Pilke, N. & N. Nissilä (red.) 2016. Tänkta termer. Terminologihänsyn i nordiskt perspektiv.
VAKKI Publications 5. Vasa, (100–136)

Post och bank – om lagring av terminologi, den svenska Rikstermbanken och den finska Vetenskapstermbanken

Henrik Nilsson

Terminologicentrum, TNC

Antti Kanner & Tiina Onikki-Rantajääskö

Helsingfors universitet, Vetenskapstermbanken i Finland

Tässä kaksikielisessä artikkelissa kuvaillaan kahden kansallisen termipankin, ruotsalaisen Rikstermbankenin ja suomalaisen Tieteen termipankin piirissä tapahtuvaa termityötä. Artikkelin alussa keskustellaan kansallisten termipankkien erityispiirteistä sekä niiden koostamiseen ja ylläpitoon liittyvistä haasteista. Esitellyt kaksi kansallista termipankkia poikkeavat toisistaan monin tavoin, muun muassa kokoamistapansa, sisältönsä ja teknisen toteutuksensa puolesta. Ruotsalaisen Rikstermbankenin sisällöstä ja teknisestä toteutuksesta vastaa yksi organisaatio, Terminologicentrum TNC, joka arvioi ja tallentaa kaikki termipankkiin vietävät tietueet. Suomalainen Tieteen termipankki puolestaan on Suomen Akatemian ja Helsingin yliopiston yhtenen infrastruktuurihanke, jonka sisällön tuotantoon on kutsuttu mukaan Suomen koko tiedeyhteisö, ts. sisältö tuotetaan talkootyönä wiki-alustalle. Kansallisten termipankkien olemassaololla on suuri merkitys tiedon saatavuuden näkökulmasta, sillä ne mahdollistavat tehokkaan terminologisen tiedonhaun samanaikaisesti useilta eri aloilta.

Nyckelord: nationell termbank, termbank, termdatabas, termpost, termbase, terminology database, terminological record, Rikstermbanken, Vetenskapstermbanken, Bank of Finnish Terminology in Arts and Sciences, BFT, nichesourcing

1 Inledning

”Resultatet av terminologiarbete lagras bäst i post och bank!” kan man säga – och det är inte bara en putslustig formulering utan faktiskt en helt sann utsaga. Struktur och systematik hör nära ihop med terminologiarbetet, och särskilt viktigt blir det när det färdiga resultatet av terminologiarbetet, terminologiska data av olika slag, ska lagras och presenteras. De minsta beståndsdelarna – termerna, definitionerna, kommentarerna etc. – struktureras enligt upprättade och genomtänkta format i *termposter* som därefter samlas ihop och görs sökbara i *termbanker*. Termbanker kan göras mycket omfattande med komplexa strukturer eller ganska enkla, med få olika typer av data, och med hjälp av eller utan särskilda så kallade *terminologihanteringssystem* (TMS:er). En termbankstyp som

ställer särskilda krav är den *nationella termbanken* – både vad gäller innehåll, teknik och användning. I den här artikeln beskrivs arbetet med den svenska Rikstermbanken och den finska Vetenskapstermbanken (Bank of Finnish Terminology in Arts and Sciences, BFT). Dessa två termbanker är intressanta både för att de är förhållandevis aktuella, men också för att de utvecklats på olika sätt: Rikstermbanken (se avsnitt 3.1) på det mer traditionella sättet med *en* ansvarig institution som matar in terminologiska data och Vetenskapstermbanken (se avsnitt 3.2) på ett sätt där experter genom så kallad *nichesourcing* tillsammans bygger upp termbanken i en wiki-baserad plattform.[18]

2 En databas – en termdatabas – en termbank?

Ett kortregister, som det som började byggas upp på Terminologicentrum (TNC) redan 1941, kan ses som en tidig version av en termbank med olika sökmöjligheter. Men det är kanske först när de datorbaserade hjälpmedlen gör sin entré i ordlistearbetet, i TNCs fall i slutet av 1960-talet med en stordator, som det blir naturligt att tala om databaser, fyllda med terminologiska data som det går att göra sökningar i. Tidigare gjordes en distinktion mellan ’termbank’ och ’termdatabas’, där ’termdatabasen’ var samlingen av ’termposter’ och ’termbanken’ samlingen av ’termdatabaser’:

- ’termdatabas’, ’terminologisk databas’: ”samling termposter som lagras i flera filer och som kan bli åtkomliga för sökning, bearbetning, komplettering och sortering”
- ’termbank’: ”centralt organiserat system av termdatabaser”

I *Terminologins terminologi* (Nordterm, 2008) finns dock enbart *termdatabas* med (med synonymen *terminologisk databas*) med definitionen ”databas som innehåller terminologiska data”. Distinktionen mellan ’termbank’ och ’termdatabas’ verkar alltså ha blivit mindre vanlig, kanske beroende på programvaruutvecklingen på området. Dessutom har ju TMS:erna, dvs. ’terminologihanteringssystemen’ (från eng.

[18] Se vidare Hoffman och Nistrup Madsen i denna volym för mer information om TMS:er och termbanker.

’terminology management systems’) också börjat dyka upp som helhetslösningar för att bygga upp och sköta termbanker. Ibland har även andra skillnader framförts, till exempel att ’termbank’ enbart skulle avse själva samlingen av terminologiska data, utan någon sökfunktion (vilken däremot skulle ingå i termdatabasen).

Bowker (Kockaert & Steurs 2015: 308ff.) gör en distinktion mellan ’term banks’ och ’term bases’ där de förstnämnda sägs vara större och mer heltäckande både vad gäller innehåll och antal element medan de sistnämnda är mer begränsade och oftare med grund i ett ad hoc-arbete med några få begrepp snarare än i ett systematiskt arbete med begreppen inom ett helt fackområde. Med den beskrivningen hamnar också ’term bases’ ganska nära översättningsminnen. Wright och Budin (2001: 878) gör en liknande distinktion, men betonar att ’term banks’ ofta är en institutionell produkt som kräver offentlig finansiering medan ’termbases’ är databaser som läggs upp av enskilda företag och organisationer.

Inom ramen för denna artikel finns inte möjlighet att uttömmande beskriva termbankens hela natur, men några grundläggande principer är viktiga att ha med sig för fortsättningen: all terminologihantering, inklusive termbanksuppbyggnad, bör präglas av struktur och systematik, och generellt kan man säga att det ofta behövs mer struktur än man tror. Dessutom bör terminologihanteringen följa principerna om *begreppsorientering* och *termautonomi* (Schmitz 2005). Begreppsorientering innebär att begreppet ska vara utgångspunkt i arbetet och att varje termpost bara ska innehålla terminologiska data om ett begrepp. Termautonomi innebär att man för varje term som används för ett visst begrepp (och alltså registreras i en och samma termpost) måste kunna upprepa ev. språklig information (böjning, status etc.). På så sätt kan varje term hanteras självständigt i termbanken.

3 Den nationella termbanken

Till att börja med skulle man kunna se på den nationella termbanken som en termbank med ”nationell nytta”, dvs. en termbank med en ”sammantagen samhällsnytta som uppstår genom en infrastrukturinvestering för alla medborgare i hela landet”

(Riksrevisionen 2012: 21f.). Den definitionen stämmer överens med fallet Rikstermbanken, men är kanske ändå något för enkel. Vetenskapstermbanken i Finland kan också sägas vara nationell (Tieteen kansallinen termipankki på finska), och den är med i Vägvisaren för Finlands forskningsinfrastrukturer i samarbete med FIN-CLARIN. Benämningen *nationell* skulle också kunna tolkas på andra sätt, till exempel att

- staten står bakom, tar ansvar och finansierar,
- det finns en länk till ett nationellt terminologicentrum,
- den grundar sig på en ”nationell” begreppsvärld,
- den innehåller vissa språk (enspråkig, endast ”nationalspråk”),
- dess innehåll uppvisar viss kvalitet,
- den är tillgänglig (gratis, anpassad),
- den är omfattande (till exempel täcker all terminologi inom nationen, inget ”utländskt” etc.),
- den har en viss status (vilket påverkar användningen av dess innehåll),
- det är ett bra marknadsföringsepitet.

Vid ett seminarium inriktat på just nationella termbanker, Termintra, som hölls i Oslo 2012, diskuterades kriterierna för ”nationella termbanker” och följande tre kriterier antogs som nödvändiga:

- viss täckning,
- status (erkänd av yrkesverksamma och en språk- eller terminstitution),
- tillgänglighet (öppen och inte föremål för äganderättstänkande).

Att benämningen *nationell* sedan kan vara politiskt känslig berördes också. Unesco har definierat ’national terminology database’ som ”database containing mono- or multilingual terminological data […] established at country level” (Infoterm, 2005: 45). Men det geografiska kännetecknet som anges här räcker inte till för att bestämma begreppet ’nationell termbank’; i Wales har man till exempel medvetet kallat termbanken *nationell*, medan man i till exempel Frankrike inte skulle kunna använda uttrycket

nationell om man förväntar sig många användare. I en genomgång av termbankstyper lägger McNaught (1988: 114) till ytterligare några kännetecken: en nationell termbank ”attempts to serve a general purpose role in coordinating the creation and use of terminologies within a country, and hence is theoretically multifunctional, multilingual and exploited by widely differing kinds of users”. Och just denna vida målgrupp utgör en av många utmaningar vid uppbyggnaden av en nationell termbank, inte minst finansiellt, vilket påpekades redan 1983 av dåvarande TNC-medarbetaren Kjell Åström (1983: 185):

> If linguistic data banks have national coverage and fulfil an important need in any particular country’s linguistic heritage and development it stands to reason that operations should in part be financed as a part of a cultural programme sponsored by the nation at large.

Ser man till det nordiska perspektivet är insikten om behovet av termbanker som medel för tillgängliggörande av resultatet av terminologiarbetet inte ny. Arbetet med det nordiska termpostformatet (NTRF[19]) inom ramen för Nordterm var inte bara en viktig grundplåt i utvecklandet av termbanker utan också en internationell föregångare som blev viktig för arbetet med termpostformat inom ISO som MARTIF, GENETER och numer TBX[20].

TNC gav ut den första kommersiella cd-rom:en i Sverige, men *Orðabanki* på Island kan nog sägas vara den första *nationella* termbanken i Norden när den introducerades. I Finland har TSKs *TEPA* funnits tillgänglig ett längre tag och sedan några år finns *Vetenskapstermbanken* (se avsnitt 3.3). Den senaste regeringen i Finland har även lanserat planer på en nationell portal, där alla terminologiska resurser kan samlas ihop. Den samiska termbanken *Sátni* hör till en av de nyare, och den visar också på hur termbanker kan vara ett viktigt verktyg i kampen mot bland annat domänförlust och för domänåtererövran. Introduktionen av Rikstermbanken 2009 var en milstolpe, och de senaste åren har arbetet med motsvarande termbanker i Danmark och Norge intensifierats. Utredningar i Danmark (bland annat ”Sprog til tiden” 2008) har tryckt på vikten av en nationell, fritt tillgänglig terminologisk resurs, en dansk termbank, *DTB*. Vid Copenhagen

[19] Se vidare www.nordterm.net/filer/NTRF/Ntrf.pdf
[20] Se vidare ISO 30042:2008: Systems to manage terminology, knowledge and content – TermBase eXchange (TBX) och referenserna.

