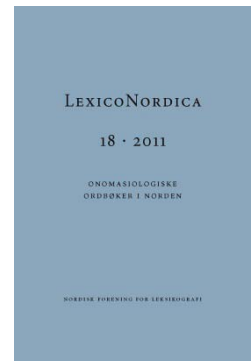


LexicoNordica

Titel: FinnWordNet och kulturbundna ord
Forfatter: Nina Martola
Kilde: LexicoNordica 18, 2011, s. 111-134
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

Betingelser for brug af denne artikel

Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

FinnWordNet och kulturbundna ord

Nina Martola

In this paper I analyze the beta version of the Finnish WordNet from a societal and cultural point of view. FinnWordNet has been created through translation of the Princeton WordNet (Lindén & Carlson 2010). Certain lexicographical problems are to be expected due to the method.

Essential terms referring to Finnish society structure and Finnish culture cannot be found in the present FinnWordNet. The English definitions for words that can be found are not adequate from a Finnish viewpoint and the concept hierarchies are only partly relevant.

Constructing a WordNet is a laborious process and starting by translating is not a bad alternative. The linking to the English WordNet comes 'for free'. If the work had started by building up an independent WordNet for Finnish the linking process would have been very time-consuming. But pure word translation from L1 to L2 without any supplementing work does, of course, not result in an adequate description of the target language L2.

1. Inledning

Den första versionen av WordNet vid Princetons universitet kan sägas ha förelegat 1985, och under årens lopp har ordnätet bearbetats och utvecklats på många olika vis (Miller 1998). För närvarande (augusti 2011) är det version 3.1 som är aktuell och som ligger fritt tillgänglig på webben (<http://wordnetweb.princeton.edu>). Ett ordnät är i korthet enligt Miller (1995:39):

an online lexical database designed for use under program control. English nouns, verbs, adjectives, and adverbs are

organized into sets of synonyms, each representing a lexicalized concept. Semantic relations link the synonym sets

Den amerikanska WordNet-databasen har varit incitamentet till ett stort antal ordnät på andra språk. Ett sjukspråkigt ordnät, Euro-WordNet, skapades i Europa under åren 1996–1999. Sedermera har WordNet-versioner kommit till för ett flertal språk. (<http://www.illc.uva.nl>)

Under slutet av 2009 och början av 2010 skapades ett ordnät för finska språket. Metoden man valde för detta var översättning av det amerikanska ordnätet (se närmare Lindén & Carlson 2010) och översättningsarbetet utfördes manuellt på mycket kort tid. En av tankarna med att skapa ett finskt ordnät genom översättning var att nätet då genast skulle fungera som en tvåspråkig källa. Betaversionen av FinnWordNet (i fortsättningen FWN) ligger i skrivande stund fritt tillgänglig på webben. I sina konklusioner över processen konstaterar Lindén & Carlson (2010:139) att behovet av att förbättra FWN kommer att utvärderas. De räknar med att åtminstone undersöka vilka (högfrekventa) finska ord som saknas och vilka betydelser som saknas för existerande ord. Vidare nämner de att en separat studie kommer att göras för att utvärdera om det finns delar av FWN som inte har adekvat ontologisk struktur för finska.

Denna undersökning är en helt fristående analys av vissa aspekter av den första versionen av det finska ordnätet, och den baserar sig i huvudsak på den version som var tillgänglig våren 2011. Ändringar och förbättringar som gjorts efter det är inte systematiskt beaktade. Jag gör analysen utifrån min bakgrund som tvåspråkslexikograf, inte med språkteknologiska ögon. Jag kommer att koncentrera mig på en typ av ord som mig veterligen inte har diskuterats särskilt mycket i WordNet-sammanhang, nämligen ord som är nära förknippade med det samhälle språket används i.

FWN ska alltså kunna fungera som en tvåspråkig ordbok, men

jag antar att ett lika viktigt (om inte ännu viktigare) syfte är att det finska ordnätet ska fungera som en beskrivning av det finska språket och det finska samhället. Här ser jag som tvåspråkslexikograf stora problem. Utan ett stort anpassningsarbete kan en vänd ordbok – vilket FWN i detta skede måste betraktas som – knappast tjäna som en adekvat beskrivning av det nya källspråket. Detta är trivialt i lexikografiska sammanhang, och det kan verka som att slå in vidöppna dörrar att ägna frågan en studie. Men FWN ska utvecklas vidare, och jag hoppas att jag kan bidra till arbetet genom att peka på konkreta problem som en lexikograf lätt hittar. Finska språket förtjänar ett gediget ordnät.

I det följande kommer jag först att se på definitionerna (i WordNet-termer *glosser*) för ett antal finska ord som på olika vis ansluter sig nära till det finska samhället. Definitionerna är inte översatta till finska utan föreligger på engelska. Det är bara själva orden, eller närmare bestämt ordgrupperna, *synseten*, som översatts till finska i det finska ordnätet.¹ Ekvivalensproblematik tar jag däremot inte systematiskt upp, eftersom jag är fel person att bedöma finsk-engelska motsvarigheter.

FWN är ju ett onomasiologiskt verk, och därmed ägnar jag en rätt stor del av studien åt att diskutera begreppshierarkierna, som i likhet med definitionerna är direkt övertagna från Princetons WordNet.

FWN kan laddas ner till den egna datorn. Det har jag inte gjort, utan jag har utnyttjat ordbasen via det sökgränssnitt som ligger fritt tillgängligt på webben. Ord nätet går att använda som en tvåspråkig ordbok. Om man klickar för *Show translations* visas de engelska ursprungsorden till varje översatt ord både i betydelseöversikterna och i begreppshierarkierna.

