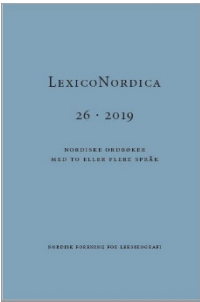


LexicoNordica

| | | |
|------------|---|---|
| Titel: | Korpus och enspråkig ordbok som bas för ny estnisk-finsk ordbok |  |
| Forfatter: | Sofia Björklöf & Saarni Laitinen | |
| Kilde: | LexicoNordica 26, 2019, s. 17-34 | |
| URL: | https://tidsskrift.dk/index.php/lexn/issue/archive | |

© 2019 LexicoNordica og forfatterne

Betingelser for brug af denne artikel

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

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

Korpus och enspråkig ordbok som bas för ny estnisk-finsk ordbok

Sofia Björklöf & Saarni Laitinen

The new Estonian-Finnish online dictionary, *Viro–suomi-sanakirja. Eesti-soome sõnaraamat*, published in January 2019, has been compiled using the new corpus-based Estonian monolingual dictionary. This article examines how editing the dictionary is made easier due to it being based on the corpus-based monolingual dictionary, while also discussing how this editing process can result in some problems. Technology helps in compiling bilingual dictionaries, but do the advantages of using a monolingual dictionary as a basis for a bilingual one outweigh the resulting problems? As there are quite a few emerging issues, it is important that sufficient time and labour resources are given to projects of this sort.

1. Inledning

Den nya ordboken *Viro–suomi-sanakirja. Eesti-soome sõnaraamat* är den första estnisk-finska ordboken som publicerats enbart elektroniskt. Ordboken baserar sig på den nya estniska definitionsordboken (*Eesti keele sõnaraamat* 2019) som för sin del är korpusbaserad. Därmed är den estnisk-finska ordboken också den första i sitt slag som baserar sig på korpus. Syftet med denna artikel är att behandla och diskutera hur detta å ena sidan underlättar redigeringsarbetet och å andra sidan leder till olika problem. Författarna av artikeln har varit redaktörer för den estnisk-finska ordboken.¹

1 Huvudredaktörer Heikki Hurta, Margit Langemets, redaktörer Sofia Björklöf, Anu Haak, Külli Kuusk, Saarni Laitinen, Ilona Paajanen, Ulla Vanhatalo.

2. Bakgrund

Initiativtagaren till skapandet av en ny estnisk-finsk ordbok var Institutet för det estniska språket (Eesti Keele Instituut) i Tallinn. Ordboken redigerades i samarbete med Institutet för de inhemska språken i Helsingfors. Projektet påbörjades i september 2016 och avslutades i december 2018. Ordboken publicerades den 9 januari 2019. Att den publicerades just då beror delvis på att Finland fyllde hundra år 2017 och Estland 2018 och att de båda instituten ville uppmärksamma de här jubileumsåren med det nya verket.

Ordboken har redigerats med redigeringsprogrammet EELex, som skapades vid Institutet för det estniska språket för ungefär tio år sedan. Den tekniska sidan har överlag skötts av esterna. I redaktionen har både estnisk- och finskspråkiga lexikografer ingått, de flesta med målspråket finska som modersmål, vilket gör ordboken till den första estnisk-finska ordboken med en tvåspråkig redaktion.

Ett viktigt skäl till att den aktuella ordboken utarbetades var att de tidigare ordböckerna var föråldrade. Den första estnisk-finska ordboken, *Virolais-suomalainen sanakirja*, sammanställdes för hundra år sedan av en finsk språkvetare, Lauri Kettunen (1917). Den följande, *Eesti-soome sõnaraamat* med 36 000 ord, publicerades i Estland 1972. Det gavs ut ett par nya upplagor av den och år 2015 publicerades en reviderad upplaga med 9 000 nya uppslagsord (*Eesti-soome sõnaraamat*, sammanlagt 45 000 ord). Den nyaste upplagan är trots allt lite föråldrad och också den innehåller bara få exempel, fraser och sammansatta verb. Redaktionen var dessutom enbart estnisk. Det behövdes trots den reviderade upplagan därför en ny modern ordbok vars redaktörer har målspråket finska som modersmål och som skulle innehålla fraser och sammansatta verb som lemman samt rikligt med exempel.

