðsjúkdómalæknir* som islandsk ekvivalent, medan adjektivet ville hatt føreleddet *húðsjúkdóma-* som næraste parallell, sjølv om ekvivalensen ikkje er fullstendig. *Galax* og *galakse* har dei to potensielle ekvivalentane *stjörnuþoka* og *stjörnukerfi*, men det er ikkje problemlaust å velje mellom desse. Vidare kan ikkje *immigration* og *immigrasjon* ekvivalerast med eitt islandsk ord, trass i at omgrepet er kjent: ein kan seie *innflutningur fólks* (eigentleg 'import av menneske'), men i demo-

grafiske rapportar er det nemninga *aðfluttir frá útlöndum* ('migrerte frå utlandet') som ver brukt.

Eit overflatefenomen som formell parallellitet er likevel ikkje noko premiss fo

fullekvalens, tvert imot er det mange fallgruver i høve til det å setje sin lit til slike formmessige likskapstrekk på det same semantiske området i to språk, jamfør mellom anna såkalla «falske vener». Eit kjent døme på dette fenomenet er svensk *kjol*, som har norsk *skjørt* og islandsk *pils* som ekvivalentar, og altså ikkje norsk *kjole* eller islandsk *kjóll*, som derimot er ekvivalentar til svensk *klänning*. Eit anna døme er svensk *glasögon*, som er ekvivalent med norsk *briller*, og ikkje *glassøye*, som på norsk tyder 'et kunstig øye av glass (eller et eventuelt annet egnet materiale)'. Den islandske ekvivalenten *gleraugu* svarar derimot formelt så godt som semantisk til det svenske ordet.⁴

Like fullt kan identisk eller nær identisk form ofte vere eit fyrste haldepunkt for eventuell ekvivalens, i det minste mellom så nærskylde språk som svensk og norsk, og har slik trass alt hatt ein viss praktisk verdi i ekvivaleringsarbeidet: Det svenske lemmaet *immigration* vil naturleg ha sin norske ekvivalent i *immigrasjon*, medan det svenske lemmaet *invandring* har *innvandring* som norsk ekvivalent. Dette gjev parallellitet i registra, men ein må sjølv sagt gå varsamt fram. Eit naturleg ynske om størst muleg grad av overlapping mellom det svenske, det norske og det islandske LEXIN-materialet kan føre til at ein kan kome til å ignorere dei til tider ørsmå stil- og registerskilnadene som det er ein del av mellom språka. Teoretisk sett kan det nemleg vere slik at medan ein i svensk allmennspråk berre brukar *immigration*, brukar ein i norsk allmennspråk berre *innvandring*. Eit anna døme er dette: I svensk språk er *varg* den allmennspråklege nemninga for 'ulv', medan det kan synast som om *ulv* berre vert brukt i litterært eller arkaisk språk. I norsk og islandsk er det motsett; her er *ulv* eller *úlfur* det frekvente. Slike allmennspråklege ulikskapar i register må reflekterast i det ekvivalerte materialet.

I det norske prosjektet vart det bestemt at ekvivalens i lemmafeltet skulle vere ei absolutt kriterium for om det skulle byggjast vidare på ein svensk ordbokpost eller ikkje. Dersom eit svensk lemma ikkje kunne ekvivalerast, vart dermed heile ordbokposten utelaten frå det nye norske kjeldespråkmaterialet. Programmerarane i prosjektet vurderte dette som den sikraste løysinga med tanke på føreståande halvautomatiske omsetjingar. Den negative konsekvensen ved dette var at eventuell ekvivalerbar informasjon i ordbokposten også gjekk tapt.

I andre tilfelle var det berre det svenske lemmaet i ein ordbokpost som kunne ekvivalerast, og ikkje resten av informasjonen i han. Det er ulike årsaker til dette. Semantisk sett hadde den svenske informasjonen klare motsvar på norsk, men problemet var at den ikkje kunne ekvivalerast i samsvar med lemmaet. Eit døme på dette er den svenske ordbokposten med lemmaet *gris*, med mellom anna idiomet *köpa grisen i sekken*. I det korresponderande norske idiomet støytter vi på eit anna dyr – ein *kjøper katta i sekken* – og det norske idiomet må i den norsk LEXIN-basen difor plasserast under eitt eller fleire av lemmaa *katt*, *kjøpe* eller *sekk*. Sjølv om det semantiske innhaldet i hovudsak er identisk, så kunne ikkje idiomet verte ståande i den opphavlege posten. Det same gjeld det svenske idiomet *ha skinn på näsan*, som tyder å vere myndig og fast, og det norske idiomet *ha bein i nesa*. Dette idiomet kunne ekvivalerast fordi det var oppført under lemmaet *näsa*, og ikkje *skinn*. Same ulikskap finst i islandsk, der bae idioma formelt tilsvavar dei norske, høvesvis *kaupa köttinn i sekknum* og (*hafa*) *bein i nefinu*. I det islandske prosjektet har ein så langt valt å ekvivalere bae idioma på grunnlag av semantisk parallelli-

