


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	Parallellkorpuser som resurs i leksikografiskt arbete	
Forfatter:	Håkan Jansson	
Kilde:	Nordiska Studier i Lexikografi 11, 2012, s. 329-339 Rapport från Konferens om leksikografi i Norden, Lund 24.-27. maj 2011	
URL:	http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive	

© Nordisk forening for leksikografi

Betingelser for brug af denne artikel

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

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

Søgbarhed

Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition' og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.

Parallellkorpusen som resurs i lexikografiskt arbete

Håkan Jansson

This paper discusses the use of parallel corpora as a tool in bilingual lexicography. After a brief background sketch with reference to literature and (lack of) reported ongoing projects, the text briefly considers the notion of a parallel corpus, before it reaches the main subject: a hands-on study of the opportunities connected to the use of parallel corpora in lexicography.

The corpus used for the study is a relatively small, about 115,000-token, ad-hoc corpus compiled of texts from the Nordic Council. The languages investigated are Icelandic on the one hand, and Danish, Norwegian and Swedish on the other. One advantage of parallel corpora seems to be the possibility to compare concordances of n-grams between the studied languages. The two Icelandic verbs probed as examples were shown to be part of hitherto unnoticed equivalent pairs, where one or both sides turned out to be MWEs. Co-occurring patterns between active or passive voice and certain translation equivalents have also been noticed.

Nyckelord: korpusbaserad lexikografi, parallellkorpus, tvåspråkig lexikografi, n-gram, flerordsuttryck

Inledning

Den lexikografiska litteraturen har sedan ett antal år noterat möjligheterna att använda parallellkorpusar i lexikografiskt arbete. Två exempel på tvåspråkiga lexikonprojekt där parallellkorpusar i någon mån varit aktuella är dels *The Oxford-Hachette French dictionary* (Corréard et al. 2007), dels den kommande *New English-Irish Dictionary* senast beskriven i Atkins et al. (2010). Än så länge har det dock inte publicerats några beskrivningar som presenterar bruket och nyttan av parallellkorpusar i dessa projekt.

Det här bidraget avser att beskriva arbetet med en förhållandevis liten korpus sammansatt av texter på isländska, danska, norska och svenska från Nordiska Rådet. Studien kommer att utgå från översättningar till isländska, där två

isländska verb i alla dess böjningsformer väljs ut för närmare undersökning med utgångspunkt i de parallella texterna. Särskilt intresse riktas mot återkommande ordsamband. Korpusens parallella texter används för att hitta kandidater till ekvivalentpar, som sedan i någon mån jämförs med de förslag till ekvivalenter och konstruktioner som påträffas i lexikon.

Bakgrund

Allt sedan Svensk ordbok (SOB) (1986) och Collins-COBUILD English language dictionary (COBUILD) (1987) kom ut, har arbetet med korpusar som bas för enspråkig lexikografi blivit en alltmer självklar förutsättning. I tvåspråkig lexikografi har utvecklingen mot korpusbruk inte varit lika tidig och tydlig, men Svensén (2004: 59) rapporterar att *Oxford-Hachette French dictionary* (Corréard & Grundy 1994) för första gången 1994 använde två enspråkiga korpusar, en engelsk och en fransk, till stöd för arbetet med källspråket och målspråket var för sig. För att fullt ut kunna jämföra två korpusar vid själva valet av översättningsekvivalenter krävs dock någon typ av parallellstruktur mellan korpusdata i källspråket och i målspråket. Atkins & Rundell (2008: 473 ff.) beskriver tre olika nivåer av korpusstöd vid arbetet med tvåspråkiga lexikon:

1. *Parallella konkordanser*. Man undersöker konkordansen för det lemma som ska analyseras i ett fönster, samtidigt som man undersöker konkordansen för den tilltänkta översättningsekvivalenten i ett annat fönster. Det finns inte någon matchning mellan texterna i korpusarna.
2. *Parallella korpusar* med matchande texter. Ingen av de ingående texterna är en översättning av den andra, men de är matchade med hänsyn till genre och/eller domän eller ytterligare relevanta faktorer.
3. *Parallellkorpus* med länkning på meningsnivå. Här är en av texterna en översättning av den andra, och meningarna har länkats till varandra. De översättningsekvivalenter som dyker upp här, har valts ut av en översättare med tanke på att de ska passa i den aktuella kontexten.