3. Tänkta användare

Traditionella och rentav självklara ordboksanvändare är översättare och studerande, vilket gäller också denna ordbok. Dessutom pendlas det mycket mellan Finland och Estland. Förutom de som arbetar i grannlandet har även turister och andra resande nytta av ordboken. Finskan och estniskan är nära besläktade östersjöfinska språk, som har rikligt med gemensamma ord men samtidigt gott om falska vänner. Både språkens och ländernas närhet bidrar till att det finns många som vill lära sig grannspråket.

Också de som bor mer eller mindre permanent i grannlandet är potentiella användare. I slutet av 2018 bodde det ca 49 700 estniskspråkiga i Finland och estniskan är därmed det näst största invandrarspråket i Finland efter ryskan som talas av 79 200 personer (Statistikcentralen 2019). Det finns en estnisk skola i Helsingfors och några skolor med estniska klasser. År 2011 bodde det ca 2 600 finskspråkiga i Estland och ca 167 000 estländare kunde finska på någon nivå (Statistikaamet 2011a; 2011b). I universitetsstaden Tartu finns det också en finsk skola.

4. Elektronisk publicering

Att den estnisk-finska ordboken publicerades enbart elektroniskt gav en hel del möjligheter. Man kunde ge många exempel eftersom man inte behövde oroa sig för brist på utrymme. Den elektroniska ordboken kunde också genast göras fritt tillgänglig för användarna. En av den elektroniska publiceringens avigsidor är att det var möjligt att ge ut ordboken med bara ca 29 000 kontrollästa artiklar samt ca 10 000 artiklar som inte ännu blivit kontrollästa. Detta tycker vi är mycket problematiskt bl.a. med tanke på ordbokens pålitlighet trots att de okontrollerade artiklarna är tydligt utmärkta. För tillfället uppdateras ordboken fortlöpande men tyvärr tar det en lång tid att kontrolläsa alla artiklar.

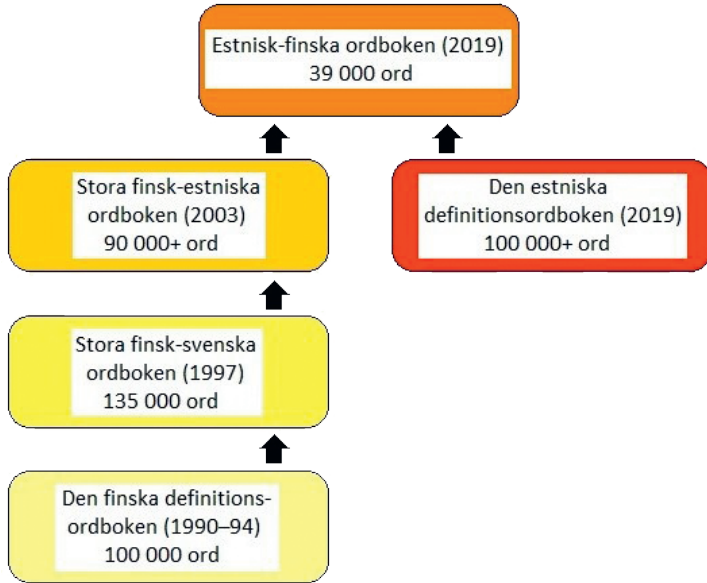
Ordbokens gränssnitt är responsivt, vilket gör den bekväm att använda med mobiler och surfplattor. Folk är nuförtiden vana vid att ha mobilen med sig hela tiden och söka information på nätet.



Figur 1: Startsidan på datorskärmen.

5. Ordböcker som källor

Att den nya estnisk-finska ordboken är korpusbaserad betyder egentligen att den baserar sig på den nya estniska definitionsordboken, som i sin tur är korpusbaserad. Utöver det utgicks det från gamla ordböcker: finska uppslagsord togs automatiskt från den stora finsk-estniska ordboken och de blev potentiella ekvivalenter i den estnisk-finska ordboken. Den stora finsk-estniska ordboken baserar sig på den stora finsk-svenska ordboken, som i sin tur baserar sig på den finska definitionsordbokens version från 1990-talet (se figur 2). Potentiella ekvivalenter krävde mycket redigeringsarbete, eftersom en stor del av dem var oanvändbara och måste gallras bort. Dessutom måste redaktörerna ofta lägga till ekvivalenter. Alla ekvivalenter kom automatiskt in i den första betydelsegruppen och måste därefter placeras i rätt grupp. (Metoden diskuteras närmare i avsnitt 7.)