tet, med tilvising til rett lemma når det islandske idiomet tilhører eit anna lemma enn det tilsvarende svenske, til dømes *köttur* ('katt') i staden for *gris*, som i dømet over. Tilvisinga tener då som ei påminning om at idiomet seinare skal leggjast inn under rett oppslagsord. Det beste ville sjølvstakt ha vore å finne ein felles metode for å ta vare på sambandet mellom slike parallellar i basen på tvers av lemma. Slik kunne ein også teke betre vare på sambandet til dei talrike målspåka.

For at den overførte ordbokbasen skal vere representativ for dei nye kjeldespråka, er det mykje å ta omsyn til, både under ekvivaleringsarbeidet og når resultatet av det vert vurdert med omsyn til eventuelle manglar eller skeivskapar. For det fyrste må ein vere merksam på at ord og uttrykk som i og for seg kan ekvivalerast, ikkje treng ha same relevans i andre kjeldespråk som i svensk. Det kan her vere snakk om særsvenske omgrep, som namn på svenske landskap (til dømes *Dalarna*) eller institusjonar (til dømes *Svenska akademien*). I slike tilfelle vert den aktuelle ordbokposten eller informasjonen inaktivert. I andre tilfelle er det snarare snakk om å gje lemmaet større eller mindre vekt med omsyn til kva relevans det har i det nye kjeldespråket eller språksamfunnet, til dømes ved å leggje til eller utelate informasjon. Mange dyr som er vanlege i Skandinavia finst ikkje på Island, mellom andre *ulv* og *elg*, men dei har likevel islandske nemningar, og mange av desse er heller ikkje uvanlege i bruk, til dømes i litteraturen. Statusen til slike ord i dei ulike språka er likevel ulik, så det kan vere aktuelt å redusere talet på døme, samansetjingar og anna i dei islandske ordbokoppslaga.

Den motsette situasjonen kan også oppstå, det vil seie at eit lemma har større relevans i det nye kjeldespråket og difor krev ein meir utførleg omtale i form av ein definisjon eller ei forklaring, fleire døme, og så vidare. I svensk er ordet *lava* utelukkande ein geologisk term som har fått ein kort definisjon i LEXIN-basen, og det same er tilfelle i den norske. På islandsk er derimot ekvivalenten *hraun* eit sentralt ord som fyrst og fremst viser til eit naturfenomen som i høg grad pregar det islandske landskapet, og det er difor mellom anna mange samansetjingar som bør vere representerte i den islandske basen. Tilsvarende har ord som *hytte* og *tur* større relevans for norsk kultur og språk enn for svensk og islandsk, noko dei tre ordbokbasane bør reflektere. Dette kan gjerast på fleire måtar: (1) Eit ord er med i ein base, men ikkje i dei andre; (2) Dersom eit ord er meir sentralt i eitt av språka, kan det krevje ein meir detaljert definisjon og/eller fleire tydingsledd; (3) Talet på døme, idiom, samansetjingar og så vidare kan ventast å variere med omsyn til oppslagsordet sin relevans i det aktuelle språket. Vidare må ein vere merksam på at det truleg finst relevante norske og/eller islandske ord som ikkje går fram som ekvivalentar til svensk, og kjeldespråkmaterialet må difor etter ekvivaleringsarbeidet samanliknast med heimlege kjelder og supplerast dersom det skulle vise seg å vere slike manglar.

I det norske prosjektet vart det arbeidd systematisk på visse omgrepsområde. Mellom anna er strukturen i det svenske helsevesenet så ulikt det norske at det vart naudsynt å inaktivere ein del postar og etablere nye postar med særnorske lemma. For å sikre eit representativt utval av norske samfunnsomgrep, vart det samla inn materiell frå offentlege instansar, bankar og andre, ekserpert terminologi, og på grunnlag av dette oppretta nye norske postar. Tilsvarende er det også på Island ekserpert ord og uttrykk frå informasjonsmateriell om mellom anna offentlege institusjonar for å kontrollere lemmautvalet i basen. Ein del andre nye norske ordbokpostar fungerer som reine supplement. Mellom anna vart det lagt inn rundt 200 frekvente datatermar i samsvar med retningsliner frå Norsk språkråd. Om det svenske prosjektet ønskjer det, står dei fritt til å ekvivalere det nye norske ordbok-