Vid arbete med den första och andra typen av korpusar måste ordboksredaktören själv ha en idé om vilken översättningsekvivalent som ska användas, medan den tredje typen genom konkordansen kan ge förslag på översättningsekvivalenter.

Benämningarna på dessa tre typer av korpusar är inte helt entydiga i den internationella litteraturen. Typ 2, *parallella korpusar* kallas ibland för *parallellkorpus*, medan typ 3 *parallellkorpus* ibland kallas för *översättningskorpus* (*translation corpus*). Den undersökning som här presenteras kommer uteslutande att

gälla korpusar av typ 3, och de kommer genomgående att benämnas *parallellkorpusar*.

Komparativa fördelar respektive nackdelar med arbete med de tre ovanstående korpus typerna har diskuterats av bl.a. McEnery et al. (2006) och Atkins & Rundell (2008). En viktig fördel med den tredje typen är att den kan ge förslag på översättningsekvivalenter och konstruktionsmönster. En nackdel kan vara att översatta texter ofta kan innehålla uttryck och konstruktioner som är mer typiska för originaltexten än för det språk den ska översättas till. Detta har bl.a. påtalats av Gellerstam (1996) samt av Borin & Prütz (2001). Man måste dessutom vara medveten om att den enskilde översättaren inte sitter inne med någon gudomlig sanning i frågan om vilken av flera tänkbara ekvivalenter som är mest lämplig i olika fall. Men i och med att översättare alltid arbetar med hela texter som är relaterade till en viss kontext, kan studier i deras val av översättning ge lexikografen uppslag till kontextuella perspektiv på valet av lexikonekvivalenter.

Det mesta som har skrivits om parallellkorpusar har skrivits inom språkteknologi, lingvistik eller inom vissa specifika språk. I Norden är särskilt kontrastiva studier inom anglistik relevanta. Bland beskrivningar av metoder och möjligheter med parallellkorpusar kan särskilt Johansson & Hofland (2007) och Hasselgård (2010) vara av intresse här. Såväl Johansson & Hoflands som Hasselgårds texter kan dessutom ge en grund för att förstå metoden bakom arbetet med flera stora parallellkorpusar där ett av de ingående språken är nordiskt. Bland de korpusar som nu finns tillgängliga med ett ingående nordiskt språk är *Europarl* (da-en, fi-en samt sv-en; se vidare Koehn 2005 för beskrivning av Europarl). Vidare kan *English-Norwegian Parallel Corpus*, beskriven av Johansson, Ebeling & Oksefjell (1999/2002), nämnas. Ytterligare en korpus, beskriven av Altenberg, Aijmer & Svensson (2001), är *The English-Swedish Parallel Corpus*.

Det finns dock ännu inte någon större korpus där nordiska språk jämförs med varandra. Arbete med att ta fram en sådan korpus pågår dock vid Språkbanken vid Institutionen för svenska språket vid Göteborgs universitet. De principer som styr arbetet med denna korpus är desamma som de som ligger till grund för SALT-projektet – som beskrivs av bl.a. Rawoens (2003). De ingående språken i den internordiska korpusen är danska, färöiska, isländska, norska och svenska.

Material och metod

Materialet i den här studien kommer från en mindre *ad hoc-korpus* bestående av texter från Nordiska rådet. Beslutet att välja texter från Nordiska rådet är i huvudsak grundat på att texterna finns lätt tillgängliga på Internet och att de är fria ur copyrightsynpunkt. Korpusen består av nyhetsbrev från Nordiska Rådet från perioden 2006–2009 samt en kortare rapport, *Friskare nordiska hav* (2010). Gemensamt för alla texter är att de är översatta *till* isländska¹. Språket i källtexterna varierar mellan danska, norska och svenska i nyhetsbreven, medan språkparet i rapporten är svenska-isländska. Storleken på korpusen är ca 115 000 löpord fördelat på ca 15 000 ordformer. Domänen för korpusen bedöms vara samhällsdebatt medan genretillhörigheten är facktext/populärvetenskap.