Figur 2: Andra ordböcker som källa för den estnisk-finska ordboken.

Den stora finsk-estniska ordboken (*Suomi–viro-suursanakirja*) gavs ut som bok år 2003. På nätet har den funnits gratis från och med november 2017. Den är den nyaste och största finsk-estniska ordboken som finns.

6. Korpus

Korpusen som den nya estnisk-finska ordboken baserar sig på är sammanställd vid Institutet för det estniska språket, vi finskspråkiga redaktörer har inte arbetat på den. Korpusen uppdateras regelbundet; år 2017 bestod den av 1,3 miljarder textord. Korpusarbetet påbörjades på 1980-talet vid Tartu universitet och Tartu-korpusen med ca 250 miljoner textord användes som bas för den som nu sammanställs vid Institutet för det estniska språket. Den nuvarande korpusen baserar sig på många slags estniska texter samlade

från internet, bl.a. tidnings- och tidskriftsspråk samt myndighets-språk, men också t.ex. kommentarer till tidningsartiklar. Nästan ingen skönlitteratur finns dock med. De flesta texterna är skrivna från och med början av 1990-talet.

Verktyget för korpusarbetet som görs vid Institutet för det estniska språket är Sketch Engine. Allt som är skrivet på estniska på internet laddas ner; det finns dock några filter, t.ex. blogginlägg filtreras bort.

6.1. Fördelar

När det gäller språkparet estniska och finska är denna ordbok som redan konstaterats den första som är korpusbaserad. Vi upplever att användning av en korpus som bas för en tvåspråkig ordbok har flera goda sidor. För det första är korpusen i fråga väldigt stor, med mer än en miljard textord, och för det andra är den aktuell (vilket också skapar problem, se avsnitt 6.2). För det tredje tror vi att det kan vara nyttigt för användaren att exemplen är från levande språkbruk; texterna i korpusen är skrivna av väldigt många olika individer, också helt vanligt folk. För det fjärde möjliggör korpusen en indelning av orden i frekvensgrupper. De mest frekventa 35 000 orden har delats in i sju frekvensgrupper och de första grupperna finns med i sin helhet i ordboken.

Artiklarna *higine* 'svettig' och *nägu* 'ansikte' fungerar som exempel på levande språkbruk. I satsen *ta oli näost täitsa higine* 'han/hon/hen var helt svettig i ansiktet' finns det estniska ordet *täitsa*, som är den talspråkiga varianten för *täiesti* 'helt'. I estniskan kan ordet *nägu* 'ansikte' också betyda 'min'. I satsen *sul on sihuke pai-poisi nägu* 'du har/uppvisar en oskyldig uppsyn' har man använt det talspråkliga ordet *sihuke* 'sån'. Vi finskspråkiga redaktörer har översatt exemplen så exakt som möjligt och om exemplet innehåller talspråk har vi försökt använda talspråk också i översättningen.

Att blogginlägg, som ofta utgörs av mer talspråksliknande tex-

ter, filtreras bort leder till att korpusen och därmed också ordboken inte innehåller så många talspråkliga ord och uttryck som den kunde. Det är ändå glädjande att det trots allt finns en del sådana ord och uttryck i ordboken för att de har använts i andra slags texter som ingår i korpusen.

6.2. Problem

Det finns naturligtvis också problem med att använda en korpus. Vi har märkt att vissa ord som är vanliga i talet kan saknas eller se mindre frekventa ut eftersom korpusen enbart innehåller skrivna texter. Också gammalt och litterärt ordförråd kan saknas, eftersom de flesta texterna är från år 1990 och framåt. Myndighetsspråk är däremot överrepresenterat.