1 Noga taget borde det egentligen, benämningen *WordNet* ~ *ordnät* till trots, handla om begrepp som ska översättas (jfr Ruppel 2011), men översättarna har utfört uppdraget som ordöversättning.

2. Betydelsebeskrivningarnas relevans

Ett språk är nära förbundet med det samhälle det används i, och samhälls- och kulturtypiska ord brukar vålla tvåspråkslexikografer problem. I det följande kommer jag därför att undersöka hur begreppen ”president” och ”län” samt begrepp som hänför sig till skolsystem och folkslag är behandlade i FWN.

2.1. Begreppet ”presidentti” (’president’)

Betydelsebeskrivningen för ordet *presidentti* ’president’ i FWN visas i exempel 1.

1. Yhdysvaltain presidentti, presidentti, ylipäällikkö, ** -- (the person who holds the office of head of state of the United States government; ”the President likes to jog every morning”)
2. presidentti, ** -- (the chief executive of a republic)
3. presidentti, ** -- (the judge who presides over a supreme court)
4. Yhdysvaltain presidentti, presidentti, ylipäällikkö, ** -- (the office of the United States head of state; ”a President is elected every four years”) ***

Exempel 1.²

Betydelserna 1 och 4 hänför sig alltså till USA:s president, vilket är bakvänt ur ett finskt användarperspektiv. Betydelse 2 är den allmänna betydelsen och kan i själva verket snarast ses som ett överbegrepp för begreppet ”president” i olika länder (Finland,

2 Svensk övers. ’Förenade staternas president, president, överbefälhavare’. Asteriskparen är länkar till sidor där användaren kan lägga till egna synonymförslag. Funktionen har kommit till efter januari 2011.

USA, Frankrike, Tyskland, Chile osv.). Att USA:s president ges prominens i det amerikanska ordnätet kan accepteras om hierarkisystemet är upplagt så, men i ett finskt ordnät borde i så fall den finländska presidenten på motsvarande sätt lyftas fram, inte den amerikanska.

2.2. Begreppet ”lääni” (’län’)

Ordet *lääni* ’län’ anges ha bara en betydelse: ”the domain controlled by a feudal lord”. Förklaringen hänför sig till ordet *fiefdom* som kommer upp om man ber om den engelska översättningen. Det är visserligen en adekvat översättning av ordet *lääni* men inte i dess moderna betydelse, och ska ett ordnät beskriva finska språket måste självfallet den moderna användningen beaktas. Ord som hör nära samman med *lääni* är t.ex. *lääninhallitus* ’länsstyrelse’ och *maaherra* ’landshövding’. De hittas inte i FWN. Men också länen är numera historia i Finland. År 2010 avskaffades län och länsstyrelser och i stället indelades landet i regioner förvaltade av regionförvaltningsverk, *aluehallintovirasto*. Ordet saknas i FWN. I ett helt uppdaterat finskt ordnät borde det alltså anges tre olika betydelser eller användningar för ordet *lääni*: 1) den feodala enheten, 2) en tidigare förvaltningsenhet i Finland och 3) en förvaltningsenhet i Sverige. Ordningföljden kan diskuteras.

2.3. Begrepp inom skolsystemet

Det följande beskrivningsområde jag ska analysera är det finländska skolsystemet. Jag börjar med en kortfattad redogörelse för hur systemet ser ut i Finland och anger de finska benämningarna på de olika begreppen och markerar med [Ø] om orden inte hittas i FWN. I Finland har vi allmän läroplikt, *oppivelvollisuus* [Ø] (alltså inte skolplikt). De elever som genomför sin läroplikt går i grundskola (årskurserna 1–9) *peruskoulu*. I den finska skollagen

och skolförordningen avser *peruskoulu* numera bara de enskilda skolorna medan själva utbildningen för årskurserna 1–9 benämns *perusopetus* [Ø] 'grundläggande utbildning'. Grundskolan delas in i *alakoulu* (årskurs 1–6) och *yläkoulu* [Ø] (årskurs 7–9). I allmänspråket förekommer fortfarande i någon mån också de tidigare benämningarna *ala-aste* 'lågstadium' och *yläaste* 'högstadium'. Efter genomgången läroplikt går eleverna vidare till andra stadiets utbildning *toisen asteen koulutus* [Ø] och söker då antingen till gymnasium, *lukio*, eller till en yrkesläroanstalt *ammattioppilaitos* [Ø], i dagligt tal *ammattikoulu* 'yrkesskola'.

Definitionerna för de skolord som återfinns i FWN är inte adekvata ur finsk synvinkel eftersom de hänför sig till det amerikanska skolsystemet. För *peruskoulu* anges det att det avser *ala-aste*, *alakoulu*, *oppikoulu* och definitionen lyder "a school for young children; usually the first 6 or 8 grades" (definitionen gäller am. *grammar school*). Ordet *oppikoulu* 'läroverk' hänför sig till ett gammalt finländskt skolsystem som avvecklades på 70-talet. Ordet *yläaste* anges ha två ungefärliga betydelser: 1) "a secondary school usually including 7th and 8th grades" och 2) "a school for students intermediate between elementary school and college; usually grades 9 to 12". De finska förklaringarna (synonymerna) är *yläaste ja lukio*, *lyseo*, *lukio*, *yläaste* ('högstadium och gymnasium', 'lyceum', 'högstadium', vilket antagligen fungerar förhållande väl som närmaste motsvarigheter för att beskriva det amerikanska systemet men som blir ytterst rörigt i fråga om finska skolor.