Business School har ett treårigt projekt kring en dansk term- och kunskapsbank avslutats och förutom insamling, automatiserad sammanslagning och kvalitetssäkring av existerande semantiska resurser, planerar man även att utvinna terminologisk information ur texter med hjälp av automatisk extraktion. I Norge sker arbetet med den nationella termbanken i projektet *Termportalen* (inom ramen för CLARINO, ett norskt projekt för forskningsinfrastruktur) och där har man tagit sin utgångspunkt i tidigare termbanker (bland annat termbanker som utvecklats av Norsk termbank och RTT) som görs tillgängliga på ett och samma ställe.

Vidgar man perspektivet och lämnar Norden är det intressant att konstatera att det ofta är mindre språk som varit tidiga med att lagra och publicera i termbanker på internet: *Téarma* i Irland, *Euskalterm* i Baskien, *Cercaterm* i Katalonien och *Maes-T* i Wales. Där har behov funnits av att snabbt få spridning på nya termer och frågan om ”nationell” kan där bli särskilt kontroversiell. Även vad gäller termbanker har Kanada varit föregångsland med *Termium* och senare även *Grand dictionnaire terminologique (GDT)*. Men också andra större länder, som Frankrike (*FranceTerme*), och Turkiet (*Bilim ve Sanat Terimleri Ana Sözlüğü*) har haft stora termbanker en längre tid, och i Litauen har *Terminų bankas* fått en alldeles särskild ställning eftersom den i stor utsträckning är tvingande och styr användning av vissa termer, till exempel i offentlig dokumentation. Det senaste tillskottet bland nationella termbanker är *Struna* i Kroatien som sedan 2015 ingår i en större terminologiportal med andra terminologiska källor.[21]

Här behandlas inte vidare stora termbanker inom den privata sektorn (till exempel *Scania Lexicon* i Sverige) eller inom stora internationella organisationer (till exempel *UNTERM* för FN, *ILOTERM* för ILO, *METEOTERM* för WMO) men där finns förstås också många erfarenheter att hämta. Särskilt viktiga har termbankerna inom EU varit – från de tidigaste via *Eurodicautom*, *TIS* och *Euterpe* till nuvarande *IATE* som numer finns för fullständig nedladdning. Nämnas bör också *EuroTermBank* som bidragit mycket till termbanksutveckling, också genom sin applikationsutveckling och sin dokumentation (EuroTermBank, 2006).

[21] Se vidare ”Terminologen (utgave 2): Terminologi – ansvar og bevissthet”, Språkrådet, Norge

I ett europeiskt perspektiv är det intressant att konstatera att mycket få nationella termbanker (och även regionala och internationella termbanker) använder sig av färdiga programvaror (så kallade *OTS:er, off-the-shelf-produkter* som kan ”tas direkt från hyllan och användas”) utan i de flesta fall har egenhändigt utvecklade programvaror tagits i bruk. Om detta beror på att nationella termbanker har så speciella krav, till exempel beroende på den breda användargruppen, är svårt att svara på. Det kan också handla om att det är enklare att få finansiering för utveckling av en programvara än för import och förvaltning av innehåll. Det saknas inte kriterier och metoder för utvärdering av terminologihanteringssystem, men kanske behöver dessa anpassas till termbankstypen ”nationell termbank”?

3.1 Rikstermbanken

Idén om en gemensam termbank i Sverige är inte ny. När arbetet inleddes med ett kortregister 1941 inleddes egentligen också arbetet med en (primitiv) termbank. Termdok, TNCs första system för datortillämpning i terminologiarbete började användas 1969 och den första versionen på cd-rom utkom 1987. TNC hade då under lång tid deltagit i internationell standardisering om datakategorier, utbytesformat med mera och haft kontakter med Kanada där termbanken Termium hade utvecklats.

TNCs arbete med att skapa en nordisk gemensam termbank inom ramen för EU-projektet Nordterm-Net samt en flerårig (2000–2002) medverkan i utvecklingen av IATE, EU:s institutionsgemensamma termbank med tillhörande programvara gav ytterligare värdefulla erfarenheter inför arbetet med Rikstermbanken. Rikstermbanken blev också en av grundpelarna i det genomarbetade förslag till en svensk terminologisk infrastruktur (TISS) som TNC presenterade 2003. Särskilt viktiga blev de förslag som 2005 lades fram i IT-propositionen ”Från IT-politik för samhället till politik för IT-samhället” (2005: 206):

> Termdatabaser bör samlas i en rikstermbank. Uppbyggnaden av en central termbank, en rikstermbank, är väsentlig för tillgången till och kvalitetssäkring av svenska termer. En bra utformad och kvalitetssäkrad termbank blir på sikt ett effektivt verktyg för fackexperter, medier, företag,

> myndigheter, universitet och högskolor och översättare, som snabbt kan behöva få tillgång till aktuell svensk terminologi, samt för utvecklingen av nya IT-tjänster.

och i ”Bästa språket – en samlad svensk språkpolitik” (2005:30):

> En rikstermbank behövs för att säkra tillgång till termer av hög kvalitet. Uppbyggnaden av en nationell central termbank, en **rikstermbank**, är en förutsättning för en smidig tillgång till och kvalitetssäkring av svenska termer inom olika verksamhetsområden.

Näringsdepartementet beslutade sedan om Rikstermbanken under 2006. Dåvarande infrastrukturministern invigde den 19 mars 2009 Rikstermbanken som då innehöll drygt 50 000 termposter och cirka 250 000 termer, till största delen på svenska, men också på några av Sveriges minoritetsspråk samt på andra språk. Ett mångårigt arbete hade äntligen nått sitt mål: en nationell termbank fritt och gratis tillgänglig på nätet. Vägen dit var dock lång och innehöll flera beslut om både innehåll och format.

3.1.1 Rikstermbanken: innehåll och teknikval

3.1.1.1 Innehåll

Eftersom det inte fanns något färdigt recept för hur man bygger upp en nationell termbank diskuterades utgångspunkten och föresatserna grundligt. Utgångspunkten blev att innehållet i Rikstermbanken ska beskriva den svenska begreppsvärlden. Men detta skulle dock inte innebära att termbanken enbart skulle innehålla svenska termer. Ett krav som ställdes upp var att varje termpost ska innehålla antingen en term på svenska eller en term på något av Sveriges minoritetsspråk (finska, jiddisch, meänkieli, romska och samiska). Dessutom skulle materialet uppfylla ytterligare några kvalitetskriterier: det skulle vara informativt, bearbetbart, tillförlitligt och gärna aktuellt och unikt – även om de två sistnämnda kriterierna inte skulle utesluta äldre, intressant material eller att flera ordlistor från samma fackområde togs med. Dessutom skulle källor med någon form av begreppsbeskrivning prioriteras; i dag innehåller ca 75 % av Rikstermbankens termposter någon form av begreppsbeskrivning vilket är en hög siffra vid internationella jämförelser. Konsekvensen av dessa val blev att följande typer av material inte har tagits med:

- allmänspråkligt material
- material som inte innehåller svenska eller minst ett minoritetsspråk
- ofullständigt material
- terminologiskt ostrukturerat material
- svårbearbetbart material
- politiskt inkorrekt och stötande material
- upphovsrättsskyddat material

Tidigt i arbetet fick Rikstermbanken sin egen devis: ”Alla termer på ett ställe” vilket avspeglar ambitionen att göra så mycket terminologi som möjligt, från alla tänkbara domäner, tillgänglig. Samtidigt fanns insikten om att det handlar om något som aldrig blir ”färdigt”, vilket måste förklaras för användarna. En tidig, pragmatisk insikt, som också var ett resultat av tidsramarna, var att utgå från existerande terminologisamlingar och fokus kom att hamna på tre huvudtyper:

1. Terminologiska ordlistor från TNC och TSK i Finland.
2. Terminologier från olika myndigheter och organisationer i Sverige.
3. Ett urval termer och definitioner ur lagtexter, föreskrifter och Statens offentliga utredningar (SOU).

Att börja med egenproducerade ordlistor erbjuder många fördelar, framför allt vad gäller upphovsrätt och kvalitetskontroll. TNCs och TSKs ordlistor[22] är resultatet av ett mångårigt samarbete mellan experter inom olika fackområden och terminologer, som tillsammans använt beprövade metoder och principer i sitt arbete (se Buchers och Seppäläs artiklar i denna bok). För den andra kategorin planerades ett ambitiöst bedömnings- och betygssystem som senare kom att visa sig vara väl ambitiöst varför det övergavs. Dock kvarstår kriterier för bedömning av externt material och en metodik för strukturering och bearbetning inför import. Majoriteten av innehållet utgörs alltså av sådant som tidigare publicerats antingen i bokform eller på webbplatser, eller i vissa fall som del av andra dokument.

[22] Se vidare www.tnc.se och www.tsk.fi.

3.1.1.2 Teknik

Förutom innehållet var förstås teknikvalet viktigt i inledningsskedet – vilken programvara skulle användas för att lagra terminologiska data på nationell nivå? Och vilka krav skulle ställas på en sådan programvara? TNC medverkade i början av 2000-talet i utvecklingen av EU-institutionernas programvara IATE (lanserad 2007), och därför låg det nära till hands att överväga om denna programvara kunde användas. Det visade sig dock finnas upphovsrättsliga problem; dessutom skulle IATE-programvaran kräva stora drift- och underhållsresurser. Men erfarenheterna från IATE-arbetet blev användbara i utvecklandet av en egen programvara med hjälp av öppna program och med inspiration från just IATE och från andra termbanker som till exempel Eurotermbank, finska TEPA och franska FranceTerme. Datamodellen som användes för IATE motsvarar i stort den som finns i standarden ISO 16642: *Terminological Markup Framework* med en termpost (på begreppsnivå) som innehåller språklådor (på språknivå) som i sin tur innehåller termlådor (på termnivå). Arbetet blev att mappa TNCs existerande fältstruktur (enligt NTRF) med de datakategorier som används inom ISO (företrädesvis i formatet TBX). Arbetet har skett agilt och när behovet av ett nytt fält uppstått har det införts efter konsekvensbedömningar. Det innebar att alla fält placerades in i en nivåhierarki, att det beslöts vilka fält som får kopplas till vilka och vad de skulle få innehålla etc. Dessutom utvecklades ett metadataformat för varje källa: bibliografiska data om källan (titel, utgivningsår, utgivare med mera), beskrivande data om källan (bakgrund, viktiga upplysningar med mera) och tolkningsspecifikationer som skulle gälla för en hel källa.