I den estniska definitionsordboken och därmed också i vår ordbok har man beslutat ordna betydelsegrupperna enligt frekvens, vilket vi anser är problematiskt. Traditionellt har betydelsegrupperna ordnats så att den konkreta och ursprungliga betydelsen kommer först och de bildliga eller annars yngre betydelseerna efter den (Kamppi 2014:152). Enligt Wojciechowska (2012:70) är dock betydelseernas frekvens i korpusen nuförtiden det vanligaste sättet att ordna betydelsegrupperna.

Också de estniska exemplen härstammar från korpusen. De kan vara problematiska med tanke på innehållet, mängden eller placeringen p.g.a. att redaktören som valde dem har förhastat sig. Ett problem för finskspråkiga redaktörer är att exemplen kan ha för knapp kontext, vilket kan försvåra översättandet och t.o.m. leda till felaktiga tolkningar.

6.2.1. Betydelsegrupperna i frekvensordning

När betydelsegrupperna ordnas enligt frekvens i korpusen, kan ordningen ändå bli traditionell tack vare att betydelseernas frekvens i korpusen är sådan. I artikeln *jäine* är ordets första betydelse 'isig'

och andra betydelse 'iskall' (se figur 3). Den bildliga betydelsen av 'iskall' hör således till den andra gruppen och kommer sist. Vi anser att betydelsernas ordning i artikeln *jäine* är bra eftersom det kan vara pedagogiskt att ange betydelserna med den konkreta och ursprungliga först. Problemet är att betydelsegrupperna som sagt alltid ordnas enligt frekvensen, vilket leder till innehållsmässigt olika resultat i olika artiklar. Ordningen ser därför slumpmässig ut och blir inte transparent för användaren. Det finns naturligtvis problem också med andra sätt att ordna betydelsegrupperna, men ofta vore det möjligt att nå ett resultat som är mer koherent och därför också begripligare för användaren (Meer 2002:514).

jäine M <adj>

1 *jääst koosnev või jääga kaetud, jäätunud (mitte veekogu kohta)*
jäinen, iljanteinen

- *õine külm oli maandumisraja jäiseks muutnud halla oli muuttanud kiitoradan jäiseksi / halla oli tehnyt kiitotiestä iljanteisen*

2 *külm kui jää, jääkülm*
jäinen, jäätävä, hyinen, hyytävä

- *kõrvetava päikese käes maitstes jäine kokteil lausa taevalikult pahtavassa auringossa jäinen cocktail maistui suorastaan taivaalliselta*
- *jäist tuult trotsides rühiti edasi tarvottiin eteenpäin hyytävää tuulta uhmaten*

■ *(tunnete, psüühilise sfääri kohta:) äärmiselt ükskõikne, kalk*
jäätävä KUV., *hyytävä* KUV.

- *külaline võeti vastu jäise viisakusega vieras otettiin vastaan jäätävän kohteliaasti*

Figur 3: En fungerande ordning på betydelsegrupperna (utdrag ur artikeln *jäine*).

Uppslagsordet *hiiligus* representerar ett motsatt fall. Ordet har ursprungligen betytt 'glans' i konkret betydelse, men denna betydelse har hamnat sist (se figur 4). Den bildliga betydelsen 'prakt' an-

ges först. Särskilt med tanke på att ordboken i fråga är tvåspråkig, skulle det vara användarvänligt att betydelsegrupperna ordnades traditionellt. Att utgå från korpusen är förstås väldigt konsekvent, men användaren har ingen aning om det eftersom man inte heller nämner denna princip i förordet.

Frekvensordningen har motiverats bl.a. med att användarna oftast antas söka just den vanligaste betydelsen, och att de snabbt vill hitta det de söker. Den frekvensbaserade ordningen snabbar visserligen upp sökandet av enskilda betydelser (vilket enligt kritikerna är det enda som denna slags ordning duger till). Däremot säger frekvensen ingenting om semantiska länkar och mönster. Den frekvensbaserade ordningen gör det svårare för användaren att bekanta sig med och lära sig betydelserna och få en sammanhängande uppfattning om betydelsen. Man borde hellre sträva efter att skapa en sådan struktur som hjälper språkinlärare och andra ordboksanvändare att förstå att betydelserna ofta är förknippade med varandra. Forskningen i språkinlärning visar att ordens betydelser är lättare att begripa, lära sig och komma ihåg när de kognitiva sambanden mellan betydelserna framhävs tydligt. (Meer 2000:420, 2002:510–515 och 2004:808; Kamppi 2014:149)

hiilgus ^M (s)