I det moderna finska skolsystemet utgör gymnasierna separata utbildningsenheter. I det gamla läroverkssystemet, dit också ordet *lyseo* 'lyceum' hänför sig, hade man mellanskola och gymnasium inom en och samma skola. Mellanskolan var fem- eller sexårig, och eleverna sökte till läroverk efter fyra eller fem år i folkskola. Ordet *lyseo* lever visserligen kvar i namnet på en del skolor, men som skolterm är beteckningen försvunnen utom med avseende på tiden fram till 70-talet.

Skolsystemen i olika länder avviker från varandra och ekvivalenterna kan aldrig bli mer än ungefärliga, men i ett finskt ordnät borde definitionerna beskriva det finska systemet, och system från olika tider borde hållas i sär.

Därmed aktualiseras också den principiella frågan om i vilken grad ett ordnät ska beakta även tidigare förhållanden. Det är en komplicerad fråga och det finns inget självklart svar, utan svaret är beroende av vad ordnätet är avsett att användas till.

2.4. Begreppen ”suomalainen” (’finne; finsk’) och ”karjalainen” (’karelare; karelsk’)

Den följande typen av ord jag ska studera är folkslagsbeteckningar. Ordet *suomalainen* anges som substantiv ha betydelsen ”a native or inhabitant of Finland” medan betydelseangivelse för adjektivet är ”of or relating to or characteristic of Finland or the people of Finland”. Det framgår inte att *suomalainen* också kan ha den snävare innebörden ’finskspråkig’, en betydelse som anges i den moderna, enspråkiga ordboken *Kielitoimiston sanakirja* (KT).

Ordet *karjalainen* ’karelare; karelsk’ har definitionen ”a member of the Finnish people living in Karelia in northwestern European Russia”. Här saknas uppgift om att ordet liksom *suomalainen* också kan vara adjektiv. Betydelseangivelsen är inte riktigt adekvat. Karelarna i Ryssland är inte ett finskt folk utan ett folk besläktat med finnarna. Men dessutom kan *karjalainen* syfta på invånarna i det finländska landskapet Karelen, som i sin tur hör till det finska folket. Karelen införlivades till största delen med Sovjetunionen vid andra världskriget och befolkningen evakuerades i stor utsträckning till Finland. Delar av landskapet hör fortfarande till Finland, och namnen *Pohjois-Karjala* (Norra Karelen) och *Etelä-Karjala* (Södra Karelen) finns kvar som benämningar på områden i Finland. I områdena talas så kallade *karjalaismurteet* ’karelska dialekter’. En adekvatare benämning är *sydostfinska dialekter*. (Torikka m.fl. 2004).

2.5. Betydelsebeskrivningen för allmänna ord i FWN

De begrepp jag sett på i det föregående har gett en rätt negativ bild av FWN, men min angreppspunkt – att undersöka begrepp som är typiska för det finska samhället – har förstås också varit orättvis. Mer allmänna uppslagsord kan förväntas ge ett bättre utslag, och i rättvisans namn har jag därför kastat en blick på betydelsebeskrivningarna för tio slumpmässigt valda ord på *a-*, *h-* och *p-* och jämfört FWN med två andra ordböcker, den enspråkiga finska ordboken Kielitoimiston sanakirja (KT) och den finsk-engelska ordboken Uusi suomi-englanti suursanakirja (USE).

KT och USE (som redigerats oberoende av varandra) visar tämligen stor samstämmighet i fråga om antalet angivna betydelser i de undersökta artiklarna, medan siffrorna för FWN överlag är högre. Vissa (kortare) artiklar stämmer förhållandevis väl överens. En del av betydelseangivelserna i FWN blir upprepningar, eftersom de härrör från olika engelska ord. Ordet *puhdas* 'ren' utgör motsvarighet till de engelska orden *pure*, *clean* och *clear*, och likartade betydelseangivelser återkommer därför flera gånger. En del av angivelserna ligger också långt borta från uppslagsorden betydelsemässigt. Det gäller bland annat nyansen *clear* för *puhdas* (klar luft är inte riktigt detsamma som ren luft). En av motsvarigheterna till verbet *ajaa* är *course* 'hunt with hounds'. Verbet *ajaa* kan visserligen användas med avseende på drivande hundar men inte om jägare som jagar hare eller räv med stövare. Verb som *haiskah-taa* 'lukta' (intransitivt) och *vaikuttaa* 'verka, förefalla' borde inte uppträda som synonymer till *havaita* 'observera' (transitivt).

Överlag verkar betydelsebeskrivningarna för mer allmänna ord mer adekvata i FWN än de för de kulturbundna orden. Men också de artiklarna skulle kräva en genomgång, eftersom FWN bär tydliga spår av att vara en vänd ordbok.

3. Begreppshierarkierna

FWN är ett onomasiologiskt verk och begreppshierarkierna utgör en viktig del av innehållet. Jag ska därför ägna de följande avsnitten åt dem. Det finska ordnätet har kommit till genom ren översättning. Ingen onomasiologisk bearbetning har tills vidare gjorts (våren 2011). Hierarkierna är övertagna från det amerikanska ordnätet med de styrkor och svagheter det uppvisar. I och med att betydelsebeskrivningarna för de genomgångna kulturella och samhällliga orden inte till alla delar är adekvata i FWN kan man vänta sig att också hierarkierna är problematiska.