Ett viktigt tekniskt vägval gällde huruvida ”alla ägg skulle läggas i samma korg” eller ej, dvs. att verkligen ha alla termer på ett ställe rent fysiskt. Fördelarna med att samla allt terminologiskt material i en databas på det sättet är att det ger kontroll (över innehåll) och underlättar kvalitetssäkring, konsekvens och likformighet (vilket förenklar jämförelser). Det blir också enklare att harmonisera innehållet. Samtidigt är det arbetskrävande och medför en dubbellagring av material vilket i sin tur medför uppdateringsproblem och en hel del administration (påminnelser om uppdateringar etc.). I fall där det redan finns flera stora termbanker skulle det vara möjligt att ha en distribuerad databaslösning, där

sökningen kan ske genom ett gemensamt gränssnitt, men där alla terminologiska data ligger på ett ställe, hos leverantören. Rent praktiskt kan det förhålla sig så, vilket var fallet i Sverige kring 2009, att det inte finns så många andra (välstrukturerade) termbanker, och den valda lösningen med en central termbank minskade då de (tekniska) krav på leverantörer som annars hade behövt ställas.

En särskild teknisk utmaning som också påverkar användarna är användningen av permanenta länkar, dvs. att varje termpost har en unik identitet. I nuläget ändras denna identitet när ordlistan som termposten ingår i importeras igen, vilket gör att ev. länkar som gjorts till termposten i uppsatser, på webbplatser etc. inte längre kommer att gälla. Det är därför en stor utmaning att införa permanenta länkar utan att behöva ändra hela arbetssättet med import av material.

3.1.2 Arbetet med Rikstermbanken i 10 steg

Organisatoriskt skapades inom TNC ett Rikstermbankssekretariat bestående av tre personer som kunde ägna tid åt innehållet och tekniken och därtill hörande uppsökande, underhållande och vidareutvecklande verksamhet. Att hitta lämpligt material för en nationell termbank är en stor utmaning, en annan är att importera det – särskilt som ”färdiga recept” eller välutvecklade teorier och metoder saknades. Arbetet med att samla och importera material i Rikstermbanken kan beskrivas som en process i tio steg där varje steg innebär olika utmaningar och möjligheter till automatisering:

1. inventering och förbedömning av publicerat terminologiskt material
2. officiell förfrågan till materialägaren om import i Rikstermbanken
3. mottagande av materialet
4. formatering av materialet
5. granskning av materialet
6. återkoppling till leverantören
7. testimport av material
8. eventuella justeringar av materialet
9. slutlig import av materialet
10. uppdatering av materialet

Den tidiga bedömnings- och betygssystemet kom som sagt inte till användning, men metoder att kontinuerligt avsöka vad som publicerats liksom kriterier för bedömning av innehållet utvecklades under steg 1. Det tidiga skedets diskussioner kring upphovsrättsliga frågor relaterade till terminologi ledde till ett arbetssätt i steg 2 där alla tänkbara leverantörer skulle få en förfrågan, även om detta skulle innebära en ganska stor administrationsbörda med påminnelser och övertygande argument. En tidig idé om ett juridiskt avtal genomfördes inte heller eftersom det inte ansågs behövas – och eftersom förfrågan i de allra flesta fall fick ett jakande svar utan förbehåll. Visst material, till exempel termer och definitioner ur lagar och offentliga utredningar, ansågs tillhöra kategorin offentlig information som får vidareutnyttjas.

För steg 3 visade det sig i de allra flesta fall inte finnas något elektroniskt material att leverera; i enstaka fall skickades en word-, pdf- eller excelfil och i undantagsfall något mer avancerat (xml-, rdf- eller databasfiler). Inhämtandet kom alltså att i stor utsträckning bestå av en klipp och klistra-metodik varefter materialet i steg 4, med halvautomatiska hjälpmedel i form av makron och dylikt formaterades till NTRF-format.

Rikstermbanken planerades inte att från början bli ett fullödigt terminologi-hanteringssystem där redigering av termposter skulle kunna ske genom ett webbgränssnitt utan i stället lagrades varje källa som en egen textfil som därefter kunde importeras i databasen. Granskningen i steg 5 genomfördes i huvudsak manuellt enligt traditionella kriterier, men utvecklandet av ett importkontrollverktyg hjälpte också till att upptäcka formella brister i materialet. I Sverige, som i många andra länder, har post- och bankväsendet ändrats betydligt på senare år. Även om förändringen inte varit lika drastisk inom terminologiarbetet så har programvaror för uppbyggnad av termbanker blivit både enklare och billigare. Trots detta används ofta(re) program som inte är särskilt utvecklade för just terminologilagring (till exempel ordbehandlings- och kalkylprogram), vilket i sig inte behöver vara dåligt. Dock kan man konstatera att många ordlistor fortfarande uppvisar brister i struktur; särskilt ofta finner man en klassisk lagringsstruktur som minner om tryckta ordlistor och då särskilt lexikografiskt utarbetade sådana vilket inte är lämpligt för terminologiska data. Sällan är de ordentligt anpassade till digitala sammanhang och deras möjligheter. Lösningen för att hantera konsekvenserna av att många terminologiska

ordlistor fortfarande utarbetas med lexikografisk metod blev ofta ”särslagning”: i ordlistor där alla betydelser av en term presenteras tillsammans under en uppslagsterm fick dessa slås isär till olika separata termposter. I många fall krävdes också ”sammanslagning”, till exempel när synonymer presenterats på olika ställen i en alfabetiskt ordnad ordlista eller när språkversioner publicerats i skilda filer. Sammantaget ser alltså inte någon ordlista ut som den gjorde i original efter TNCs förädlingsarbete, som också innebar vissa ändringar, uppdateringar och strykningar. Dessutom genomfördes skrivregelsmässiga justeringar genomgående på allt material enligt TNCs skrivregler (eller enligt andra språks konventioner till exempel vad gäller nyare tysk rättstavning och translitterering från ryska).

Den stora utmaningen handlade dock om kvaliteten på materialet, och ambitionen har varit att allt som läggs in i Rikstermbanken ska ha gåtts igenom av två terminologer. Oftast har det handlat om att i efterhand försöka tolka det sätt på vilket en ordlista utarbetats och hur den ska läsas. I en stor medicinsk ordlista användes till exempel uttrycket ”synonymt med” genomgående. Enbart i vissa fall avsågs dock relationen synonymi mellan termerna. Lika gärna kunde det handla om variantformer eller till och med om termer för helt andra begrepp!

När det gäller kvalitet och material från lagar och föreskrifter tvekade vi ett tag innan dessa infördes eftersom det ofta handlar om mycket insnävade definitioner som bara gäller i just en viss lag eller föreskrift (”Med X avses i detta dokument …”). Definitioner i författningar har dessutom ofta högt anseende, trots att de terminologiskt kan lämna en del övrigt att önska. Kriteriet för inkludering i Rikstermbanken blev att de skulle klara av att stå på egna ben, dvs. inte vara för knutna till sina respektive dokument och uppfylla en viss terminologisk definitionskvalitet.

Steg 6 blev särskilt viktigt eftersom det gav en möjlighet att rapportera tillbaka de eventuella brister som uppdagats i materialet till varje leverantör, vilket både kunde göra att den terminologiska kvaliteten höjdes och att den terminologiska medvetenheten ökade hos leverantörerna. Som tidigare nämnts finns ett importkontrollverktyg som reagerar på felaktiga kodningar, förekomst av identiska definitioner, felaktiga korsreferenser mellan

termposter, otillåtna värden för vissa fält etc. vilket naturligtvis har varit till stor hjälp för att förbättra kvaliteten på det terminologiska materialet. Som för många delar av Rikstermbanken har utvecklingen varit agil och nya kontrollmekanismer har införts vid behov. Efter testimporten i steg 7 och justeringar enligt de fel (som *måste* åtgärdas) och varningar (som *bör* åtgärdas) i steg 8 kan materialet slutligen importeras i steg 9.

En särskild utmaning med termbanker är förstås att se till att materialet är uppdaterat (steg 10), och särskilt gäller det i fall, som Rikstermbankens, där materialet är dubbelpublicerat (både hos leverantören och i Rikstermbanken). Mycket få leverantörer har automatiskt rapporterat in sina uppdateringar regelbundet vilket har medfört att sekretariatet kontinuerligt måste påminna dem om att göra det samtidigt som vi också själva måste kontrollera och ibland jämföra olika versioner av en och samma ordlista, vilket är tidskrävande. Ett särskilt problem utgör uppdateringar som medför terminologiska försämringar av innehållet, till exempel när en definition av ett begrepp i en ordlista ersatts av en sämre eller kanske till och med av två snarlika definitioner!

3.1.3 Rikstermbankens användare

Att Rikstermbanken skulle finnas gratis tillgänglig för alla på webben var ju en ny situation för en termbank: det var inte längre bara vi själva eller särskilt intresserade och terminologiskt kunniga som skulle använda den och tolka terminologiska data utan målgruppen är betydligt vidare än för en traditionell termbank. Mycket tid ägnades därför åt utarbetandet av Rikstermbankens gränssnitt – det måste vara mycket enkelt att förstå. Det resulterade i ett gränssnitt med en mycket grundläggande enkel sökning (med sökning bland termer och i alla andra textfält, så kallad fritextsökning) och en utökad sökning med några fler valmöjligheter för användarna. Hur träfflistan skulle presenteras var en annan stötesten. Traditionell visning i form av en tabell kändes omodernt, svårläst och platsineffektivt. I stället valde vi en presentation med några raders sammanfattande visning av det viktigaste om varje träff – ungefär som i sökmotorer som Google. Definition eller förklaring visas alltid (eventuellt avkortad om den är lång) men termer på andra språk visas inte. Tanken med detta är att användaren snabbt ska hitta rätt termpost, och sedan lockas att titta på hela termposten, för att få all information. Den sökta texten

framhävs i träfflistan (med mörkare bakgrund) så att det är lätt att se var man fick träff i termposten. Vad som presenteras om varje termpost i träfflistan styrs av var matchningen skedde. Först i träfflistvisningen står alltid den svenska termen, men om det blev träff på en term på ett annat språk visas den termen också. Sorteringen av träfflistan är noga genomtänkt: termposterna i träfflistan sorteras efter bland annat följande kriterier och nivåer:

1. efter matchande språk – svenska först, sedan engelska, tyska, franska osv.
2. efter plats – alla träffar där termen matchar först, därefter träffar i andra fält
3. efter matchningens exakthet – träff på hela termen först, därefter träff på ett ord i en flerordsterm
4. efter källans ålder – nyaste källor först

Effekten av detta blir att alla träffar där träffen är på svenska kommer först, därefter de på engelska osv. Inom alla sådana delmängder (till exempel träffar på svenska) kommer först alla träffar med matchning i termen, därefter de med matchning i andra fält. Så fortsätter det sedan på de olika nivåerna.