1 väline sära, toredus, uhkus või kuulsus

loisto, loistokkuus, hohto, hohdokkuus

kuuluisuus

- linna kunagisest hiilgusest on vaid varemед järel kaupungin entisestä loistosta on vain rauniot jäljellä
- kuulus lauljatar suri oma hiilguse tipul tunnettu laulajatar kuoli ollessaan kuuluisuutensa huipulla

2 tugev ühtlane helendus, valguskiirgus või sära

hohde, hohto, loiste, paiste

- keskpäevase päikese pimestav hiilgus keskipäivän auringon sokaiseva loiste

Figur 4: En problematisk ordning på betydelsegrupperna (utdrag ur artikeln hiilgus).

6.2.2. Problematiska exempel och exempel med för lite kontext

Vissa exempel kan bli mindre lyckade bland annat på grund av tempus eller brist på kontext. Det sammansatta verbet **alla heitma** betyder bl.a. 'underordna; kuva, förtrycka'. I artikeln fanns det ursprungligen ett exempel som löd 'esterna kuvas genom att röva och beskatta', d.v.s. i presens. Det verkade underligt med tanke på att Estland är ett modernt europeiskt land och att uppslagsordets betydelse är enbart konkret, inte bildlig. Till all lycka gick det att hitta satsen på nätet. Kontexten var en lång text om Estlands historia, som hade skrivits i presens och där den aktuella satsen ingår i ett avsnitt som skildrar 1200-talet. För att göra satsen förståelig ändrades tempuset till imperfekt: 'esterna kuvades genom att röva och beskatta'. Texten hade skrivits 1922 och lagts ut på nätet 2009, vilket visar att också äldre texter kan förekomma sporadiskt i korpusen.

7. Enspråkig ordbok som bas

Den nya estnisk-finska ordboken baserar sig som sagt på den nya estniska definitionsordboken. Definitionsordboken gjordes först – fast den publicerades lite senare, 14 februari 2019 – och i slutfasen redigerades de båda ordböckerna parallellt. Den nya estniska definitionsordboken har publicerats endast som en del i en ny portal som samlar ihop flera resurser. Såvitt vi vet är det sällsynt att redigera en tvåspråkig ordbok med en enspråkig som bas. I vårt fall tog man den enspråkiga ordbokens databas och lade till fält för finska ordförklaringar, ekvivalenter och översättning av exempel. Man tog också bort några fält, t.ex. fältet för det estniska uppslagsordets etymologi.

7.1. Fördelar

Att använda en enspråkig ordbok som bas för en tvåspråkig ordbok har en del goda sidor. Det gör redaktionsarbetet snabbare eftersom basen är färdig och omfattande. Det finns över 100 000 uppslagsord i den estniska definitionsordboken och vi hade i princip möjligheten att göra en lika stor tvåspråkig ordbok. Uppslagsordets betydelsegrupper har var sin ordförklaring, vilket inte är vanligt i tvåspråkiga ordböcker. Ordförklaringarna underlättar redaktionsarbetet och kan också hjälpa ordbokens användare. Förklaringarna var till stor nytta när vi diskuterade några krångliga ord med den estniska redaktionen och tidvis kunde vi även skriva om förklaringarna i samarbete med esterna.

7.2. Problem

Uppslagsordets betydelser har delats i grupper med tanke på den enspråkiga definitionsordbokens behov, och därför kan det finnas antingen för många eller för få betydelsegrupper för den tvåspråkiga ordboken. Exempelen är inte heller valda för en tvåspråkig ordbok. Allt detta kräver uppmärksamhet och bearbetningen tar tid.