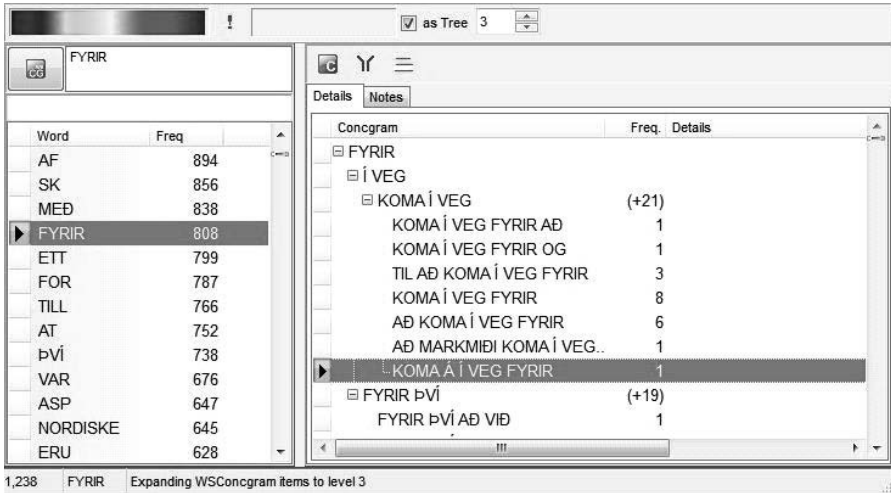
Arbetet med att föra samman texterna till en korpus har gjorts med hjälp av programvaran WordSmith 5.0, hädanefter WS5, vilken närmare beskrivs av Scott (2008). Eftersom vare sig nyhetsbrev eller rapport hade en helt identisk layout i de olika språkversionerna var det nödvändigt att för-redigera texterna, så att deras respektive textstycken hjälpligt motsvarade varandra, innan de parallelliserades på meningsnivå i WS5.

Ett område där utnyttjandet av parallellkorpusar kan vara av avgörande betydelse är när man vill hitta ekvivalentpar där en av, eller båda, de ingående ekvivalenterna är ett flerordsuttryck. För att särskilt kunna studera sådana ekvivalentpar användes den funktion i WS5 som kallas *Conc-gram*. WS5:s *Conc-gram*-funktion är en tillämpning av det som i litteraturen i dag normalt benämns *n-gram*, men som även varit känt under namn som *bi-gram* och *skip-gram*; se t.ex. Cheng et al. (2006) för en översikt över utvecklingen av funktionen i olika applikationer.

Trots att isländska i det här fallet är målspråk, kom alla sökningar att göras med utgångspunkt i isländska. Det är annars vanligt att man gör sökningar som dessa med utgångspunkt i källspråket, men eftersom källspråken förutom svenska även inkluderade norska och danska var isländska här det enda gemensamma språket.

Eftersom ekvivalentparen kan bli betydligt fler och betydligt mer komplicerade vid flerordsuttryck koncentrerades undersökningen på isländska flerordsuttryck med verb *koma fyrir* och *standa fyrir*. Figur 1 visar ett exempel på hur en sökning på *fyrir* i olika kombinationer i *Conc-gram*-funktionen i WS5 kan se ut. Lägg särskilt märke till det markerade resultatet *koma á í veg fyrir* som visar att den sökta ordkombinationen inte måste vara kontinuerlig med de ingående komponenterna i någon viss ordning.

¹ Enligt e-postkommunikation med Michael Funch, huvudredaktör för nyhetsbrevet.



Figur 1. Översikt över uppfångade flerordsuttryck med fyrir i experimentkorpusten

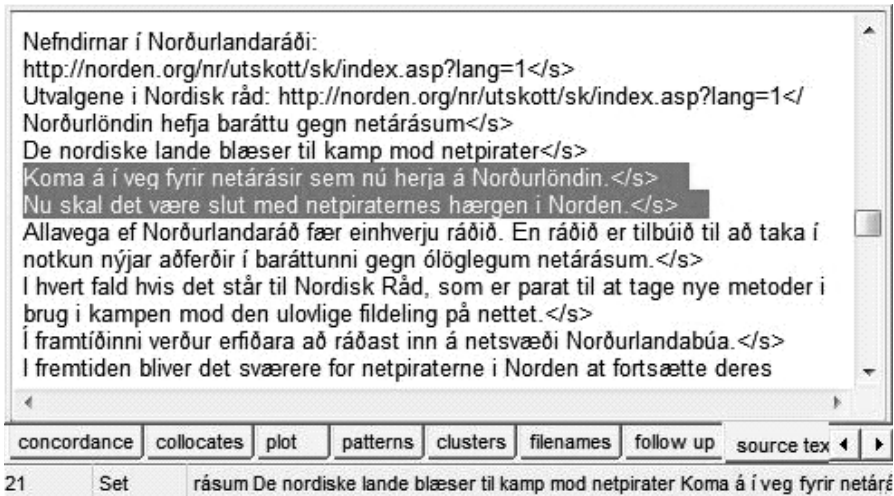
När man har hittat ordkombinationer av ett slag man vill undersöka kan man klicka sig vidare till en KWIC-konkordans (se figur 2). Vi ser här att den ordkombination som var markerad i figur 1 återkommer i kontext i figur 2.