Det finns flera olika typer av begreppshierarkier i FWN. De typer jag kommer att diskutera är hypernymer, hyponymer och ko-hyponymer (systrar). Exempelen jag visar är hämtade direkt ur FWN, och hierarkierna går från det mest specifika högst upp till det mest generella längst ner. I de fall då inte hela kedjan varit relevant för diskussionen har jag kortat av den. För icke-finskkunniga läsare har jag låtit de engelska motsvarigheterna ur FWN ingå i exemplen.

3.1. Begreppshierarkidjup och särskiljningsnivå

Mycket påfallande när man slår upp kultur- och samhällsbundna ord i FWN är de långa hypernymkedjorna. Begreppshierarkierna är med andra ord djupa.

Ber man FWN om hypernymerna för begreppet ”presidentti” får man fyra hierarkier för de fyra olika betydelsena. För betydelse 1 (USA:s president) ser hierarkin ut som i exempel 2.

TEMATISKE BIDRAG

Yhdysvaltain presidentti [President of the United States, United States President], presidentti [President], ylipäällikkö [Chief Executive]

⇒ valtionpäämies [head of state], valtionjohtaja [chief of state]

⇒ edustaja [representative]

⇒ välittäjä [negotiator], neuvottelija [negotiant], sovittelija [treator]

⇒ viestijä [communicator]

⇒ henkilö [person, individual, someone, somebody], ihminen [mortal, soul]

⇒ eliö [organism], olento [being]

⇒ elävä olio [living thing, animate thing]

⇒ kokonaisuus [whole, unit]

⇒ esine [object], fyysinen esine [physical object]

⇒ fyysinen entiteetti [physical entity]

⇒ entiteetti [entity], kokonaisuus [entity]

⇒ aiheuttaja [causal agent], syy [cause], kausaalinen agentti [causal agency]

⇒ fyysinen entiteetti [physical entity]

⇒ entiteetti [entity], kokonaisuus [entity]

Exempel 2.

Hierarkin för ”presidentti” påminner starkt om de taxonomiska hierarkierna för t.ex. växter och djur, men jag ställer mig frågande till de många nivåerna. Behövs exempelvis hela serien *viestijä–välittäjä–edustaja*, och på vilka grunder har de tre ordens inbördes hierarkiska position definierats?

Men hierarkin i exempel 2 är dessutom svårtolkad. Det förefaller som om det handlade om två alternativa hierarkier, dels en serie med tolv nivåer (sökordsnivån inberäknad), dels en serie med fyra nivåer. Å andra sidan ligger *aiheuttaja* och *eliö* på samma avstånd från vänster och det kan möjligen betyda att bara de tre översta hierarkinivåerna är alternativa. Något omnämnande om alternativa hierarkier har jag inte hittat på FWN:s instruktions-sida. Eftersom *entitet* och *fysisk entitet* återfinns i båda serierna ska trädets ändå knappast tolkas som ett enda hierarkiträd med femton nivåer.

För betydelse 4 ”presidentti” med avseende på presidentinstitutionen i USA ser hierarkin ut som i exempel 3.

Yhdysvaltain presidentti [President of the United States], presidentti [President], ylipäällikkö [Chief Executive]
 ⇒ presidentintoimi [presidency], presidenttikausi [presidency], presidentinvirka [presidentship]
 ⇒ asema [position], toimi [post, billet], paikka [berth, spot, place], virka [office], työpaikka [situation]
 ⇒ elinkeino [occupation], liiketoiminta [business], työ [job], toimiala [line of work], ala [line], kiinnostuksen kohde [line]
 ⇒ aktiviteetti [activity], toiminta [activity]
 ⇒ teko [act, deed], työ [act], toimenpide [act], inhimillinen toiminta [human action, human activity]
 ⇒ tapahtuma [event]
 ⇒ psykologinen ominaisuus [psychological feature]
 ⇒ abstrahointi [abstraction], käsitteellistäminen [abstraction], abstrakti entiteetti [abstract entity]
 ⇒ entiteetti [entity], kokonaisuus [entity]

Exempel 3.

I detta exempel är det påfallande att både statiska (”psykologisk egenskap”, ”position”) och dynamiska begrepp (”händelse”, ”aktion”) ingår i en och samma hierarki som överbegrepp för varandra.

Med avseende på USA:s president får ordet *presidentti* alltså två olika hierarkier, en för den personsyftande användningen och en för den institutionssyftande. Men om man väljer att skilja på person och institution borde samma tudelning i konsekvensens namn göras också för betydelserna ’överhuvud i republik’ och ’ledare för högre rättsinstans’, vilket alltså inte är fallet.

Strängt taget kan skillnaden mellan tjänsteposition och innehavare anses gälla de flesta arbeten, inte bara högre tjänstemän. Om en arbetsgivare i en platsannons söker en lärare, en sekreterare, en städare, en lastbilschaufför, en rörmokare eller liknande så kan benämningarna i annonsen knappast anses syfta på en regel-

rätt person utan snarast på uppgiften personen behövs till.

Frågan är om det inte vore ändamålsenligare att lägga in tudelningen person–position på en högre nivå i begreppshierarkin. Ska skillnaden kodas in borde det i varje fall göras konsekvent för alla ord som den kan aktualiseras för.

I det danska ordnätet DanNet (<http://andreord.dk>) görs ingen skillnad person–institution. Ordet *præsident* anges i betydelseöversikten ha tre betydelser ('domare', 'ledare' och 'statsöverhuvud') och för betydelsen 'statsöverhuvud' är de överordnade begreppen sex till antalet (exempel 4).

statsoverhoved → leder → person → væsen → dyr → organisme

Exempel 4.