Felt i Termbase	Ekvivalentar
Svensk lemma	hund
Ekvivalent norsk lemma	hund
Generert engelsk omsetjing	dog
Svensk ordklasse	subst
Ekvivalent ordklasse	subst
Generert engelsk omsetjing	noun
Svensk definisjon	(hus)dyret <i>Canis familiaris</i>
Ekvivalent norsk definisjon	(hus)dyret <i>Canis familiaris</i>
Generert engelsk omsetjing	the (domesticated) animal <i>Canis familiaris</i>
Svensk døme	gå ut med hunden
Ekvivalent norsk døme	gå tur med hunden
Generert engelsk omsetjing	take the dog for a walk
Svensk idiom	vara som hund och katt
Ekvivalent norsk idiom	være som hund og katt ("passe svært dårlig sammen")
Generert engelsk omsetjing	be like cat and dog, agree like cats and dogs
Svensk samansetjing	jakthund
Ekvivalent samansetjing	jakthund
Generert engelsk omsetjing	sporting dog
Svensk samansetjing	hund~mat -en
Ekvivalent samansetjing	hunde~mat -en
Generert engelsk omsetjing	dog food
Svensk samansetjing	hund~valp -en
Ekvivalent samansetjing	hunde~valp -en
Generert engelsk omsetjing	puppy

Tabell 1.

materialet.

På Island vil kontrollfasen verte omfattande. I tillegg til konstatering av ekvivalens mellom språka, vil den innebere kontroll av at alle relevante omgrep frå LEXIN-basen er dekte og at den overførte ordbokbasen dekkjer det islandske ordforrådet på ein god nok måte. Ved ekvivalering til norsk har ein hatt støtte i den formelle parallelliteten mellom svensk og norsk, i det at ein i fyrste runde har teke stilling til eventuell ekvivalens mellom det svenske ordet og det formelt parallelle norske ordet. Dersom det synte seg at det svenske og det norske ordet ikkje berre var formelt parallelle, men også med omsyn til deira tyding, funksjon og status i språket, vart ordet valt som norsk ekvivalent, trass i at det fanst andre norske synonym som også kunne veljast på grunnlag av deira innhald og funksjon.

Den store formelle parallelliteten inneber også at det i stor grad er ord-til-ord-ekvival-

ens mellom svensk og norsk – det er med andre ord sannsynleg at eit innhald som vert uttrykt med eitt substantiv i svensk, vert uttrykt med eitt substantiv i norsk. Den formelle parallelliteten mellom islandsk og svensk eller norsk er langt mindre, mellom anna på grunn av den islandske purismen, nokre døme er allereie nemnde over. Det er svært vanleg at det til svenske og norske adjektiv svarar eit føreledd i islandsk, til dømes *internationell* vs. *alþjóða-*, *individuell* vs. *einstaklings-*, *industriell* vs. *iðnaðar-*, og så vidare. Som oftast er det semantisk ekvivalens mellom det svenske og det norske adjektivet og det islandske føreleddet, men spørsmålet er om dei kan skrivast opp som ekvivalentar trass ulike morfo-syntaktiske eigenskapar. Dei fleste føreledda er dessutan genitivsformer av dei tilsvarende substantiva, mellom andre *einstaklingur* ('individ') og *iðnaður* ('industri'), så ein må også tenkje på tilknytninga til dei aktuelle lemmaa. Verb som det svenske *påverka* og norske *påvirke* kan nemnast som døme på ord som ikkje svarar til eitt einskilt islandsk ord, men ein kollokasjon: *hafa áhrif á* ('ha innverknad på'). Sjølv om ein ikkje ønskjer å gjere eit slikt ordsamband til oppslagsord ved å setje det opp som ekvivalent til det svenske og norske lemmaet, skal innhaldet likevel takast med i basen, mest sannsynleg som idiom under *áhrif*. For å ta vare på det semantiske sambandet i slike tilfelle, er det lagt inn ei tilvising til eit spesialfelt i databasen. Parallelt er det lagt inn ei tilvising frå ordbokposten med oppslagsordet *påvirke*, som ikkje har nokon direkte ekvivalent i islandsk, til oppslagsordet *áhrif*. Seinare vert det kontrollert at ordsambandet *hafa áhrif á* kjem med. Det er dessutan planar om å samanlikne resultatet av ekvivaleringsarbeidet med heimlege kjelder, mellom anna lister frå frekvensundersøkingar, for å sikre at sentrale islandske ord kjem med i basen.