TNCs egna undersökningar av användning och statistik visar att den enkla sökningen används i de allra flesta fall och att användarna verkar vara nöjda med den. Det som efterfrågats är sökning per källa, vilket medvetet inte implementerades från början. Förutom en viss rädsla för kopiering av hela ordlistor om den möjligheten gavs handlade det även om typerna av material: eftersom definitioner ur lagar och föreskrifter fanns med kunde man tänka sig att användare skulle välja dessa i första hand och därmed inte se andra, och kanske lämpligare, definitioner. Dessa farhågor har inte infriats och det är troligt att en sådan sökmöjlighet kommer att införas; särskilt intressant blir det eftersom Rikstermbanken i nuläget saknar ämnesklassifikation och en sökning utifrån källtitel ger åtminstone någon vägledning till vilket fackområde det rör sig om. Man kan också konstatera att den huvudsakliga användningen sker på vardagar under arbetstid och också i ganska stor utsträckning av användare utanför Sveriges gränser. Byggområdet är efter flera år fortfarande det område där flest termer och definitioner eftersöks. Referenser till

Rikstermbanken börjar också dyka upp på fler och fler ställen, bland annat i Wikipedia, där även diskussioner om definitioner och termer i Rikstermbanken förs (se bild 1).

Referenser [redigera | redigera wikitext]

1. ^ [a b c d e f] Price, Monica; Walsh, Kevin (2005). *Bergarter och mineral*. Bonniers Naturguider. Översatt av Erik Jonsson. Stockholm: Albert Bonniers Förlag. ISBN 91-0-010458-2 Originalets titel: *Rocks and Minerals* (Doring Kindersley Ltd).
2. ^ [a b] Liten Geologisk Encyklopedi /s.38,68/- utgiven av: Sveriges lantbruksuniversitet, Institutionen för mark och miljö. kan hittas på internetadressen(27feb2013): http://www.geonord.org/ugs/LitenGeologiskEncyklopedOkt2011Rev2.pdf
3. ^ [a b] internetsida: http://www.rikstermbanken.se/rtb/visaTermpost.html?id=199775 & http://www.rikstermbanken.se/rtb/visaTermpost.html?id=166275 & http://www.rikstermbanken.se/rtb/visaTermpost.html?id=191469 - Rikstermbanken, Sveriges nationella termbank, huvudinternetsida: http://www.rikstermbanken.se , läst datum:27 feb 2014 - Rikstermbanken är förvaltad av Terminologicentrum TNC , sveriges nationella centrum för terminologi och fackspråk: huvudinternetsida: http://www.tnc.se
4. ^ [a b c] http://www.ne.se/serpentin & http://www.ne.se/serpentinserien & http://www.ne.se/krysotil & http://www.ne.se/antigorit - från Nationalencyklopedin på nätet - http://www.ne.se - läst datum: 27 feb 2014

Bild 1. Exempel på när Rikstermbanken används som referens i Wikipedia

Trots möjligheten att kommentera varje termpost är det få som kommenterar, och de få kommentarer som lämnas är oftast inte av renodlat terminologisk karaktär.

Även om det redan från början betonades att Rikstermbanken byggts upp, och förvaltas, av TNC har Rikstermbanken med tiden fått en egen identitet (vilket naturligtvis kan bero på det faktum att den ju har en egen webbplats och att brev som rör Rikstermbanken signeras med Rikstermbankssekretariatet). Marknadsföringsmässigt har denna utveckling fördelar – en viktig aspekt när det gäller att skapa en nationell terminologiresurs – men en annan konsekvens har blivit att många ser Rikstermbanken som en autonom helhet, dvs. nästan som en egen organism med *ett* innehåll snarare än som en samling av mer eller mindre disparata källor som presenteras i samma form och med samma utseende på ett ställe. Och med denna status verkar också följa att Rikstermbanken (dvs. TNC) får ett ansvar för *allt* det material som termbanken innehåller. Rikstermbankens karaktär tolkas alltså mer som normativ än deskriptiv – och frågan är om, när och hur den ska övergå till att bli det.

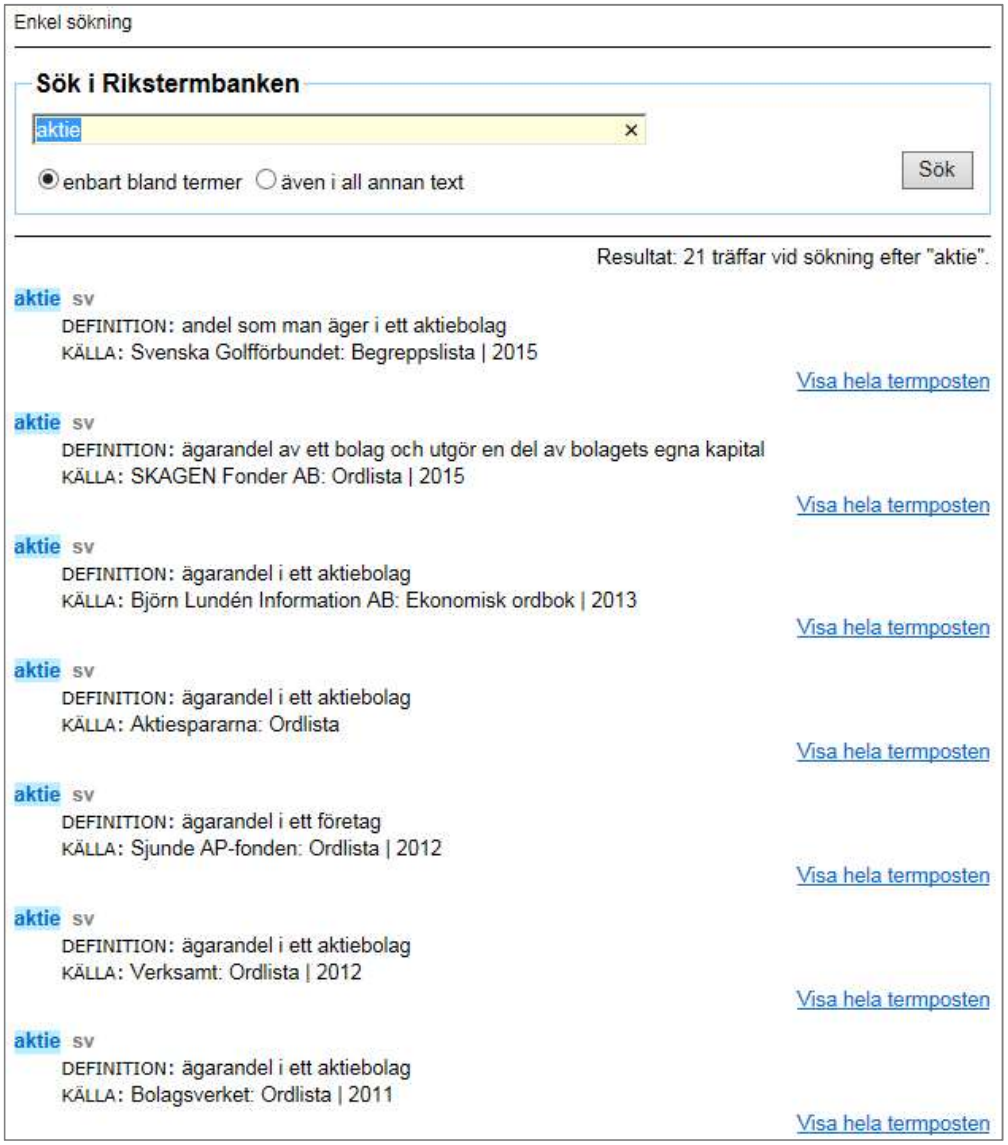

Bild 2. Exempel på flera definitioner av samma begrepp i Rikstermbankens träfflista

Ett av de tänkbara utmärkande kännetecknen för en nationell termbank är just att innehållet har en viss status, till exempel att det måste användas som i fallet med Litauens nationella termbank *Terminų bankas*. I motsats till detta har Rikstermbanken snarast varit deskriptiv till sin natur, dvs. i den presenteras vilken terminologi som finns inom olika fackområden. Det medför att flera ordlistor från ett och samma fackområde ingår och alltså kan det finnas till exempel olika definitioner för samma begrepp (se begreppet ’aktie’ i bild 2).

Vill man att en nationell termbank ska vara normativ och alltså ge en definition av varje begrepp kan detta uppfattas kontroversiellt. Samtidigt får användaren på det sättet möjlighet att jämföra och välja den definition som passar bäst i sammanhanget. När man samlar in så mycket material från olika källor i en och samma termbank är det ofrånkomligt att det kommer att finnas vad som skulle kunna kallas dubbletter. Ofta handlar det dock inte om identiska kopior (även om man ibland kan ana att det lånats från olika håll utan att detta uttrycks explicit i källan), utan om en skala från identiska termposter till termposter där endast termen är identisk. Att harmonisera längs hela denna skala är en utmaning och kan handla om att presentera de identiska i en gemensam termpost med flera källor till att verkligen jämföra kännetecken och göra begreppsanalys. Kanske måste man också i slutändan tillåta att olika definitioner, som lyfter fram olika aspekter av ett och samma begrepp, kan förekomma sida vid sida i en nationell termbank.

Vetenskapstermbanken i Finland har från början skapat sin egen identitet för att framhäva att det gäller forskarna i hela landet, och samarbete med Fin-Clarin betonar det också. Vetenskapstermbanken är snarare deskriptiv, men det är möjligt att den antogs vara mer preskriptiv än var den egentligen är. Expertgrupperna har frihet att bestämma om de vill rekommendera någon termvariant. Implicit är regeln att den förstnämnda varianten på begreppssidan är den lämpligaste, men det är också möjligt att välja rekommendation i rullgardinsmenyn. Andan på wiki-webbplatser är att de alltid är under utveckling, men det är möjligt att informationen på dem tolkas som att den är färdig och permanent. Forskarsamfundet har ett stort ansvar för att uppdatera informationen, men det är inte säkert att de vill ta det ansvaret i praktiken.

3.1.4 Rikstermbanken i framtiden

En nationell termbank är aldrig färdig, den kan bara fortsätta att utvecklas. Det är en bra bit kvar tills ”all” terminologi finns med i Rikstermbanken. Därför är ständig påfyllning med nytt material en viktig del i den framtida utvecklingen – och en viktig fråga för framtiden rör mängden material och varifrån detta material ska hämtas. Det finns till exempel mycket terminologi i svensk standard som skulle berika Rikstermbanken. Påfyllning med viss typ av material begränsas av upphovsrättsliga och marknadsmässiga

faktorer, men det är möjligt att öppna för kommersiella lösningar med till exempel abonnemang och inloggning som skulle ge tillgång till mer material. Samtidigt är det viktigt att Rikstermbanken fortsätter att vara en fri och öppen resurs. Parallellt med påfyllning bör det nuvarande materialet i Rikstermbanken också konsolideras och harmoniseras, till exempel genom att tematiska nätverk för samordning och konsolidering av terminologi inom olika fackområden bildas. Detta är en viktig insats för kvalitetsutvecklingen av Rikstermbankens innehåll.