Man kan förstås ställa frågan om *president* i satser av typen ”i Finland väljs det president vart sjätte år” passar in i begreppsserien ”person”–”organism”, men å andra sidan är innehavande av olika positioner och tjänsteställningar i samhället så intimt sammankopplat med begreppet ”person” att man kan fråga sig om inte ett ’potentiell samhällsfunktion’ utgör en viktig betydelsekomponent i ordet.

De FWN-egenskaper som här diskuterats är ärvda från den amerikanska förlagan och det finns kanske goda argument för att begreppsserierna ser ut som de gör. För en utomstående betraktare verkar de dock inte särskilt stringenta. Med tanke på en fortsatt bearbetning av FWN borde en diskussion föras om de direkt övertagna hierarkierna är optimala eller om justeringar kunde genomföras. Problemet är förstås att kopplingen till det amerikanska ordnätet försvagas om ändringar görs. Å andra sidan antar jag att kopplingen kan försvagas också vid revideringar av Princeton WordNet.

3.2. Perspektiv och precisionsnivå

Det som är att vänta när ett ordnät skapat för ett språk översätts till ett annat språk är att perspektivet blir fel ur det nya språkets synvinkel.

Avsaknad av finsk infallsvinkel i FWN är påfallande om man ser på hierarkin för begreppet ”karjalainen” ’karelare’, som ju är en folkslagsbeteckning nära förknippad med den finska kultursfären. Hypernymserien visas i exempel 5. Endast de nedersta nivåerna är medtagna.

Sense 1

karjalainen [Karelian, Carelian]

⇒ venäläinen [Russian]

⇒ alkuasukas [native, indigen, aborigine, aboriginal], maakalainen [indigene]

⇒ henkilö [person, individual, someone, somebody], ihminen [mortal, soul]

[...]

Exempel 5.

En karelare betecknas alltså rätt och slätt som ryss, vilket i och för sig är korrekt i fråga om betydelsen ’medborgare i Ryssland’. Men i ett finskt ordnät saknar man en alternativ hierarki för karelarnas tillhörighet till de finsk-ugriska folken. Karelarna och andra finsk-ugriska grupper innehar en särställning i finskt perspektiv jämfört med andra folk i Ryssland.

Anmärkningsvärt i hierarkin i exempel 5 är också att ryssar definieras som ”urinvånare”, medan hypernymkedjan för t.ex. finnar, danskar och tyskar är ”europé–invånare–person” och för t.ex. turkar, indier och israeler ”asiat–invånare–person”. Bakgrunden är möjligen att inte alla ryssar är européer och att ”ryss” å ena sidan borde ingå i två hypernymkedjor, å andra sidan är övergripande för ”europé” och ”asiat”.

TEMATISKE BIDRAG

Systemangivelserna för ”karjalainen” utgår förstås också från ”ryss” som överordnat begrepp, och i raden av ko-hyponymer återfinns därmed beteckningar för olika folkslag i Ryssland, både finsk-ugriska och andra (exempel 6). Att ”karelare” också är systerbegrepp till ”finne” i egenskap av beteckning på finsk-ugriska folk framgår ingenstans.

- venäläinen [Russian]
- ⇒ tataari [Tatar]
- ⇒ udmurtti [Udmurt], votjakki [Votyak]
- ⇒ komi [Komi]
- ⇒ tseremissi [Cheremis, Cheremiss], Mari [Mari]
- ⇒ inkeriläinen [Ingrian, Inger, Ingerman]
- ⇒ karjalainen [Karelian, Carelian]
- ⇒ osjakki [Ostyak], hanti [Khanty]
- ⇒ mordvalainen [Mordva, Mordvin, Mordvinian]
- ⇒ nganasani [Nganasan]
- ⇒ selkupi [Selkup], ostjakkisamojedi [Ostyak-Samoyed]
- ⇒ samojedi [Samoyed]
- ⇒ vepsäläinen [Veps, Vepse, Vepsian]
- ⇒ voguli [Vogul], mansi [Mansi]
- ⇒ jeniseiläinen [Yeniseian]
- ⇒ isovenäläinen [Great Russian]
- ⇒ moskovalainen [Muscovite]
- ⇒ georgialainen [Georgian]
- ⇒ siperialainen [Siberian]
- ⇒ tsetseeni [Chechen]

Exempel 6.

Systrarna till ”suomalainen” ’finne’ är över sextio till antalet och de representerar en brokig skara med ”eurooppalainen” ’europé’ som överbegrepp. En stor, entydig grupp systrar är benämningar på invånare i nu existerande stater som t.ex. *tanskalainen* ’dansk’, *italialainen* ’italienare’ och *valkovenäläinen* ’vitryss’. I tillägg till dessa finns det invånarbetesbeteckningar på nyligen upplösta stater

som *jugoslavialainen* 'jugoslav' och *tsekkoslovakialainen* 'tjeckoslovak'. Däremot är det tveksamt om angler, saxare, franker och kelter för att inte tala om teutoner och langobarder hör hemma under begreppet "europé". Jämförbara med beteckningar för moderna statsmedborgare är de i varje fall inte.