To norske målspråksprosjekt: halvautomatisk vs. manuell omsetjing

Norsk–engelsk ordbok

Pilotprosjektet med den norsk–engelske ordboka (34 000 ord) starta i 2000 og var fullført hausten 2001. Engelsk vart valt som testspråk av fleire årsaker: Fyrst og fremst opna det for at alle deloppgåver i pilotprosjektet kunne gjennomførast av dei same personane som hadde teke del i ekvivaleringsarbeidet, og som difor hadde god kjennskap til prosjektet og databaseapplikasjonen *Termbase* (sjå neste kapittel). Vidare var engelsk eit målspråk alle meistra, noko som var viktig når det norske kjeldespråksmaterialet skulle koplast med dei svenske engelsk-omsetjingane. Ikkje mindre viktig skulle dette språkvalet seinare vise seg å vere for kvalitetskontrollen av feil og systematiske feiltypar i dei genererte omsetjingane. Valet av engelsk vart også vurdert som positivt for målgruppa: For mange innvandrara fungerer engelsk som eit *lingua franca* eller mellomspråk, og det finst eit stort utval ordbøker mellom engelsk og ulike minoritetsspråk som kan vere til hjelp når ein skal lære seg norsk.

Koplinga av dei svenske elektroniske engelsk-omsetjingane med det norske kjeldespråksmaterialet tok kort tid. Ekvivalensprosenten for dei ulike informasjonstypene var i overkant av 80, og ein tilsvarende del av omsetjingane kunne dermed genererast. All særnorsk informasjon i form av nye lemma og ny informasjon i ekvivalerte ordbokpostar vart seinare omsett manuelt. Alle omsetjingar vart granska før og etter prosessen, og det vart gjort berre mindre korreksjonar. Dei fleste ordbokpostane var korrekte, som dette utdraget frå ordbokposten *hund* viser:

Felt i Termbase	Ekvivalentar
Svensk lemma	hund (illustrert)
Ekvivalent lemma bokmål (BM)	hund
Ekvivalent lemma nynorsk (NN)	hund
Ekvivalent islandsk lemma	hundur
Svensk ordklasseinformasjon	substantiv
Norsk ordklasseinformasjon	substantiv
Islandsk ordklasseinformasjon	nafnorð, karlkyn
Svensk definisjon	(hus)dyret Canis familiaris
Ekvivalent definisjon BM	(hus)dyret Canis familiaris
Ekvivalent definisjon NN	(hus)dyret Canis familiaris
Ekvivalent islandsk definisjon	húsdýrið Canis familiaris
Svensk dømme med lemmaet	gå ut med hunden
Ekvivalent dømme BM	gå tur med hunden
Ekvivalent dømme NN	gå tur med hunden
Ekvivalent islandsk dømme	fara út með hundinn
Særnorsk dømme BM	hunden biter ikke
Særnorsk dømme NN	hunden bit ikkje
Ekvivalent islandsk dømme	hundurinn bitur ekki
Svensk idiom med lemmaet	vara som hund och katt
Ekvivalent idiom BM	være som hund og katt ("passe svært dårlig sammen")
Ekvivalent idiom NN	vere som hund og katt ("passe svært dårleg saman")
Ekvivalent islandsk idiom	vera eins og hundur og köttur "(um fólk) eiga mjög illa saman")
Særnorsk idiom på BM	gå i hundene ("gå til grunne")
Særnorsk idiom på NN	gå i hundane ("gå til grunne")
Ekvivalent islandsk idiom	fara í hundana ("fara illa út úr lífinu, lenda á botninum")

Felt i Termbase	Ekvivalentar
Særnorsk idiom BM	man skal ikke skue hunden på hårene ("man skal ikke dømme noen etter deres (dårlege) ytre")
Særnorsk idiom NN	ein skal ikkje skode hunden på håra ("ein skal ikkje dømme nokon etter deira (dårlege) ytre")
Ekvivalent islandsk idiom	–
Svensk samansetjing	jakthund
Ekvivalent samansetjing BM	jakthund
Ekvivalent samansetjing NN	jakthund
Ekvivalent islandsk samansetjing	veiðhundur
Svensk samansetjing	Hund~mat –en
Ekvivalent samansetjing BM	hunde~mat –en
Ekvivalent samansetjing NN	hunde~mat –en
Ekvivalent islandsk samansetjing	hunda~matur -inn
Svensk samansetjing	hund~valp -en (illustrert)
Ekvivalent samansetjing BM	hunde~valp -en
Ekvivalent samansetjing NN	hunde~kvalp -en
Ekvivalent islandsk samansetjing	– (er ekvivalent som annan informasjonstype, sjå under)
Ekvivalent islandsk omsetjing	hvolpur -inn (er lagt inn som eige oppslagsord)
Særnorsk samansetjing BM	hunde~bånd -et (illustrert)
Særnorsk samansetjing NN	hunde~band -et
Ekvivalent islandsk samansetjing	hunda~ól -in
Særnorsk samansetjing BM	hunde~slede -n (illustrert)
Særnorsk samansetjing NN	hunde~slede -n
Ekvivalent islandsk samansetjing	hunda~sleði -inn
Særnorsk samansetjing BM	hunde~spann -et (illustrert)
Særnorsk samansetjing NN	hunde~spann -et
Ekvivalent islandsk samansetjing	hunda~eyki -ð