N	Concordance	File	%
1	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
2	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
3	samstarf til að koma í veg fyrir mansal, fíkniefnasmygl, netglæpi,	2009-09-01.txt	96%
4	til að draga úr og koma í veg fyrir olíuleka og afleiðingar þeirra á	Renare_hav.txt	95%
5	en er ekki jafnmikilvægt að koma í veg fyrir minni háttar losun í tengslum við	Renare_hav.txt	92%
6	en er ekki jafnmikilvægt að koma í veg fyrir minni háttar losun í tengslum við	Renare_hav.txt	92%
7	munu löndin einnig koma í veg fyrir stjórnsýsluhindranir í viðskiptum	2009-02-16.txt	90%
8	borde utökas. Vilji menn koma í veg fyrir frekari mengun og eitrun	Renare_hav.txt	77%
9	í sveitarfélögum en að koma í veg fyrir leka frá ökrum? Á det då bättre att	Renare_hav.txt	64%
10	til kamp mod netpirater Koma á í veg fyrir netárásir sem nú herja á	2009-02-02.txt	56%
11	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
12	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
13	öryggisráðstafanir til að koma í veg fyrir að lagning gasleiðslu frá Rússlandi	2006-09-11.txt	55%
14	tilraun ESB til þessa til að koma í veg fyrir hnignun vatns og vatnakerfa.	Renare_hav.txt	45%
15	tilraun ESB til þessa til að koma í veg fyrir hnignun vatns og vatnakerfa.	Renare_hav.txt	45%
16	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
17	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
18	skattaparadísum til að koma í veg fyrir undanskot í hinu alþjóðlega	2009-05-25.txt	42%
19	kvinnehandel Hvernig má koma í veg fyrir verslun með konur? Hvordan kan	2006-10-02.txt	34%
20	með það að markmiði að koma í veg fyrir grófa glæpastarfsemi eins og	2009-09-07.txt	31%
21	með það að markmiði að koma í veg fyrir grófa glæpastarfsemi eins og	2009-09-07.txt	31%

At the bottom of the table, there are tabs for 'concordance', 'collocates', 'plot', 'patterns', 'clusters', 'filenames', 'follow up', 'source text', and 'notes'. Below the table, a snippet of text is visible: '21 | Set | rásum De nordiske lande blæser til kamp mod netpirater Koma á í veg fyrir netárásir sem nú herja á Norðurlöndin. Nu skal det'

Figur 2. Kwic-konkordans med koma í veg fyrir som sökbegrepp

Från konkordansen i figur 2 kan man i sedan i sin tur klicka sig till de länkade parallelltexterna (se figur 3).



Figur 3. Länkade parallelltexter

I dessa länkade parallelltexter kan man så undersöka hur verksamma översättare träffar sina val om översättningsekvivalenter. Dessa val kan sedan jämföras med resultaten av de olika översättningarna som görs av ordboksredaktionerna.

Resultat: undersökning av isländska flerordsuttryck

Den här studien bygger på en närmare undersökning av isländska flerordsuttryck som bildas med två vanliga isländska verb som bas: *koma* [sv. komma] och *standa* [sv. stå]. Experimentkorpuser innehåller 148 belägg för verbet *koma* i alla dess ordformer, varav 15 fall av flerordsuttrycket *koma í veg fyrir* [sv. förhindra]. *Koma í veg fyrir* är den ekvivalent som Svensk-isländsk ordbok (Holm & Davíðsson 1982), hädanefter SvIO, ger till uppslagsordet *förhindra*; jämför även ISLEX (2011) under uppslagsordet *koma*. Följande språkprov ur parallellkorpuser ger en tydligare bild av grundbetydelsen:

- (1) Är det då bättre att satsa på kommunala reningsverk eller på att *förhindra* urlakning från åkrarna?
 Borgar sig þá frekar að koma upp hreinsunarstöðvum í sveitarfélögunum en að *koma í veg fyrir* leka frá ökrum?