I ko-hyponymlistan för "suomalainen" ingår vidare beteckningar på invånare i områden som inte utgör stater t.ex. *kreetalainen* 'invånare på Kreta', *balkanilainen* 'balkanbo', *saamelainen* 'same', *pohjoismaalainen* 'nordbo' och *tirolilainen* 'tyrolare'. De avser alla i och för sig också moderna européer, men ska den typen av beteckningar anses som systrar till *suomalainen* borde också ett mycket stort antal andra invånarbeteckningar komma med.

Den märkligaste ko-hyponymen till "suomalainen" är beteckningen *sahib*. Visserligen har benämningen förekommit med avseende på européer i Indien, men ordet fungerade som tilltalsord med innebörden 'herre' (SO *sahib*) och inte som en folkslagsbeteckning.

Avsaknaden av underbegrepp för "suomalainen" 'finne; finländare' är också flagrant med tanke på att ett betydande antal amerikanska invånarbeteckningar återfinns i FWN. För det första har vi beteckningar för invånare i de olika landskapen, exempelvis *hämäläinen* 'tavastlänning', *savolainen* 'savolaxare', *pohjalainen* 'österbottning' med flera. En lite annan typ representerar *suomenruotsalainen* 'finlandssvensk'. Hyponymer är också alla inbyggarbeteckningar till stadsnamn: *helsinkiläinen* 'helsingforsare', *tamperealainen* 'tammerforsare', *vaasalainen* 'vasabo'. I och för sig är det knappast påkallat att ta med invånarbeteckningar för alla städer i Finland i ett finskt ordnät, men ser man till frekvens i moderna finska texter är *helsinkiläinen* och *tamperealainen* vida vanligare än exempelvis *permiläinen* 'permjak' som finns med i FWN.

Men också i detta fall finns det en historisk dimension. Benämningen *suomalaiset* avser tillika en av de stammar som har befolkat Finland. I det avseendet är alltså "finne" ett systerbegrepp till t.ex.

”tavast” och ”karelare”, inte hypernym som i fråga om moderna förhållanden. I ett finskt ordnät borde med andra ord en historisk betydelse med tillhörande hierarkier anges separat.

Det sista exemplet på perspektivproblem som jag ska ta upp är ko-hyponymerna för ”presidentti” i betydelsen ’statsöverhuvud’ (exempel 7)

- valtionpäämies [head of state], valtionjohtaja [chief of state]
 - ⇒ pääministeri [chancellor, premier, prime minister]
 - ⇒ pääministeri [Prime Minister, premier], PM [PM]
 - ⇒ presidentti [president]
 - ⇒ Yhdysvaltain presidentti [President of the United States, United States President], presidentti [President], ylipäällikkö [Chief Executive]
 - ⇒ hallitsija [sovereign], kruunupää [crowned head], monarkki [monarch]

Exempel 7.

Ett land har normalt ett och endast ett statsöverhuvud (*valtion päämies, valtionjohtaja*), antingen en president eller en monark beroende på statsskick. Regeringschefen (statsminister, premiärminister, förbundskansler) är inte statshuvud (utom möjligen ställföreträdande) och ”pääministeri” är inte en ko-hyponym till ”presidentti”. I USA är presidenten tillika regeringschef, och premiärminister saknas. I den meningen kan ”president” med syftning på USA:s president alltså faktiskt betraktas som ko-hyponym både till ”monark” och till ”stats- och premiärminister”. Men i exempel 7 handlar det inte explicit om USA:s president.

De anförda exemplen visar inte enbart på ett avlägset perspektiv på finsk kultur och det finländska samhället utan de aktualiserar också ett problem som måste vara generellt för ordnät, nämligen hur noga man ska definiera begrepp. Att exempelvis definiera langobarderna som européer utan någon vidare kategorisering blir ytterst oprecist. Går man in för underkategorisering kan det

dock ha betydande återverkningar på ett stort antal ord. Samtidigt inställer sig frågan vad ett ordnät ska beskriva: ska det i första hand beskriva ett visst språk i ett visst samhälle eller ska det vara en allmän beskrivning som täcker så mycket som möjligt av ”världen” utan att gå in i detaljer.

3.3. Översättningsrelaterade problem

Som jag nämnde i inledningen hör engelsk-finsk översättning inte till mitt kompetensområde, så kvaliteten på själva översättningsarbetet undandrar sig min bedömning. Vissa missar har jag i alla fall stött på i FWN under undersökningens gång, och felaktigheterna kan ha återverkningar på begreppshierarkierna. I exempel 8 är *Baltic-Finnic* översatt med *itämerensuomalainen* och *Fennic, Finnic* med *itämerensuomalainen kieliryhmä* (m.fl.).

Sense 2

suomi [Finnish, Suomi]

⇒ itämerensuomalainen [Baltic-Finnic]

⇒ itämerensuomalainen kieliryhmä [Fennic, Finnic], itämerensuomalaiset kielet [Fennic, Finnic], ei-ugrilainen kieliryhmä [Non-Ugric], ei-ugrilaiset kielet [Non-Ugric]

⇒ suomalais-ugrilainen kieli [Finno-Ugric], suomalais-ugrilainen kielikunta [Finno-Ugric]

⇒ uralinen kielikunta [Uralic], uralinen kieli [Uralic language]

⇒ ural-altalaiset kielet [Ural-Altai]

⇒ kieli (yli) [natural language], kieli [tongue]

[...]

Exempel 8.