Felt i Termbase	Ekvivalentar
Svensk morfologi	hunden hundar
Morfologi BM	hunden hunder hundene
Morfologi NN	hunden hundar hundane
Islandsk morfologi	hundurinn hunds(ins) hundar(nir)
Svensk lemma	röt~månad
Ekvivalentert lemma bokmål (BM)	hunde~dager
Ekvivalentert lemma nynorsk (NN)	hunde~dagar
Ekvivalentert islandsk lemma	hunda~dagar
Svensk ordklasse	subst
Ekvivalentert ordklasse BM	subst
Ekvivalentert ordklasse NN	subst
Ekvivalentert islandsk ordklasse	nafnorð, karlkyn
Svensk definisjon	tiden mellom 23 juli og 23 augusti, som k�nnetecknas av fuktig v�rme
Ekvivalentert definisjon BM	perioden mellom 23 juli og 23 august (ofte med fuktig varme)
Ekvivalentert definisjon NN	perioden mellom 23 juli og 23 august (ofte med fuktig varme)
Ekvivalentert islandsk definisjon	t�mabilið fr� 23. j�l� til 23. �g�st
Svensk morfologi	-m�naden
Morfologi BM	-dagene
Morfologi NN	-dagane
Islandsk morfologi	-dagarnir
Grammatisk informasjon BM	bare i flertall
Grammatisk informasjon NN	berre i fleirtal
Grammatisk informasjon islandsk	bara � fleirt�lu
Nytt lemma BM	hunde~galskap
Nytt lemma NN	hunde~galskap
Ekvivalentert islandsk lemma	hunda~�ði

Felt i Termbase	Ekvivalentar
Ordklasse BM	subst
Ordklasse NN	subst
Ekvivalent islandsk ordklasse	nafnorð, hvorugkyn
Definisjon BM	en smittsom og dødelig virussykdom hos hunder og andre pattedyr som også kan ramme mennesker, rabies
Definisjon NN	ein smittsam og dødeleg virussjukdom hjá hundar og andre pattedyr som også kan ráke menneske, rabies
Ekvivalent islandsk definisjon	smitandi og banvænn veirusjúkdómur í hundum og öðrum spendýrum sem líka getur borist í menn
Morfologi BM	-galskapen
Morfologi NN	galskapen
Islandsk morfologi	-æðið

Tabell 2 Utdrag frå ordbokpostane *hund*, *hundedagar* og *hundegalskap* i redigeringsgrensesnittet i *Termbase*.

Ei viktig oppgåve i møtet med det tospråklege ordbokmaterialet der målspråket engelsk var relatert til svensk, var å avdekkje manglande samsvar mellom det overførte kjeldespråket og målspråket som skuldast delevivalens mellom det opphavlege og det nye kjeldespråket, og deretter vurdere eventuell semantisk innsnevring eller utviding av dei aktuelle omgrepa. Semantisk utviding av eit omgrep er i stor grad eit spørsmål om å slå saman ulike tydingar av polyseme ord. I tilfelle der ein etablert norsk ekvivalent til eit svensk lemma vart vurdert å ha ei vidare tyding enn det som kan lesast ut av den svenske definisjonen, stod valet mellom å utvide den norske definisjonen, slik at den dekkjer den samla tydinga av ordet, eller å opprette ein ny post som dekkjer den eller dei tydingane som ikkje var med i den ekvivalerte ordbokposten. Det å opprette nye postar for å dekkje tydingar som ligg svært nær allereie dekte tydingar kan vere like uheldig som å leggje inn for mange tydingar i éin post. Sett frå ståstaden til dei som skal ha oversyn over og kompensere for eventuelt manglande ekvivalens mellom omgrep i eit overført kjeldespråk og eit målspråk, er likevel oppretting av nye postar å føretrekkje. Ein eventuelt utvida definisjon gjer det ofte naudsynt å leggje inn fleire ord i lemmafeltet på målspråket. Alle eksisterande omsetjingar av det opphavlege definisjonsfeltet vil då seinare måtte utvidast manuelt i samsvar med den nye definisjonen på det overførte kjeldespråket.