De femton fall av *koma í veg fyrir* som finns i korpusen har också *förhindra* som enskilt vanligaste källspråksord. Andra ord som översätts med *koma í veg fyrir* är *stoppa* exempel (2) och *forebygge* exempel (3):

(2) stoppa (sv)

/.../ minska och stoppa oljeutsläppen och deras effekter.

/.../ draga úr og koma í veg fyrir olfuleka og afleiðingar þeirra /.../

(3) forebygge (no)

/.../ forebygge menneskehandel, ulovlig handel med narkotika, internetkriminalitet, miljøkatastrofer og pandemier /.../

/.../ koma í veg fyrir mansal, fíkniefnasmygl, netglæpi, umhverfisslys og farsóttir /.../

Bland de övriga källspråksformuleringar som översätts med *koma í veg fyrir* hittade vi uttryck som *bekjempe* exempel (4) och *komme /.../ til livs* exempel (5), vilket visar att alla parallella uttryck i korpusen inte är direkta kandidater till att bli översättningsekvivalenter i ett lexikon, utan snarare får ses som ett försök att skapa dynamisk ekvivalens från översättarens sida:

(4) bekjempe (no)

/.../ bekjempe den grove kriminaliteten som menneskehandel er,

/.../ koma í veg fyrir grófa glæpastarfsemi eins og mansal,

(5) komme /.../ til livs (no)

Hvordan kan man *komme* kvinnehandel *til livs*?

Hvernig má *koma í veg fyrir* verslun með konur?

Det finns även friare formuleringar än dem vi ser i (4) och (5) som leder till *koma í veg fyrir*. Ett intressant sådant exempel är (6), där *oppná fordeler* [rörande] *förhindringer* omtolkas till *koma í veg fyrir /.../ hindranir* med omvänd polaritet:

(6) Ved å samles om ett ernæringsmerke vil alle landene også *oppná fordeler når det gjelder grenseforhindringer* ved handel.

Með því að sameinast um eitt næringarmerki munu löndin einnig *koma í veg fyrir stjórnsýsluhindranir* í viðskiptum milli landanna.

Den vanligaste ekvivalenten till *koma í veg fyrir* är alltså *förhindra*, och om man undersöker detta lemma i samma korpus ser man att det förekommer tio gånger, och översätts fyra gånger med *koma í veg fyrir*, fyra gånger med *stöðva* [sv. stoppa] (7) respektive en gång vardera med *hindra* (8) [sv. hindra] och *sporna við* (9) [sv. spjärna emot].

(7) stöðva (isl)

Nordiskt projekt för att *förhindra* skatteflykt fortsätter
Samstarf um leiðir til að *stöðva* skattaundanskot heldur áfram

(8) hindra (isl)

som huvudsakligen sprids via /.../ båtbottnfärger som *förhindrar* algpåväxt.
sem dreifist aðallega með /.../ botnbátamálningum sem *hindrar* þörunga-
vöxt.

(9) sporna við (isl)

/.../ och hur vi ska *begränsa* och *förhindra* framtida katastrofer.
/.../ og hvernig hægt er að *sporna við* stórslyssum í framtíðinni.

Korpusexemplen har visat att *koma í veg fyrir* är en högfrekvent flerordsenhet, som, om den av redigeringsprinciper inte ges lemmastatus, åtminstone borde ges en framträdande plats i mikrostrukturen. Ekvivalentparet *förhindra* ↔ *koma í veg fyrir* finns med i såväl SvIO som ISLEX, men analys av korpusexemplen visar att det kan vara motiverat att även ge *stoppa* och *förebygga* som ekvivalenter till *koma í veg fyrir*. Det är viktigt att understryka att existensen av parallella uttryck i källspråk och målspråk i korpustexterna inte i sig utgör ett bevis för att visst ord ska få en bestämd översättningsekvivalent. Översättarnas val måste analyseras. Exempel (2) och (3) ovan – med *stoppa* och *förebygga* – visar på översättningar som är tydliga kandidater till att tas med i lexikon, medan (4) och (5) är resultat av arbete med dynamisk ekvivalens och därmed inte relevanta för en ordboksartikel.