En intressant utökningsmöjlighet finns i svenskspråkigt material från Finland. Även om Rikstermbanken har den svenska begreppsvärlden som utgångspunkt finns säkerligen en hel del material som inte är uppenbart knutet just till det finska samhället och som därför skulle kunna publiceras även i Rikstermbanken.[23] Med utökade resurser skulle det förmodligen gå att ”hinna ifatt” ordlisteutvecklingen, dvs. inläggningen av material skulle ske fortare än nyproduktionen av ordlistor. Därför kan man tänka sig, som man gjort i arbetet med Dansk termbank i Danmark, att också excerpera terminologisk information ur textmassor och korpusar, mer eller mindre automatiskt. I viss mån innehåller dagens Rikstermbanken sådant material, men det stammar i ringa utsträckning från några automatiska excerperingar. Ska den normativa karaktären öka och mängden likartade termposter minska kommer också den totala mängden termposter att minska på sikt vilket måste hanteras både i marknadsföringen och i förhållande till användarna och deras önskemål.

Tekniskt finns också mycket att utveckla. En nationell termbank, med ett genomarbetat och kvalitetssäkrat terminologiskt innehåll, bör kunna utvecklas för att bli navet i en större semantisk resurs. TNC har föreslagit att Rikstermbanken vidareutvecklas så att termbanken kan ge ett ännu bättre stöd till myndigheterna i utvecklingen av den svenska e-förvaltningen. Informationsförmedling via myndigheters webbplatser, tillgänglighet till e-tjänster med mera förutsätter nämligen ett systematiskt ordnat innehåll som har en begreppslig och terminologisk precision.

[23] Se till exempel exemplet om den gemensamma valutan i Strandvik, Nissilä och Pilke i denna volym.

En väl fungerande e-förvaltning ställer stora krav på samverkan, samordning och samarbete mellan både människor och datorer. Detta i sin tur ställer krav på en effektiv fackspråklig kommunikation (mellan människor) och interoperabilitet (mellan datasystem). Interoperabilitet enligt European Interoperability Framework (EIF) bör ses ur tre olika perspektiv: teknisk, semantisk och organisatorisk. Även om alla tre anses vara lika viktiga dominerar fortfarande de rent tekniska aspekterna, medan den semantiska interoperabiliteten, som troligen utgör en mycket större utmaning, oftast hamnar i skymundan. För att Rikstermbanken ska kunna göra största möjliga nytta i e-förvaltningen och andra IT-satsningar behöver den vidareutvecklas och bli mer känd.

En sammanhållen nationell semantisk resurs kräver att de olika ingående resurserna sammankopplas och görs samtidigt sökbara. De fördelar som uppnås därigenom är att de olika resurserna blir simultant jämförbara. De bör också vara nedladdningsbara så att de direkt kan tillämpas i praktiska applikationer och på sikt även göras tillgängliga som länkade öppna data (LOD). De kan också bidra till utveckling av nya tjänster och utgöra underlag för forskning. Den gemensamma webbplatsen kan därutöver innehålla goda exempel på hur semantiska problem lösts (till exempel ”best practice” och ”business cases”), metodbeskrivningar, standardbaserade format, mallar med mera.

Sökmöjligheterna i Rikstermbanken kan utökas genom nya funktioner, bland annat införande av lagringsbara sökprofiler samt nya sökkriterier, till exempel samtidig sökning på källa och ämnesområde. Åtkomsten behöver förbättras genom anpassning till nya medier (appar till mobiltelefoner, surfplattor etc.). Likaså måste möjligheten att referera till Rikstermbankens innehåll förenklas, vilket förutsätter permanenta länkar.

En annan funktion med en potentiellt stor betydelse för den framtida e-förvaltningen är det som man skulle kunna kalla ”term tracking”, dvs. att man för en viss term vid ett visst tillfälle får fram en sammanställning över de dokument i vilka termen används. En viktig grundprincip för att detta ska kunna åstadkommas är att varje termpost i Rikstermbanken förses med en generell resursidentifierare (URI) som följer med när termen ”lämnar” Rikstermbanken. För att rätt utnyttja denna möjlighet måste författaren av ett dokument ”hämta” termerna hon eller han använder när hon eller han skriver, ur Rikstermbanken,

och integrera dem i sitt ordbehandlingsprogram. Med ett ”term tracking-system” skulle exempelvis föreslagna ändringar av en viss term kunna föregås av en konsekvensbedömning (hur många dokument som berörs av den termförändringen och vilka kostnader det medför etc.). Detta kan vara intressant för lagstiftare som i dag inte enkelt kan överblicka vilka konsekvenser det får för informationsspridningen när man ändrar på en term i en lagtext (ett exempel är den oklarhet som uppstod när ”förtidspension” ändrades till ”sjuk- och aktivitetsersättning”).

Om Rikstermbanken i nuläget oftast används enbart för sökning, bör utvecklingen gå mot en mer interaktiv användning. Interaktivitet vad gäller digitala resurser har ökat i fokus. En mer interaktiv användning av Rikstermbankens material skulle kunna innebära att materialet återanvänds och tillgängliggörs i andra applikationer, för forskning etc. En förutsättning för detta är nya överenskommelser med leverantörerna (som i nuläget enbart godkänt publicering i nuvarande Rikstermbanken). En alternativ lösning skulle snarare vara att materialet inte kopieras och lagras på egen dator utan i stället hämtas in vid visningstillfället och då visas i nya kombinationer utifrån egna önskemål på till exempel leverantörernas webbplatser. Vidare kan interaktiviteten uppmuntras genom att det ges möjlighet till uppladdning av egna terminologisamlingar som kan uppdateras och korrigeras direkt i Rikstermbanken, ett smidigt alternativ till interna system som ibland är oflexibla. Rikstermbanken skulle också kunna bli ett renodlat terminologi-hanteringssystem där leverantörerna själva skulle kunna hantera sin terminologi (även om ett sådant skifte skulle kräva en stor utbildningsinsats för att terminologihanteringen skulle bli smidig).

3.2 The Bank of Finnish Terminology in Arts and Sciences (BFT)

The Bank of Finnish Terminology in Arts and Sciences (BFT) is an infrastructure project funded by the Academy of Finland and the University of Helsinki. The project has two full-time employees – a coordinator specialized in wiki programming (one of the writers of this article[24]) and a post doc researcher specialized in terminology – and a pool of part

[24] Writer 2 until 3/2015

time research assistants. Occasionally, other post doc researchers have been temporarily employed to activate a new field. The project is led by one of the writers of this article and is coordinated at the department of Finnish, Finno-Ugrian and Scandinavian Studies, University of Helsinki. The main cooperative partner is the Finnish Society for Science and Letters and its member societies. The BFT has been accepted to the national road map of infrastructures as a cooperative partner of FIN-CLARIN for the years 2014–2017.

The aim of the BFT is to build a permanent and continuously updated terminological database for all fields of research in Finland. To this end, the project opened a wiki-based website at the beginning of 2012. The BFT wiki platform is an open collaborative website, which means that access to the site and participation in the discussion forums is open to all. Content is created by a form of crowdsourcing, where the participation is limited to a particular group of people. This form of crowdsourcing has also been called *nichesourcing* (e.g. de Boer et al. 2012). In the BFT, researchers of each field can contribute by acquiring an expert status role on the website. The project's staff activate expert groups in different fields, provide support with terminology work, enter available terminologies (with copyrights) into the term bank, develop the website, inform the stakeholders about the term bank and give lectures, run courses and provide hands-on sessions concerning terminology work and practices of the term bank. The project was launched with three pilot fields: botany, jurisprudence and linguistics, but presently it welcomes every interested (sub)field with expert members prepared to commit to do terminology work in the term bank. Today, it includes 30 (sub)fields, with 38 000 term pages (that is concept pages, see figure 2 below, more than one 45 % of which contain definitions or explanations), and ca. 300 000 lemmas. The number of experts is approximately 500, most of whom, however, are only sporadically active.

The goals of the BFT are linked to the domains of language in the fields of arts and science and Finnish society. The multilingual properties of the term bank enhance the parallel use of different languages. The wiki interface is available in Finnish, Swedish, and English, and is currently being translated into Sami languages and Finnish Romani. The selection of languages given for term equivalents is unrestricted as different fields of research have

different needs for their selection of relevant languages. At present, term equivalents are provided in as many as 31 different languages in the BFT.

Thus while any single subject field on the wiki platform reflects the concept systems of one language, that language is not necessarily Finnish. Moreover, subject fields dealing with the same field but in a different language can be interlinked, and the expert groups working in different languages can benefit from each other's work across language barriers. To date, in addition to Finnish language subject fields, there are also fields established in Swedish, the other national language of Finland, and Romani, one of the minority languages in Finland. Moreover, there are fields in English in such cases where the expert group is international. We will address issues related to the choice of field language shortly.

The BFT has its roots in explicit concerns about the use of the Finnish language in academia. The Finnish language policy project commissioned by the Centre for Languages of Finland and the University of Helsinki concluded in its final report (Suomen kielen tulevaisuus 2009) that for academic discourse, and its relationship to other domains, the availability of up-to-date terminology is crucial if Finnish is to be used. It also became clear that there are many fields of science where, in practice, research is predominantly written in English. Furthermore, the use of English has spread throughout higher education to the extent that in some fields and universities, especially in economics[25], the number of thesis' written in English are catching up or exceeding those written in Finnish (47 % in University of Oulu Business School 2013-2015, 83 % in Aalto University School of Business). New terminology is only available in English, and teachers have difficulty finding equivalents in Finnish, even for basic courses. This, in turn, influences school books: how can pupils be informed about new inventions if there are no terms in Finnish? The question of terms thus concerns not only academic spheres, but also the media and the discourse of decision makers and everybody interested in research-based argumentation. The language policy project recommended that a survey

[25] Cf. University of Oulu Library's Publication Database Jultika (http://jultika.oulu.fi) and Aalto University eThesis portal (http://epub.lib.aalto.fi/en/ethesis).

be conducted into the current state of the terminologies of different fields, and urged for more support for terminology work and more centralized services. As a partial answer to these demands, the idea of the BFT was developed within the language policy group.[26]

The BFT also strives to strengthen terminological awareness across the fields of arts and sciences. When different fields share a common platform for their terminology work, it is possible, even unavoidable, to compare the extent to which different fields use the same terminology. Do the scopes of terms overlap? Is it possible to use the same definitions across different fields? Or are there hidden conceptual differences behind the apparent usage of the same terms? A wiki-based platform enables discussion of the terminological choices in the BFT, both within and between different fields. Consequently, the need to establish interdisciplinary fields becomes inevitable.