I nokre ekvivalerte ordbokpostar er det norske definisjonsfeltet redigert slik at det avvik frå det svenske. I dei fleste av desse tilfella har dei norske definisjonane vorte supplerte med meir informasjon. Eit døme på dette er den svenske ordbokposten med lemmaet *dart*, ekvivalent til *darts*. Den svenske synonymdefinisjonen 'pilkastningsspel' har fått sitt nor-

ske motsvar i parafrasen 'et spill der man kaster piler mot en skive'.

Definisjonsfelt er også endra i tilfelle der informasjon om lemmaet har vorte flytta frå andre informasjonsfelt i ordbokposten til definisjonsfeltet. Dette er gjort på grunnlag av ei prinsippavgjerd tidleg i det norske prosjektet om at all informasjon som tener til omgrepsavklaring, som antonym, informasjon om bruksområde og eventuell biletleg bruk eller tilvising til andre postar for meir informasjon, av omsyn til brukaren bør presenterast samla i definisjonsfeltet.

Norsk–tamilsk ordbok

Tamilane i Noreg har hatt ein prekær mangel på ordbøker i mange år. Det einaste som eksisterte var ei mindre ordliste som var del av eit læreverk i norsk for innvandrarakar. På bakgrunn av dette avgjorde oppdragsgjevaren for prosjektet at det skulle lagast ei norsk–tamilsk ordbok trass i at det ikkje fanst tamilske omsetjingar i det svenske prosjektet ein kunne byggje på, og at alt omsetjingsarbeid difor måtte gjerast manuelt.

Det tamilske alfabetet består av 247 teikn, inkludert diakritiske variantar. Sjølv om både tal og form skil seg mykje frå latinske teikn, kunne dei tamilske omsetjarane etter mindre justeringar bruke *Termbase* i arbeidet, noko som viste seg å ha mange praktiske føremoner. Den viktigaste av desse var at omsetjarane fekk direkte tilgang til engelskomsetjingane av kjeldespråksmaterialet. Dette var fyrst og fremst til hjelp ved omgrepsavklaringa, men medverka også på andre måtar til å lette omsetjingsarbeidet.

Arbeidet med ordboka starta sommaren 2001 og den nettbaserte utgåva vart lansert våren 2003. Ordboka er illustrert og inneheld rundt 6 500 ord. Det er planar om å utgje ordboka også i trykt utgåve.

Om ombruk av omsetjingar i norsk LEXIN

Det er så langt gjennomført to svært ulike målspråksprosjekt i det norske LEXIN-prosjektet: ei norsk–engelsk ordbok med rundt 34 000 ord (såkalla MAXI-utgåve) og ei norsk–tamilsk ordbok med rundt 6 500 ord. Medan den fyrste ordboka i hovudsak kunne genererast ved å kople rundt 80 prosent av det norske kjeldespråksmaterialet med elektroniske svenske engelsk-omsetjingar, måtte den andre omsetjast manuelt.

Med så ulike utgangspunkt seier det seg sjølv at ressursbehovet i dei to delprosjekta måtte verte svært ulikt og at det difor kan hende er feil å setje dei opp mot kvarandre. Det er likevel verdt å merkje seg at arbeidet med den norsk–engelske ordboka krevde berre halvparten av dei ressursane som den norsk–tamilske ordboka gjorde.

I det norske LEXIN-prosjektet har det synt seg at det kan vere store ressursar å spare på ombruk av omsetjingar. Særleg utslagsgjevande vil slik ombruk truleg vere for realisering av ordbokprosjekt for små målgrupper. Denne metoden gjer det dermed muleg å lage ordbøker for fleire og mindre målgrupper enn tidlegare.

Databaseapplikasjonen *Termbase*

Sidan det andre driftsåret har det norske LEXIN-prosjektet brukt *Termbase*, ein skreddarsydd redigeringsapplikasjon i databaseform utvikla i samarbeid mellom pro-

grammerarane og leksikografane i prosjektet. Applikasjonen vert kontinuerleg raffinert i høve til teknologiske nyvinningar og nye aktivitetar. Sidan 2001 har det islandske Lexin-prosjektet arbeidd med sitt materiale i same applikasjon med redigeringsstilgang via Internett. I redigeringsgrensesnittet er det fleispråkleg parallellvising av alle informasjonsfelte i ordbasen. I dei unikt koda felte ligg all informasjon i ordbokopplaga sortert etter språk og type, mellom anna lemma, ordklasse, forklaring eller definisjon, døme, morfologi og uttale. Materialet kan sorterast på fleire måtar og med ulike kombinasjonsløyningar. Prefiks i form av bokstavar og bokstavkombinasjonar skil informasjonen på ulike språk og informasjonstypene frå kvarandre. Resultatet av den fleispråklege samstillinga kan karakteriserast som eit leksikografisk parallellkorpus. Ordbasen i det norske LEXIN-prosjektet ligg i dag føre på svensk, bokmål og engelsk, og delar av han også på nynorsk, islandsk og tamil. I dei to næraste åra skal sorani og kurmandji (to av fleire kurdiske dialektar eller variantar) og somali også inngå som målspråk.