Ett annat isländskt högfrekvent flerordsuttryck är *standa fyrir*. Av 88 belägg för verbet *standa* i alla dess former i korpusen gäller 24 olika former av *standa fyrir*. Den svenska grundbetydelsen för *standa fyrir* är 'arrangera', 'stå för' (Jfr ISLEX, 2011 *standa*). SvIO saknar dock *standa fyrir* som ekvivalent till *arrangera*, och *stå för* finns inte med i SvIO, vare sig som eget lemma eller som kollokation under *stå*. Följande korpusexempel visar hur *arrangera/standa fyrir* typiskt används:

(10) arrangera (sv)

Nordiska ministerrådet och NordGen, /.../, *arrangerar* den 13 maj ett evenemang i FN-byggnaden /.../

Norræna ráðherranefndin og NordGen, /.../, *standa fyrir* viðburði í byggingu Sameinuðu þjóðanna þann 13. maí /.../

En genomgång i parallellkorpusen ger också vid handen att *arrangera* är det källspråksord som oftast ger resultatet *standa fyrir*. Tabell 1 (nedan) visar vilka danska, norska och svenska verbkonstruktioner som har översatts med *standa fyrir*.

Danska källspråksord	Norska källspråksord	Svenska källspråksord
organisere (2 ex)	arrangere (5 ex)	arrangera (6 ex)
arrangere	være arrangør	stå bakom
koodinere (i passiv med <i>blive</i>)	være ansvarlig	genomföra
lancere	byde ... på	ha (2 ex)
stå som vært		

Tabell 1. Verbkonstruktioner som översätts med *standa fyrir*

Tabellen visar det för norska och svenska tydliga mönstret att det är *arrangera/arrangere* som leder fram till ekvivalenten *standa fyrir*. Bland andra ekvivalentpar märks bl.a. följande:

(11) byde ... på (no)

På kvelden *byr* informasjonskontoret *på* konsert /.../

Um kvöldið *stendur* upplýsingaskrifstofan *fyrir* tónleikum /.../

(12) stå som vært (da)

/.../ Nordisk Ministerråds kontorer *står som værter*.

/.../ Norræna ráðherranefndin *stendur fyrir* námstefnunum.

(13) ha (sv)

Nordiska ministerrådet *har* också en rad olika program och aktiviteter som sträcker sig långt utanför de nordiska länderna.

Norræna ráðherranefndin *stendur einnig fyrir* fjölmörgum áætlunum og aðgerðum sem ná langt út fyrir Norðurlönd.

I exempel (11) ser vi att den norska och den isländska satskonstruktionen är identisk, även om det nog kan anses att det är en liten semantisk skillnad mellan *stå för* [isl. *standa fyrir*] och *bjuda på*. Exempel (12) visar hur skillnaden i föredragen aktantuppsättning mellan *stå som vært* och *standa fyrir* leder till att den isländska översättaren ser sig nödd till att explicit införa det implicita led, *seminarierne* [isl. *námstefnunum*], som tidigare förekommit i den danska kontexten. Beträffande exempel (13) är det inte självklart att översättaren menar att *standa fyrir* egentligen ska ses som ekvivalent till *ha*, utan snarare tolkar 'ha' som 'stå bakom', 'stå för'.

När vi nu har vänt på resonemanget och ser på hur *arrangera/arrangere* översätts till isländska framträder ett intressant mönster. Vi kan nämligen se det som Sinclair (2004:142) kallar *colligations*. De isländska ekvivalenterna är här: *standa fyrir* (12 ex.), *halda* (16 ex.), *efna till*, *hafa umsjón*, *skipuleggja*. Mönstret gäller valet mellan *standa fyrir* och *halda*. Det visar sig nämligen att i fallen med

standa fyrir, som i (14), så har källspråket i överväldigande antal ett verb i aktiv diates (10 aktiva verb, 2 passiva verb):

(14) *arrangera / standa fyrir*

De nordiske landene *arrangerer* også en rekke nasjonale seminarer under FNs kvinnekommisjons sesjon.