3.2.1 Contents and technology of the BFT

Wiki-platforms are usually designed for some form of community based content management. Consequently, the editorial process is based on transparency and openness, which require from the platform, among other things, an extensive and visible change log system, where users can track single revisions of any page and see when and why any single change to the content of a webpage was made. In addition to the complete history records of every page, the community usually requires a forum to discuss topics arising from the content of the wiki and debate preferences between competing versions. In order to maintain log and discussion data, wiki content is usually structured as independent webpages, where each webpage contains its own change log and discussions.

For the reasons mentioned above, the structure of the BFT's content is modelled on individual wiki pages with their own history records, discussion forums and stable URLs. Each concept article and each designation article is published as a separate wiki page. This way, the BFT wiki can maintain stable and transparent URLs for its contents, which

[26] The BFT project was put into practice by professor Lea Laitinen, who remains a member of the BFT steering group, as do other members of the language policy project.

is why web crawlers used by search engines can easily navigate the wiki structure. And this in turn means that any content page can function as a potential point of entrance to the website, and users can access BFT by simply googling a term. Stable, human readable URLs also makes it both easy to link BFT resources from outside and build very straightforward third party applications which use the BFT resources.

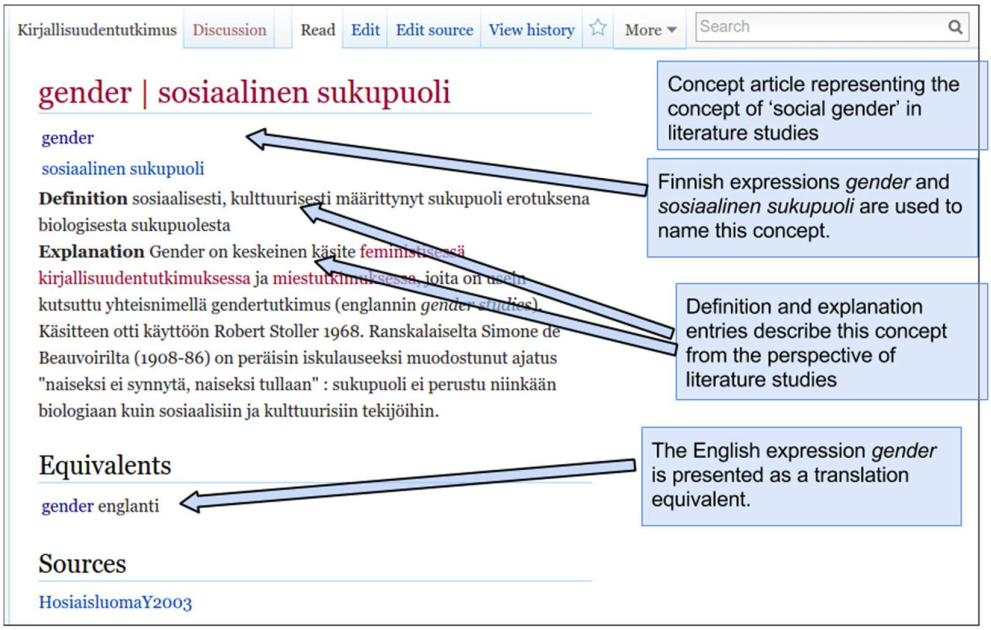

Figure 3. Concepts in BFT are represented by concept articles

In the BFT, concept and designation articles are separated into their own pages with differing layouts. Concept pages (figure 3) are intended to provide information on the concept, including synonymous terms, definitions, explanations, related concepts, translation equivalents, concept diagrams or other images. In contrast, designation pages contain linguistic, word-specific information, such as lexical category, information about word formation or etymological notes. A concept page can be linked to any number of designation pages, informing the reader that these expressions are used to denote this particular concept. Designation pages (figure 4), in turn, contain information about every concept page to which they are linked. The definitions in the designation page are formed dynamically from the contents of the respective concept pages. This structure enables

users to browse the BFT either as a concept-based terminology bank or as an designation-based lexicon, although this capacity is still to be fully implemented.

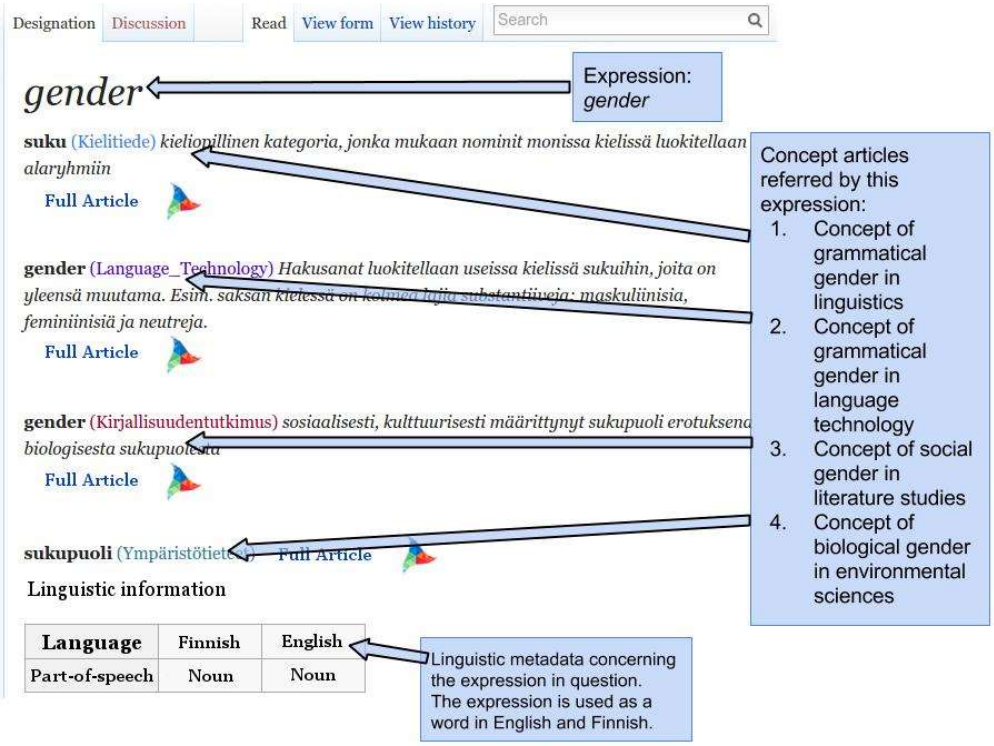

Figure 4. Example of designation page

Wiki discussion forums allow BFT experts to interact directly with a wider audience in a stable and moderated virtual environment. It allows concept forming negotiations within the expert community, which are sometimes necessary in order to satisfactorily define a more multifaceted or historically complex concept. Moreover, the forums can function as meeting points for experts seeking to understand a field outside their own discipline or define inherently multi-disciplinary concepts.

The definitions in the BFT are scientific, and thus they are primarily intended to serve the academic community. However, the explanations can also contain additional information and examples that illuminate the use and scope of the term for a wider audience. It is even possible to add a separate page with a wider description of the phenomenon or link the explanation to a Wikipedia article. Furthermore, academic and general audiences overlap

in various ways, as researchers are also laymen when it comes to terms outside their field of expertise. For both audiences, the additional value of the BFT, compared to the variable quality of content on the internet, is the reliability attained by nichesourcing of expert groups.

3.2.2 BFT challenges

To our knowledge, the idea of limited expert-based crowd-sourcing and wiki-based work is unprecedented in terminology work, at least on this scale, and as such it is not surprising that many challenges have arisen. Before addressing the greatest challenge, the activation of experts, first some words on how the BFT's approach differs from traditional terminology work. It is possible to start a new field in the BFT in very much the same way as in a more traditional terminological team, by using the expertise of the terminologist(s), by beginning with hierarchical concept diagrams with each subfield and by controlling how well definitions meet the standards of terminology. If there is a small, active and committed team working on the field, it is possible to continue this way. When the size of the expert group grows larger, however, the content likely becomes more heterogeneous.

Nevertheless, compared to more traditional terminology projects, the work of the BFT expert groups begins more at a grass-roots level, usually from a single concept and its related concepts. This means that instead of the traditional hierarchical orientation of terminology work, which emphasizes complete concept systems as a starting point, the orientation in the BFT is towards horizontal expansion, one immediately mappable micro-concept system at a time.

If the number and input of experts in each field reaches a critical mass, then it is quite possible for the content to evolve according to the community's self-correction and evaluation processes, where the discussion-based work of the experts overcomes any heterogeneity, and the collective terminology work slowly achieves the required quality standards. It is also useful to allocate different roles to some senior experts, including

monitoring the work in their respective subfields, managing the categorizations and tracking gaps in terms and definitions.

By testing various ways of launching a new field in the BFT, we have arrived at a model of what seems to be best practice. An effective way to begin a terminology project on a wiki platform seems to be to hire an experienced post doc or senior researcher to motivate the network of researchers in a given field to organize an expert group, create basic categorization for the field, survey existing terminologies and acquire copyrights to publish useful parts in the BFT content format, and begin writing definitions for the field in question. This phase usually takes between six months and one year.

There appears to be a certain threshold that needs to be crossed in the early stages of a new field, especially if the field lacks a history of terminology work in Finland's national languages. Thus far, there have been few cases where this threshold has been crossed on an entirely voluntary basis. BFT project staff can provide consultation on organizing terminology work and they can enter existing terminologies into the term bank, but each field requires expertise in the discipline to successfully begin. It is also possible to apply for external funding for an expert team that has reached an active phase in their terminology work. Thus far external funding has been attained for multidisciplinary terminology work in the humanities.

The greatest challenge is presented by the constraints of the experts' voluntary nichesourcing work. Learned societies of science and letters have performed voluntary terminology work in Finland since their establishment in the 19th century, thus the idea of terminology among academic professions is not completely alien. However, the working practices of professional researchers have changed drastically since the 19th century, which has had a clear effect on activities outside the scope of their immediate professional responsibilities, including the work of societies and terminology work. While many researchers appreciate the value of terminology work in their field in general, and that of BFT specifically, they are forced to focus on duties which are directly measured in the performance indicators of academic institutions, such as teaching and producing journal articles. Although it would not require tremendous effort to add the key

terms of an article or a lecture to the BFT while preparing it – the article could even be added to the BFT's references on the topic (to attain more citations) or the BFT's discussion forums could be used as a platform for virtual interaction on a course – researchers and lecturers do not yet see the BFT as a viable part of their jobs. There are also researchers, mostly in the natural sciences, who consider writing in Finland's national languages a waste of time. For them, science occurs through the medium of English, and researchers should not risk their careers by attempts at popularization. To reach a critical mass of experts in the BFT, it would be necessary for terminology work to be counted among the other indicators which govern the work load of professional researchers. The BFT's website in fact offers metrical tools for this: every change in the term bank leaves a trace in its history, and it is easy to follow the activities of any given researcher. The Language Policy of the University of Helsinki (2014: 55) states that the terminology work can be counted towards doctoral studies or for the workload of teaching and research stuff. If this policy comes true in practice and at other universities as well, it will be possible to connect the terminology work in the BFT to the routines of research and teaching.