I 2000 starta sju biletkunstnarar arbeidet med å utvikle nye norske Lexin-illustrasjonar på grunnlag av den svenske biletløysinga. Bakgrunnen for dette var at dei svenske illustrasjonane ikkje lenger fanst i original og at ein del av dei var forelda eller lite representative for norske tilhøve. Moderne teknologi opna for eit meir pedagogisk brukargrensesnitt i det norske og islandske prosjektet. Det er i dag utvikla eit system med interaktive koplingar mellom dei tre nettbaserte LEXIN-ordbøkene og dei 33 tematiserte biletsidene som samla illustrerer rundt 2 700 av orda i ordbøkene. Biletsidene kan brukast på to måtar: Som vedlegg til ordbøkene (i ordbøkene er alle illustrerte ord markerte med eit blome-ikon) eller som ei sjølvstendig informasjonseining som brukarane kan gå til direkte og deretter velje kva biletema dei vil sjå på. Ved å klikke på eit bileto-objekt får dei sjå kva skriftspråkleg uttrykk det har. Når eit bileto-objekt er markert, kjem det fram ei svart line som knyter det saman med det korresponderande ordet tydeleg markert i ei skjermordliste. Brukarane kan også gå motsett veg; frå dei einskilde orda i skjermordlista til bileto-objekta. For alle bileto-objekta kan det søkjast båtvegar mellom alle søkjespråka per i dag. Dette brukargrensesnittet opnar for at også analfabetar og andre med særskilde behov kan ha læringsutbytte av norsk LEXIN. Funksjonaliteten opnar for at LEXIN også kan brukast som eit læremiddel.

All informasjon i *Termbase* kan på tilsvarende måte som i redigeringsgrensesnittet også gjerast søkbar i brukargrensesnittet, men dette er eit kostnadsspørsmål. Norsk LEXIN inneheld alle rettskrivingsformer og bøyingsparadigme som er tillatne i høve til læreboknormalen for bokmål. Mange av brukarane søker på bøyde former, og morfologifelte skal difor gjerast søkbare. I det islandske prosjektet er det planar om å knyte heile bøyingsparadigme til lemmaa på eit seinare tidspunkt, sjølv om det i fyrste omgang berre er dei kanoniske kasus- og tempusformene som er lagde inn i det islandske morfologifeltet. Førremonen ved å kunne finne fram til rett lemma frå alle skrift- og bøyingsformer ville vere stor, ikkje minst for urøynde ordbokbrukarar og for personar med minimale kunnskapar i kjeldespråket.

Nord-Lexin som eit (felles)nordisk prosjekt

Det er i dag nasjonale LEXIN-prosjekt i Noreg, på Island og delvis også i Danmark i tillegg til opphavslandet Sverige. Trass i dette kan NORD-LEXIN likevel ikkje karakteriserast som eit fellesnordisk prosjekt, då det ikkje er noko aktivt samarbeid mellom prosjekta, med

unntak av det norske og det islandske. Dette samarbeidet består fyrst og fremst i bruken av ein felles databaseapplikasjon. På grunnlag av dette har det etter kvart vakse fram ei stor og stadig aukande mengd svenske, norske og islandske ekvivalentar, sjølv om det må leggjast til at denne ekvivaleringa er gjort frå svensk til norsk og islandsk, og i noko mindre grad frå norsk til islandsk, men enno ikkje vice versa. Det svensk–norsk–islandske kjeldespråkmaterialet er tilgjengeleg i ei fleispråkleg parallellvising, og ein kan søkje og redigere i materialet på dei måtane ein måtte ynskje.

Det mangesidige ekvivaleringsarbeidet i ein felles database for alle dei nasjonale prosjekta er ein stor gevinst, både med omsyn til det som heile tida har vore hovudføremålet med NORD-LEXIN – å lage ordbøker for innvandrarar – og eventuelle nye funksjonar i framtida. Det har i nokre år til dømes vore planar om ei islandsk–svensk ordbok med ei islandsk–norsk ordbok som parallelt prosjekt. Det har vore drøfta om desse prosjekta kan og eventuelt skal knytast til Nord-LEXIN.