Ordet *itämerensuomalainen* kan inte ensamt syfta på språk, utan ordet är adjektiv och ingår som bestämning i språkgruppsbenämningen *itämerensuomalaiset kielet* ’översjöfinska språk’, vilket är den finska motsvarigheten till *Baltic-Finnic*. I plural kan adjektivet

användas substantiviskt, men då avser det folkgrupp, inte språk. *Itämerensuomalaiset kielet* är synonymt med *itämerensuomalainen kieliryhmä* 'den östersjöfinska språkgruppen'. Samma begrepp återkommer alltså på två olika nivåer i exemplet. De engelska benämningarna *Fennic*, *Finnic* hänför sig till den finska grenen av de finsk-ugriska språken, där östersjöfinska språk utgör en av flera grupper.

Också i detta sammanhang kan en inkonsekvens nämnas: i begreppshierarkin för "saame" 'samiska' står överbegreppet "saamelaiskieli" 'samiskt språk' direkt under "uraliska språk". Nivån finsk-ugriska saknas alltså.

3.4. Sammanfattande om begreppshierarkierna

Genomgången av begreppshierarkierna har visat på flera problem med FWN. En del har direkt att göra med hur det finska ordnätet har kommit till. Synseten i det finska ordnätet har skapats genom direkt översättning av de amerikanska synseten utan lexikografisk eller ontologisk bearbetning, och FWN har därmed övertagit de problem som finns i förlagan. En del direkta fel har kommit till på grund av felöversättningar, men den stora svagheten med FWN – i en lexikografs ögon – är att beskrivningen av finska förhållanden framstår som mycket distanserad och rentav oinitierad. Möjligen är ordnäten tänkta att vara så allmänna att de är inbördes kompatibla och att de alltså inte ens ska röra sig på språkspecifik nivå. Men i så fall borde FWN gallra bort många säramerikanska drag.

Det som trätt fram under analysens gång som en viss överraskning är inkonsekvenser i behandlingen av olika begrepp av samma typ. Jag hade förväntat mig att en datorbaserad systematik skulle vara mycket konsekventare. Man får intrycket av att materialet som det amerikanska ordnätet bygger på varit ojämnt. En begreppsordbok förutsätter synnerligen hög grad av systematik i beskrivningarna och en allsidig täckning av hela språkssystemet eller åtminstone av de områden man väljer att ha med.

Antalet hierarkier för ett begrepp är beroende av antalet betydelser man anför. Därför borde betydelsenyanserna bedömas mycket noga, och man måste också ta principiell ställning till på vilken nivå man ska lägga beskrivningen för olika typer av begrepp.

4. Utvecklingsarbete

I Linden & Carlson (2010) påpekas det att den första versionen av FWN ska utvecklas och bearbetas. Ett led i detta är s.k. *crowdsourcing* (jfr Kann 2010:110) där användarna ombeds utvärdera synonymer. I januari 2011 var utvärderingen frivillig och de ord man sökte i FWN kom upp för utvärdering. Användaren ombads också föreslå egna synonymer.

Ur språklig synvinkel är det problematiskt att det är finska ord som ska utvärderas medan den kontext som ges är engelskspråkig. Visserligen kan det försvaras med att utvärderingen principiellt gäller begrepp, inte ord, och att bedömningen därför är språkoberoende. Frågan är dock om alla de användare som råkar komma in på FWN på webben har detta klart för sig. I vissa fall kan kontexten dessutom vara svår att förstå för den som inte har mycket goda kunskaper i engelska. Jämför också Ruppels (2011) ifrågasättande av språkoberoende begrepp.

Någon gång våren 2011 gjordes synonymbedömningen mer obligatorisk, så att man inte kommer in på sökfunktionen i FWN förrän man tagit ställning till ett slumpmässigt synonympar (man kan svara *I'm not sure*). När man sänt sin utvärdering kommer dialogrutan för sökning fram och man får ett nytt ordpar vars synonymigrad man kan bedöma om man vill. Några egna synonymförslags bes man inte längre om³.

I och med att man ska värdera synonymin hos slumpmässigt

³ Den funktionen har integrerats i själva artiklarna: via en länk (dubbelasterisk) kan man lägga till synonymer. Se avsnitt 2.

givna par blir bedömningen ibland svår. Orden man ställs inför kan representera mycket främmande områden. En principiell fråga är också hur väl crowdsourcing lämpar sig som metod för bedömning av termer och termartade ord där graden av synonymi inte kan avgöras via en gallupundersökning utan där det handlar om precisa benämningar som antingen är synonymer eller inte är det. Men regelrätta fel kanske upptäcks genom metoden.

För att bedöma ord som är mer vaga till betydelseinnehållet fungerar metoden uppenbarligen rätt väl (åtminstone enligt t.ex. Kann 2010), och många sådana synonympar kommer naturligtvis också upp i FWN.

Att be användarna utvärdera synonymer är ett led i förbättringsarbetet. Men jämsides med det borde systematiska genomgångar av olika slag utföras av en redaktion.

5. Avslutning

WordNet och dess avläggare på olika språk är främst skapade för språkteknologiska ändamål. Om DanNet sägs det att det ska vara en resurs för IT-system som arbetar med intelligent informationshantering (<http://www.wordnet.dk>). För det amerikanska WordNet anger man en mängd olika tillämpningar (Miller 1998, <http://wordnet.princeton.edu/>), men Miller (1998:xxi) nämner explicit också mänskliga användare. En version som tillhandahålls i ett webbgränssnitt bör väl vara tänkt för mänskliga webbanvändare.