3.2.3 BFT users

There are two main groups of BFT users. Researchers in the arts and sciences can add and modify the content of the term bank as registered members of expert groups dedicated to their respective fields of specialization. The larger group of users, in contrast, consists of the readers of the content provided by the experts. The content of the BFT is accessible to all via the internet, without restrictions. At this stage, something more needs to be said about each group, as they differ from each other in many ways.

First, expert groups in each field can decide their policy concerning the granting of expert status to new members. Different practices have already emerged, and they have also been purposefully tested in the pilots. The basic principle is that a new field is initiated by one or more professors or members of the board of a learned society. They can then decide whether they want to invite each new member personally or whether expert status is available to any interested researcher, university lecturer or even postgraduate student.

Usually all the members of a given expert group have the rights to modify the content of the field, but there are exceptions to this as well. The linguistics pilot project began with the widest expert group: every interested researcher from the postgraduate level up was (and is) allowed expert status. The botany pilot project began with an expert group of around 10 members, but it soon became clear that the greatest hindrance to the terminology work was the experts' limited opportunities for contributing to projects outside the scope of their immediate responsibilities. Thus, they offered membership to the field's entire academic community, composed of the Biological Society of Finland Vanamo and relevant research institutions. In contrast, the jurisprudence pilot project took a totally different route. Its terminology project is based on a centrally organized and hierarchical editorial organization consisting of a chief editor (professor emeritus Heikki E. S. Mattila), around 25 editors, and selected writers of the concept articles, whose work is supervised and allocated by their respective editors. The work began by updating articles in the paper volumes of *Encyclopaedia Iuridica Fennica* (1994–1999) and converting them into electronic form, as BFT concept pages. Concept articles are currently entered into the term bank in a centralized manner at the office of the Finnish Lawyers' Association. Jurisprudence is also an exception because it does not use the Creative Commons license otherwise adhered to in the BFT.

In spring 2014, the BFT conducted a user survey, and the results were encouraging. A key factor in whether users find and benefit from a web service is ease of use. 85 % of the survey respondents reported that the BFT platform was user-friendly and intuitive. While in terms of content the BFT project is still in its infancy, with all expert groups reporting that their work is far from completion, 88 % of respondents (see figure 5) nevertheless found the information they were seeking either completely (32 %) or partially (56 %).

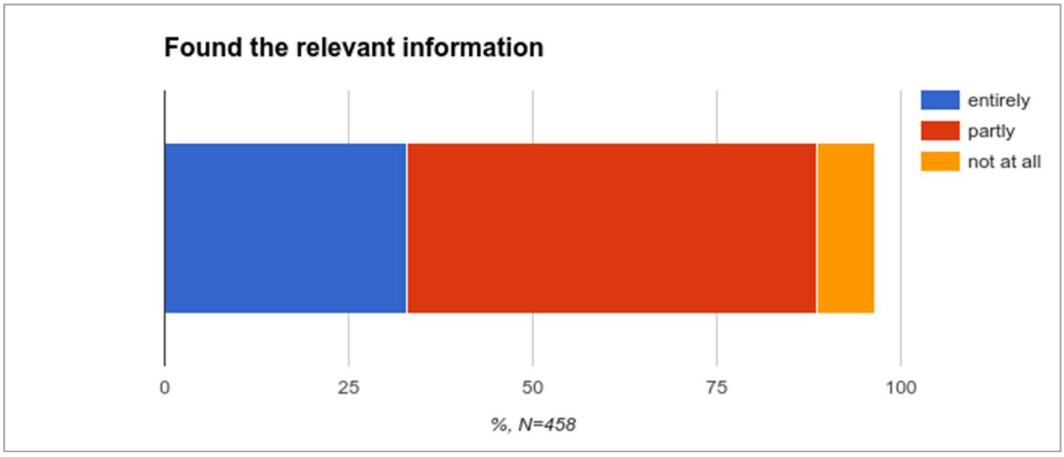

Figure 5. 2014 BFT User survey – question about whether the user found the information she/he was looking for

Most of the respondents (58 %) were university undergraduates, while the number of post-graduates and upper-secondary school students was notably smaller (11 % and 15 %, respectively). 41 % reported using the BFT for study purposes, while 14 % reported research, 8 % a school assignment and 8 % general interest as the purpose of their visit. See figures 6 and 7.

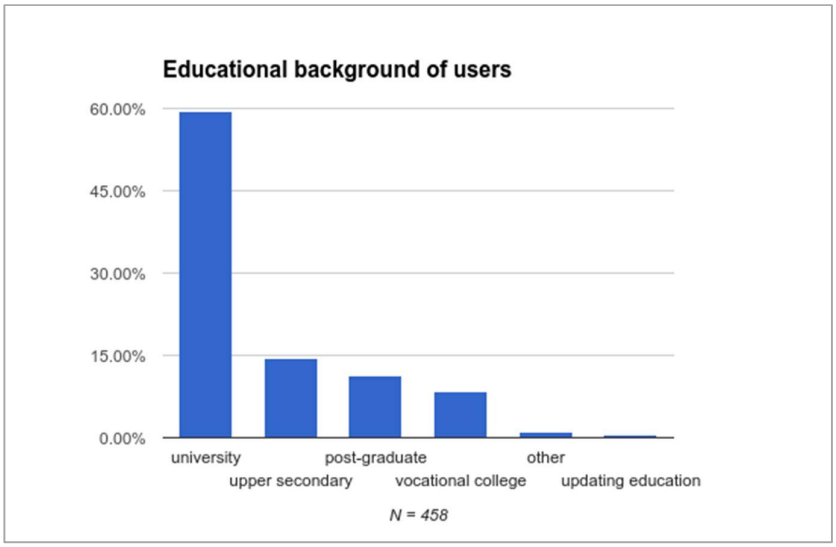

Figure 6. 2014 BFT User survey – question about educational background of the user

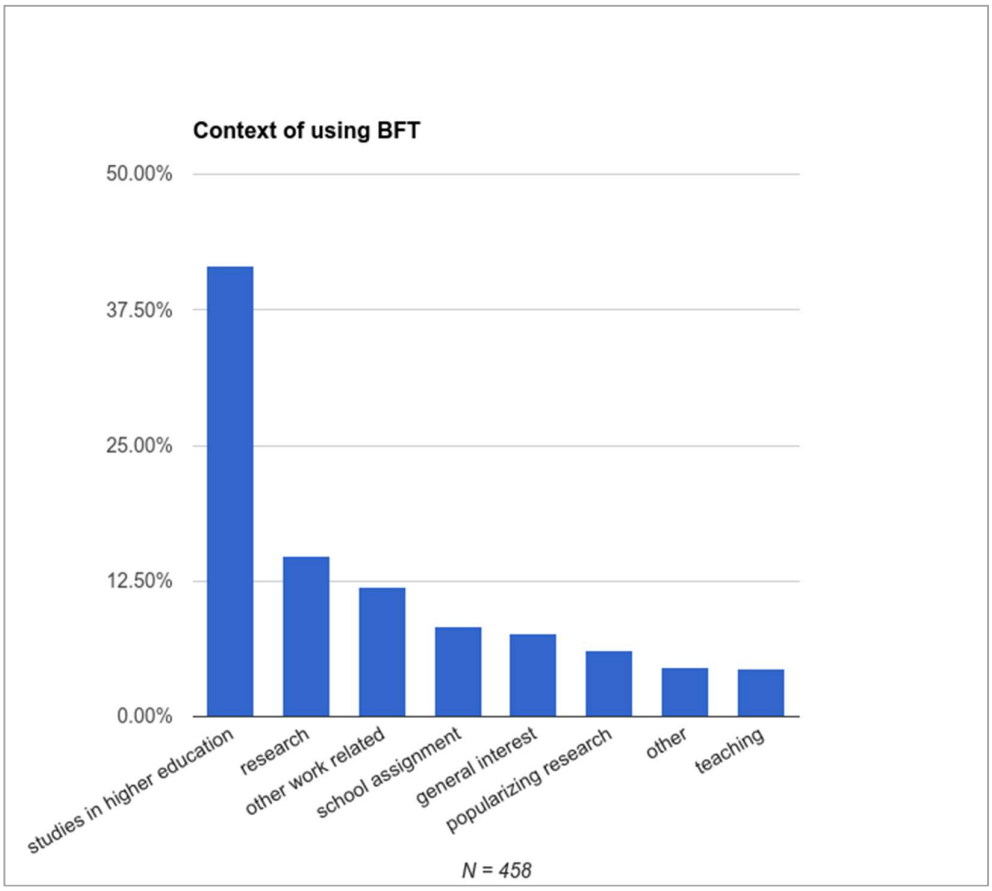

Figure 7. 2014 BFT User survey – question about the purpose of visit

At the same time the amount of content has continued to grow, with the number of pages and the number of fields of research increasing steadily. The amount of traffic in the BFT is growing steadily in terms of both visits and users, and it reached 10 000 weekly visits in autumn 2014.

3.2.4 BFT in the future

Current development plans include enhancing the personal interface used by the experts to allow them to more easily manage content dispersed across hundreds of single concept pages. Most of the wiki platforms have been designed to accommodate monitoring functions for a few dozen pages at time, but in terminology work the number of single terminological entries is far higher. In order to facilitate the monitoring process, the

Mediawiki platform's special annotation system, called semantic properties, is used to provide expert users with a larger and more comprehensive picture of the terminological content in their field.

The same interface can also be used to facilitate the work-flow within the wiki. One of the greatest difficulties in motivating a virtual community in terminology work on a wiki platform is the allocation of specific tasks that can be performed when a member has time to contribute. When experts enter the wiki, the system should provide a clear indication of how the terminology projectwould best benefit from their expertise without placing unreasonable demands on their time in the search for missing terms or imprecise definitions. These cues could be provided in the form of:

- lists of pages still lacking a certain content
- lists of search terms which fail to return a concept or designation page
- information on empty or stub pages (with not enough content to be useful) voted for completion by readers
- an overview of recent discussions within their field
- lists of candidate terms from external sources

To better facilitate the presentation of candidate terms to its experts, the BFT intends to work with large corpora of academic texts in the future. Automatic term extraction methods could be employed to scan repositories of journals, monographs, dissertations, reports and conference publications to seek gaps in the BFT's collections. The same techniques and repositories could also be used to link terms with relevant publications, to evaluate the scope and contemporaneity of the BFT's content and to provide contextual information about terms and concepts. Providing contextual information is often identified as an important, but under-developed, aspect of terminological resources.