Norrænu ríkin *standa* einnig hvert fyrir sig *fyrir* nokkrum námstefnum á kvennaráðstefnunni.

I frågan om *halda* (15) gäller det omvända (4 aktiva verb, 12 passiva verb):

(15) *arrangera / halda*

Konferansen som heter 'Energidagan', *ble* i år *arrangert* for tredje gang.

Ráðstefnan, sem bar nafnið 'Orkudagurinn', *var* nú *haldin* í þriðja skipti.

Sinclair brukar framhålla att det endast är genom korpusstudier man kan studera *collogations*, d.v.s. samvariation mellan olika grammatiska former hos sökorden och den kontext de förkommer i. Parallellkorpuser blir därmed ovärderliga för att hitta mönster som detta i valet mellan *standa fyrir* och *halda* som ekvivalent till *arrangera*.

Sammanfattningsvis kan vi säga följande om *standa fyrir*: Korpusexemplen visar att *standa fyrir* är en högfrekvent flerordsenhet som, om den inte kan få lemmastatus i kommande ordböcker, åtminstone bör få en framträdande plats i mikrostrukturen. Vi har också sett att *arrangera/arrangere* oftare översätts med isländska *standa fyrir* i aktiv diates, medan översättningen i passiv diates blir *halda*. I ett lexikon från ett skandinaviskt språk till isländska borde nog brukarna upplysas om det.

Slutord

Huvudsyftet med det här konferensbidraget har varit att visa på möjligheten med (och kanske i någon mån behovet av) att använda parallellkorpuser i tvåspråkig lexikografi. Den korpus som här har använts är naturligtvis för begränsad för fullskalig lexikografisk tillämpning – den borde vara minst 100 gånger så stor och innehålla fler typer av text. Icke desto mindre har exempelstudierna kunnat visa på hittills förbisedda perspektiv på behandlingen av de isländska orden *koma* och *standa* i relation till deras ekvivalenter i de skandinaviska språken.

Principellt har vi sett att ekvivalenspar där flerordsuttryck är inblandade är svåra att undersöka utan en parallellkorpus med möjlighet att upprätta konkordanser av s.k. n-gram. Parallellkorpuser gör det vidare möjligt att analysera de

val som kompetenta översättare har träffat vid arbetet med att finna ekvivalenter i kontext. Vi har hittat ett fall av samvariation mellan verbets diates i källspråket och den ekvivalent som valdes.

Tillgången till parallellkorpusar varierar naturligtvis beroende på vilka språkpar som ska undersökas, men mellan engelska och finska/norska/svenska finns det idag korpusar som används inom språkvetenskap i allmänhet. Det vore en angelägen uppgift att nu ta dem i bruk för lexikografen. Som nämnts ovan pågår arbete med att ta fram en korpus med texter från de fem nordiska språken. När det gäller andra europeiska språkpar har t.ex. Koehn (2005) beskrivit det arbete som pågår med att upprätta en parallellkorpus för elva europeiska språk, inklusive danska, finska och svenska, med material från Europaparlamentet. Det är således rimligt att anta det kan finnas såväl korpusar som parallelltext att tillgå redan nu, och i ännu högre grad om några år. Det som återstår, är att göra dessa resurser tillgängliga, dels genom överenskommelser som gör dem tillgängliga, dels genom att utveckla korpusverktyg som gör det lätt att söka i dem.