Vilka egenskaper som krävs av resurser för intelligent informationshantering är jag inte klar över, men jag kan inte tro annat än att samma krav bör ställas på dem som på resurser där mänskliga språkhanterare hämtar sin information: de ska vara adekvata beskrivningar av språket i fråga. Förmodligen måste kraven ställas ännu högre, för en dator bearbetar ju i ett nafs sådana mängder av material som en människa inte ens kan drömma om att gå igenom

under en hel livstid. Alla eventuella fel återverkar oändligt mycket mer. Och en dator kan inte tänka själv.

En gedigen och adekvat resurs för finska språket är FWN tills vidare inte. Som jag ser det fungerar ordnätet för närvarande knappast tillfredsställande för något annat än som källa för översättning från (amerikansk) engelska till finska. Men inte ens vid översättning vill jag slå upp i ett finskt ordnät och få begreppet ”grundskola” beskrivet för mig i amerikanska termer, och jag känner mig lite kränkt som finländare om min finska källa anger att den första betydelsen för ordet *presidentti* är USA:s president.

Problemet med att ord som är specifika för ett visst samhälle kan vålla översättningsproblem tas upp av Lindén & Carlson själva (2010:133f.). De avfärdar det dock med att de udda orden utgör bara en liten del och att 95 % av alla ord låter sig översättas. Det motiverar enligt deras åsikt översättning som metod att skapa ett ordnät för ett nytt språk. Det är till detta jag som tvåspråkslexikograf har mina betänkligheter. Jag kan i och för sig gärna tro att 95 % av orden låter sig översättas men jag ställer mig frågan vad översättningsresultatet ska användas till om produkten inte förädlas avsevärt.

Min undersökning har visat att en gammal lexikografisk sanning verkar hålla streck också för ordnät: man gör inte en ordbok från språk 1 till språk 2 genom att vända på en ordbok från språk 2 till språk 1. Det ska mycket mer redigeringsarbete till.

Därmed inte sagt att översättning inte kan vara det första steget på vägen. Som Lindén & Carlson (2010:123) slår fast: att skapa ett finskt ordnät från början hade varit att upprepa årtionden av lexikografiskt arbete som utförts vid Princetons universitet. Det är helt sant. Men varje källa har sina svagheter och översättningsarbetet borde utföras parallellt med eller åtminstone följas av en kritisk analys och framför allt av målgruppsanpassning.

Jag hoppas det ska finnas resurser att arbeta vidare med FWN så att det småningom kan bli det gedigna ordnät för finska som

det är tänkt att vara. Att bygga upp ett ordnät för ett språk är en intrikat och mödosam process oavsett i vilken ända man börjar (jfr Pedersen 2010 om DanNet). Ett ordnät som kommit till genom översättning är en god råvara för förädling. Och förädlingen av FWN kunde förslagsvis ske med hjälp av resultaten av många års mödosamt *finskt* lexikografiskt arbete. På så vis kunde finsk lexikografi och finsk språkteknologi berika varandra.

Litteratur

Ordböcker och uppslagsverk

KT = *Kielitoimiston sanakirja* (Elektronisk version) ('Språkbyråns ordbok'). Eija-Riitta Grönros (huvudredaktör), Minna Haapanen (redaktionssekreterare) m.fl. Helsinki: Kotimaisten kielten tutkimuskeskus ja Kielikone Oy, 2008.

SO = *Svensk ordbok utgiven av Svenska Akademien*. Utarbetad vid redaktionen för Svenska Akademiens samtidsordböcker, Lexikaliska institutet, Institutionen för svenska språket, Göteborgs universitet. Svenska Akademien & Norstedts 2009.

USE = Raija Hurme, Riitta-Leena Malin & Olli Syväoja 1984/1989: *Uusi suomi-englanti suursanakirja* ('Ny finsk-engelsk storordbok'). Helsinki: WSOY.

Annan litteratur

Kann, Viggo 2010: KTHs morfologiska och lexikografiska verktyg och resurser. I: *LexicoNordica* 17, 99–118.

Lindén, Krister & Carlson, Lauri 2010: FinnWordNet – WordNet på finska via översättning. I: *LexicoNordica* 17, 119–140.

Miller, George A. (utg.): 1998: *WordNet. An electronic database*. Cambridge (Mass.)/London: The MIT Press.

- Pedersen, Bolette S. 2010: Semantiske sproressourcer – mellem sprogteknologi og leksikografi. I: *LexicoNordica* 17, 163–180.
- Ruppel, Klaas 2011: Ontologi, begreppssystem och WordNet. I: *LexicoNordica* 18 (i denna volym).
- Torikka, Marja, Raija Miikkulainen & Anneli Hänninen 2004: *Karjala*. Kotimaisten kielten tutkimuskeskuksen julkaisuja 129. Helsinki. <http://kaino.kotus.fi/www/verkkojulkaisut/julk129/> (mars 2011)

Internethänvisningar

- DanNet*. <http://wordnet.dk/dannet/>, sökningar: <http://andreord.dk/> (mars 2011)
- EuroWordNet*. <http://www.illc.uva.nl/EuroWordNet/> (mars 2011)
- FWN = *FinnWordNet*. <http://www.ling.helsinki.fi> (dec. 2010–aug. 2011)
- Miller, George A. 1995: WordNet: A Lexical Database for English. I: *Communications of the ACM* Vol. 38, 39–41. citeseerx.ist.psu.edu/viewdoc/download (mars 2011)
- WordNet*. A lexical database for English. <http://wordnet.princeton.edu/> (mars 2011)

Nina Martola
forskare, fil. dr
Forskningscentralen för de inhemska språken
Berggatan 24
FI-00100 Helsingfors
nina.martola@focis.fi