The modified wiki platform also offers a starting point for international cooperation. For instance, there is a field in English for language technology in the BFT because a Nordic team is beginning terminology work in this field. Why then should not every field be

established in English? The decision, in practice, is one of language policy. If Finnish is to be fully developed as a language of the arts and sciences, much more is needed than mere translations of English terminology. Terms are central to the arts and sciences as they condense scientific classifications and frameworks of knowledge, and are thus an essential part of academic discourse. Moreover, terms and their definitions do not live a life separate from the context in which they appear or the discourse surrounding them; rather they evolve in constant interaction with these factors (cf. Temmerman 2000). Especially at the early stages of students' academic careers or when researchers seek knowledge from a field outside their own discipline, fine-grained, a more nuanced understanding is better achieved in their native language. Consequently, the development of national and minority languages in all domains of society is a question of equal opportunities. It is essential for small nations such as Finland to find talent from all levels of society, not just from those groups who have access to multilingual education.

It is also possible to display parallel definitions in, say, Finnish and English on the same page[27]. Moreover, international terminology work has the potential to change language priorities. An open question is also whether there should be a common international terminological portal for arts and sciences. At present we wish to explore the possibility of linking term banks, specifically those which utilize wiki platforms, and developing open access portals that provide links through search engines to many terminological resources across fields and languages.

4 Sammanfattning

Att bygga upp en termbank är en aktivitet som kräver resurser och en stor portion struktur och systematik. Att bygga upp en nationell termbank, som *Rikstermbanken* eller *Vetenskapstermbanken*, innebär ytterligare utmaningar vad gäller innehållet och tekniken. Uppbyggnaden av Rikstermbanken och Vetenskapstermbanken har, som artikeln visat, inneburit en hel del olika överväganden och beslut varav några kan vara användbara för

[27] This is unproblematic in the fields where terminology is harmonized, and fruitful in the fields where harmonizing is the goal of the terminology work of the expert group. Otherwise it is often the case that definitions and scopes of terminological equivalents don't match in different languages.

andra liknande termbanksprojekt. Ett agilt arbetssätt, där justeringar hela tiden kan göras, har visat sig vara fruktbart, och som så ofta leder pragmatiska omständigheter till beslut som teoretiskt kan diskuteras; några av de mest ambitiösa planerna vad gäller innehållsbedömning och administration fick till exempel överges i Rikstermbankens fall, men det har visat sig fungera bra ändå.

Nyttan med en nationell termbank har också framgått: en nationell termbank kan ge en terminologisk överblick som oftast inte varit möjlig tidigare genom att den innehåller termer och definitioner från flera olika fackområden. Samtidigt bidrar Rikstermbanken i Sverige till att effektivisera myndigheternas terminologiarbete i Sverige och Vetenskapstermbanken aktiverar forskarnas terminologiarbete, vilket också ökar kvaliteten på resultatet: Rikstermbanken är lättillgänglig, och det är enkelt att ta reda på om någon redan har definierat ett visst begrepp som man sedan kan återanvända eller ta som utgångspunkt för en egen definition. Detsamma gäller för Vetenskapstermbanken, även om man där även uppnår ett viktigt förenhetligande av terminologin inom den vetenskapliga världen, och då särskilt universitets- och högskolevärlden.

Den tidsbesparing som enkel åtkomst till terminologi ger kan inte underskattas. Men enkel åtkomst är inte liktydigt med aktiv användning. En aktiv användning av terminologin förutsätter att termerna kan integreras med andra verktyg, för till exempel ordbehandling eller innehållshantering för webben, och utbytas smidigt. Då kan ytterligare effektiviseringar uppnås. Det är tydligt att en nationell termbank aldrig blir klar.

Referenser

de Boer, Victor, Michiel Hildebrand, Lora Aroyo, Pieter De Leenheer, Chris Dijkshoorn, Binyam Tesfa & Guus Schreiber (2012). Nichesourcing: Harnessing the Power of Crowds of Experts. I: *Knowledge Engineering and Knowledge Management*. 18th International Conference, EKAW 2012, Galway City, Ireland, October 8–12, 2012. Proceedings. Lecture Notes in Computer Science. Volume 7603, 16–20. Red. Ten Teije, Anette, Johanna Völker, Siegfired Handschuh, Heiner Stuckenschmidt, Mathieu d'Acquin, Andriy Nikolov, Nathalie Aussenac-Gilles & Nathalie Hernandez. Berlin: Springer Verlag.

Bowker, Lynne (2015). Terminology and translation. I: *Handbook of Terminology*, 304–323. Eds. Hendrik Kockaert & Frieda Steurs. Amsterdam: John Benjamins.

Dobrina, Claudia & Henrik Nilsson (2007). Koll på innehåll. Vad ska in i en nationell termbank? I: *NORDTERM 15 Kunnskap og fagkommunikasjon,* 69–75. Red. Jan Hoel. Bergen: Språkrådet.

Dobrina, Claudia & Henrik Nilsson (2009). Koll på innehåll igen: Vad har kommit med i Rikstermbanken? I: *NORDTERM 16 Ontologier og taksonomier* (cd-rom), 120–132. Red. Bodil Nistrup Madsen & Hanne Erdman Thomsen. Köpenhamn: Institut for Internationale Sprogstudier og Vidensteknologi & DANTERMcentret.

EuroTermBank Consortium (2006). *Towards consolidation of European Terminology Resources. Expérience and Recommendations from EuroTermBank Project*. Riga: Tilde.

Infoterm (2005). *Guidelines for terminology policies. Formulating and implementing terminology policy in language communities*. Paris: United Nations Educational [citerat 15.6.2015]. Tillgänglig: http://unesdoc.unesco.org/images/0014/001407/140765e.pdf.

ISO (1999). ISO 12200: *Computer applications in terminology – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange.*

ISO (2003). ISO 16642: *Computer applications in terminology – Terminological markup framework (TMF).*

ISO (2012). ISO 26162: *Systems to manage terminology, knowledge and content – Design, implementation and maintenance of terminology management systems.*

ISO (2008). ISO 30042: *Systems to manage terminology, knowledge and content – TermBase eXchange (TBX).*

Iuridica Encyclopaedia Fennica I-VII (1994–1999). Helsinki Suomalainen Lakimiesyhdistys.

McNaught, John (1988). A survey of termbanks worldwide. I: *Translating and the Computer 9. Proceedings of a conference 12–13 November 1987*. 112–129. Ed. Catriona Picken. London: Aslib.

Nilsson, Henrik (2005). TISS & IATE. Svensk terminologisk infrastruktur och svensk rikstermbank. I: *Nordterm 14: Ord og termer*. 216–230. Reykjavik: Stofnun Árna Magnússonar í íslenskum fræðum.

Regeringens proposition 2005/06:2. (2005). *Bästa språket – en samlad svensk språkpolitik* [citerat 15.8.2015]. Tillgänglig: http://data.riksdagen.se/fil/8E6432AF-2A6D-42D1-A9AD-DD3097BD91D4.

Regeringens proposition 2004/05:175 (2005). *Från IT-politik för samhället till politik för IT-samhället* [citerat 15.8.2015]. Tillgänglig: http://data.riksdagen.se/fil/1D8C0013-04D2-4A6A-8083-AE45FC1081F3

Riksrevisionen (2012). Citybanans regionala medfinansiering. *RiR* 8, 21–22. Stockholm: Riksrevisionen.

Schmitz, Klaus-Dirk (2005). Developments in computational terminology management and its influence on terminology science and terminology work. In: *ΕΛΕΤΟ – 5ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία» Λευκωσία, Κύπρος, 13–15 Νοεμβρίου*, 878. Aten: Eleto.

Suomen kielen tulevaisuus (2007). *Language Policy Action Plan* [citerat 12.1.2015]. Kotimaisten kielten tutkimuskeskuksen julkaisuja 155. Helsinki. Tillgänglig: http://scripta.kotus.fi/www/verkkojulkaisut/julk7/.

Svanberg, Peter (2009). Från Termdok på cd-rom till Rikstermbanken. I: *NORDTERM 16 Ontologier og taksonomier* (cd-rom), 133–139. Red. Bodil Nistrup Madsen & Hanne Erdman Thomsen. Köpenhamn: Institut for Internationale Sprogstudier og Vidensteknologi & DANTERMcentret.

Temmerman, Rita (2000). Towards New Ways of Terminology Description. *Terminology and Lexicography Research and Practice 3*. Amsterdam: John Benjamins.

TNC (2011). *En nationell semantisk resurs för det digitala samhället.* PM till Näringsdepartementet (opubl.).

TNC (2008). *Terminologins terminologi: ordlistan*. Tillgänglig: http://www.tnc.se/Nordtermpublikationer/35-Terminologins-terminologi/flypage.tpl.html. Solna: TNC.

TNC (2004). *Ja till IATE? Testning och utvärdering av en tänkbar programvara för en framtida svensk rikstermbank.* Slutrapport projekt 25303-1 (Vinnova 2004-00473). Solna: TNC.

University of Helsinki (2014). *Language Policy of the University of Helsinki.* Unigrafia. Helsinki.

Wright, Sue Ellen & Gerhard Budin (2001). Infobox 28: TermbaseTerminology. I: *Handbook of Terminology Management, Volume 2*. 878. Eds. Sue Ellen Wright & Gerhard Budin. Amsterdam: John Benjamins.

Åström, Kjell (1983). Aspects of term bank operation. I: *Term banks for tomorrow's world : translating and the computer 4 : proceedings of a conference 11–12 November 1982*. 179–185. Ed. Barbara Snell. London: Aslib.

Termbanker

Land/region	**Termbank**	**URL**
Baskien (Spanien)	Euskalterm	http://www.euskara.euskadi.eus/r59-euskalte/eu/q91EusTermWar/kontsultaJSP/q91aAction.do
Danmark	Dansk termbank	http://dantermbank.cbs.dk/dtb (under utveckling)
Finland	Tieteen kansallinen termipankki (TTP) | Vetenskapstermbanken | BFT	http://tieteentermipankki.fi/wiki/Termipankki:Etusivu
Finland	TEPA	http://www.tsk.fi/tepa/netmot.exe?UI=engr&height=15
Frankrike	FranceTerme	http://franceterme.culture.fr/FranceTerme/
Irland	Téarma.ie	http://www.tearma.ie/Home.aspx
Island	Orðabanki	http://www.arnastofnun.is/page/gagnasofn_ordabanki
Kanada	Termium	http://www.termium.com
Kanada	Grand dictionnaire terminologique (GDT)	http://www.granddictionnaire.com/
Katalonien (Spanien)	Cercaterm	http://www.termcat.cat/ca/Cercaterm/Fitxes/
Kroatien	Struna	http://struna.ihjj.hr/
Litauen	Terminų bankas	http://terminai.vlkk.lt/pls/tb/tb.search
Norge	Termportalen	http://www.terminologi.no/forside.xhtml (under utveckling)
Sápmi	Sátni	http://www.satni.org/index.html
Sverige	Rikstermbanken	http://www.rikstermbanken.se
Turkiet	Bilim ve sanat Terimleri Ana Sözlüğü	http://www.tdk.gov.tr/
Wales	MAES-T	http://maes-t.com/