LITTERATUR

- Altenberg, Bengt, Karin Aijmer & Mikael Svensson, 2001: The English-Swedish Parallel Corpus (ESPC) : Manual of enlarged version. [tillgänglig elektroniskt: http://sprak.gu.se/digitalAssets/1333/1333431_manual_espc.pdf]. (sept 2011)
- Atkins, B. T. Sue, Adam Kilgarriff & Michael Rundell, 2010: Database of ANalysed Texts of English (DANTE): the NEID database project. I: Dykstra, Anne (ed.), *Proceedings of the Fourteenth EURALEX International Congress*. Leeuwarden. S. 549–556.
- Atkins, B. T. Sue & Michael Rundell, 2008: *The Oxford guide to practical lexicography*, Oxford.
- Borin, Lars & Klas Prütz, 2001: Through a glass darkly: Part of speech distribution in original and translated text. I: Daelemans, Walter, Khalil Sima'an, Jorn Veenstra & Jakub Zavrel (ed.), *Computational linguistics in the Netherlands 2000*. Amsterdam.
- Cheng, Winnie, Chris Greaves & Martin Warren, 2006: From n-gram to skipgram to conngam. I: *International journal of corpus linguistics*, 11. S. 411–433.
- COBUILD = Sinclair, John, 1987: *Collins COBUILD English language dictionary*, London.
- Corréard, Marie-Hélène & Valerie Grundy, 1994: *Oxford-Hachette French dictionary : French-English, English-French*, Oxford; Paris.
- Corréard, Marie-Hélène, Valerie Grundy, Jean-Benoit Ormal-Grenon & Nicholas Rolin, 2007: *Le grand dictionnaire Hachette-Oxford: français-anglais, anglais-français = The Oxford-Hachette French dictionary: French-English, English-French*, Oxford; Paris.

- Friskare nordiska hav, 2010. Nordisk Ministerråd. København. [tillgänglig elektroniskt: <http://www.norden.org/sv/publikationer/publikationer/2010-732>]. (sept 2011)
- Gellerstam, Martin, 1996: Translations as a source for cross-linguistic studies. I: Aijmer, K, B. Altenberg & M. Johansson (ed.), *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*. Lund.
- Hasselgård, Hilde, 2010: Parallel corpora and contrastive studies. I: Xiao, Richard (ed.), *Proceedings of UCCTS 2010; Using Corpora in Contrastive and Translation Studies*. Edge Hill University, Ormskirk.
- Holm, Gösta & Aðalsteinn Davíðsson 1982: *Svensk-isländsk ordbok: Sænsk-íslensk orðabók*, Lund; Reykjavík.
- ISLEX, 2011: ISLEX : en tvärnordisk, flerspråkig onlineordbok med isländska som källspråk och svenska, norskt bokmål, nynorska och danska som målspråk. Reykjavík, Bergen, Göteborg, København. [tillgänglig elektroniskt: <<http://islex.lexis.hi.is/islex/>>; <<http://islex.se/>>]. (sept 2011)
- Johansson, Stig, Jarle Ebeling & Signe Oksefjell, 1999/2002: *English-Norwegian Parallel Corpus: Manual*. [tillgänglig elektroniskt: www.hf.uio.no/ilos/english/services/omc/enpc/ENPCmanual.pdf]. (sept 2011)
- Johansson, Stig & Knut Hofland, 2007: Towards an English-Norwegian parallel corpus. I: Teubert, Wolfgang & Ramesh Krishnamurthy (ed.), *Corpus linguistics: Critical concepts in linguistics*. London ; New York.
- Koehn, Philipp, 2005: *Europarl: a parallel corpus for statistical machine translation*. I: *Proceedings of the tenth Machine Translation Summit*. Phuket, Thailand. [tillgänglig elektroniskt: <http://www.mt-archive.info/MTS-2005-Koehn.pdf>]. (sept 2011)
- McEnery, Tony, Richard Xiao & Yukio Tono, 2006: *Corpus-based language studies : an advanced resource book*, London.
- Rawoens, Gudrun, 2003: *Korpuslingvistik och kontrastiv språkbeskrivning. Ett svensk – nederländskt projekt*. *Tijdschrift voor Skandinavistiek*, 24.
- Scott, Mike, 2008: *Developing WordSmith*. *International Journal of English Studies [Special issue]: Monograph: Software-aided Analysis of Language*, edited by M. Scott, P. Pérez-Paredes & P. Sánchez-Hernández, 8, s.95-106. [tillgänglig elektroniskt: <http://dialnet.unirioja.es/servlet/articulo?codigo=2729155>]. (sept 2011)
- Sinclair, John McHardy, 2004: *Trust the text : language, corpus and discourse*, London.
- SOB = *Språkdata*, Göteborgs universitet, 1986: *Svensk ordbok : [ordens betydelse och användning] : [mer än 100.000 ord och fraser]*, Solna.
- Språkbanken: <http://spraakbanken.gu.se/> (sept 2011)
- Svensén, Bo, 2004: *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*, 2 uppl. Stockholm.

Håkan Jansson

Institutionen för svenska språket, Göteborgs universitet.

hakan.jansson@svenska.gu.se