Vi kjenner ikkje til at det finst eit tilsvarande materiale i eit slikt omfang som NORD-LEXIN frå før. Fordi materialet er ekvivalent, er samanstillinga svært informativ både med omsyn til samanlikningar av lemma, omgrepsavklaringar i form av allmennspråklege forklaringar (definisjonar, parafrasar, synonym eller antonym), lemmaet i praktisk bruk i døme (frasar, kollokasjonar, samansetjingar, avleiingar, ordspråk og idiom med forklaringar) og informasjon om uttale. I tillegg kjem for ei lang rekkje lemma meir informasjon om mellom anna stilnivå, grammatikk og syntaks. Dette kan gje forskarar og andre unike høve til å studere og analysere språklege likskapar og ulikskapar mellom dei nordiske grannespråka.

Sjølv om materialet i si noverande form er skreddarsydd for innvandrarar i Norden, meiner vi at det finst fleire nye utviklingsvegar, bruksområde og målgrupper. Det enkle språket kombinert med eit fleksibelt brukargrensesnitt og talrike tematiserte illustrasjonar opnar for at NORD-LEXIN med noko vidareutvikling også kan eigne seg for til dømes svenske, norske eller islandske barn, unge eller vaksne som skal lære seg eit av grannespråka.

Litteratur

- Bjørneset, Tove 1999: «Ordboksprosjektet NORDLEXIN–N». *LexicoNordica* 6:35–45.
- Bjørneset, Tove 2000: «Dictionaries for Immigrants and Refugees in Norway». I: Apollon, Daniel, Odd-Bjørn Fure og Lars Svåsand (red.). *Approaching a New Millennium: Lessons from the Past – Prospects for the Future*. Proceedings of the 7th Conference of the International Society for the Study of European Ideas (ISSEI), 14–18 August 2000, University of Bergen, Norway. Bergen [CD-ROM].
- Bjørneset, Tove 2001: «Introduksjon til ordboksprosjektet NORDLEXIN–N». I: Gellerstam, Martin, Kristinn Jóhannesson, Bo Ralph og Lena Rogström (red.) *Nordiska studier i lexikografi* 5. Rapport från Konferens om lexikografi i Norden, Göteborg 27–29 maj 1999. Skrifter utgivna av Nordiska föreningen för lexikografi 6. Göteborg, 44–53.
- Bjørneset, Tove 2002: «The Dictionary Project Nordlexin–n». I: Gottlieb, Henrik, Jens Erik Mogensen og Arne Zettersten (red.). *Symposium on Lexicography X*. Proceedings of the Tenth International Symposium on Lexicography May 4–6, 2000 at the University of Copenhagen. *Lexicographica Series Maior* 109. Max Niemeyer Verlag. Tübingen, 31–42.
- Bjørneset, Tove 2005: «The NORDLEXIN–N Dictionary Project: Part II». I: Gottlieb,

- Henrik, Jens Erik Mogensen og Arne Zettersten (red.). *Symposium on Lexicography XI*. Proceedings of the Eleventh International Symposium on Lexicography. May 2–4, 2002 at the University of Copenhagen. Lexicographica Series Maior 115. Max Niemeyer Verlag. Tübingen (under utgjeving).
- Gellerstam, Martin 1999. «LEXIN – lexikon för invandrare». *LexicoNordica* 6: 3–17.
- Holmboe, Henrik 1999: «Det danske bidrag til NORDLEXIN». *LexicoNordica* 6: 55–64.
- Svavarsdóttir, Ásta 1999: «LEXIN på Island? Funderingar kring isländska invandrarordböcker.» *LexicoNordica* 6: 19–33.
- Svenska ord – med uttal och förklaringar* 1994: Andra upplagan. Skolverket/Norstedts, Stockholm.

Notar

1. Ein meir utførleg omtale av LEXIN-prosjektet og bakgrunnen for det finst m.a. i Gellerstam (1999).
2. Heimeside til norsk LEXIN: <http://decentius.aksis.uib.no/lexin.html>
3. På same måte som i Sverige er det også i Noreg og på Island statleg prosjekt- og finansieringsansvar (det norske Læringscenteret og det islandske undervisningsministeriet), medan prosjektgruppene held til på HIT-senteret (etter 01.08.03: AKSIS) ved Universitetsforskning Bergen og Orðabók Háskólans i Reykjavík. Artikkelforfattarane er prosjektleiarar i dei to landa.
4. I svensk og islandsk er pluralformene *glasögon* og *gleraugu* tvitydige, i motsetnad til det norske *glassøye*, som berre kan ha den eine av tydingane.