7.2.1. Uppdelning

Artikeln *filosoofia* är ett exempel på för fin uppdelning för den tvåspråkiga ordboken (se figur 5). Det finns fyra betydelsegrupper fast det skulle räcka med färre. Till exempel består artikeln ”filosofi” i den gamla estnisk-finska ordboken endast av uppslagsordet *filosoofia* och den finska ekvivalenten *filosofia*. Enligt Svensén (2000:153) ska grupperingen av betydelser i en tvåspråkig ordbok inte återspegla källspråkets betydelsestruktur. På grund av att den estnisk-finska ordboken baserar sig på den enspråkiga ordboken har det varit svårt att följa denna princip.

filosoofia TM <s>

teadus, mis uurib maailma ja mõtlemise kõige üldisemaid seaduspärasid

filosofia

- *mingi koolkonna, teatava autori filosoofiline õpetus*

filosofia

- Aristotelese filosoofia **Aristoteleen filosoofia**

- *mingi valdkonna kõige üldisemad seaduspärad, nende uurimine ja rakendamine*

filosofia

- juhtimisfilosoofia **johtamisfilosofia**
- haridusfilosoofia **koulutusfilosofia**; kasvatusfilosofia

- *(mingi valdkonna tavade ja kommete kohta)*
(toimintatavoista, perinteistä tms.)

filosofia

- meie veinitegemise filosoofia on täiesti erinev naabrite omast
viininteon filosofiamme on aivan erilainen kuin naapureilla

Figur 5: För fin uppdelning i betydelsegrupper.

Det följande exemplet handlar också om för fin uppdelning, men i det här fallet har artikelns struktur korrigerats (se figur 6). Den övre bilden är från den enspråkiga ordboken och där ser man att det finns två betydelsegrupper. På den nedre bilden – från den tvåspråkiga ordboken – syns det att grupperna har förenats. Det estniska ordet *jalam* betyder 'fot' av t.ex. ett berg eller torn. Esterna har uppdelat betydelsen så att den första gruppen består av geografiska former och den andra av sådant som människan har byggt. Det kändes vettigt att förena grupperna, eftersom betydelseorna ändå är så nära varandra och ekvivalenten är densamma för båda grupperna.

{F2} jalam <s> <-i 2>

KOHT_LOODUS {OSA} **GEOGR** mäe, künka vms nõlva alumine osa ja selle lähem ümbrus

- *Vulkaani jalamile rajati rahvuspark.*

- **KOHT_HOONE {OSA}** samba, torni vm ehitise alumine osa või alus
- *Tuletorni jalamil peetud pidulik tseremoonia.*

F2 jalam <s>

ab GEOGR mäe, künka vms nõlva alumine osa ja selle lähem ümbrus; samba, torni vni ehitise alumine osa või alus

jonkin ympäristöstään kohoavan vierusta, lähiympäristö

juuri [> 3. vuoren tms. jalam <]

- *vulkaani jalamile rajati rahvuspark tulivuoren juurelle perustettiin kansallispuisto*
- *tuletorni jalamil peetud pidulik tseremoonia majakan juurella pidetty juhlallinen seremonia*

Figur 6: För fin uppdelning i betydelsegrupper och uppdelningen efter bearbetning (skärmbild från redigeringsprogrammet EELex).

Ibland är uppdelningen däremot inte tillräckligt fin. Uppslagsordet **algdokument** 'originalhandling' har endast en betydelsegrupp i den enspråkiga ordboken trots att den har två ordförklaringar åtskilda av ett semikolon (se figur 7). På finska har vi inget ord som skulle motsvara den första ordförklaringen. När ekvivalenten saknas ges det en ordförklaring på finska och ekvivalenten ersätts med ett understreck. I det här fallet har den ursprungliga betydelsegruppen dessutom delats upp, eftersom det finns en ekvivalent som motsvarar den andra förklaringen.

{F4} alg+dokument

ESE_SEMIO majandustehingu kohta koostatud originaaldokument (nt pangaväljavõtte, ostu- või müügiarve, töötasu arvestamise dokumendid); muu originaaldokument

▫ *Raamatupidamise algdokumendid.*

F4 alg+dokument

majandustehingu kohta koostatud originaaldokument (nt pangaväljavõtte, ostu- või müügiarve, töötasu arvestamise dokumendid) /SBjorklof/

rahankäyttöön liittyvä alkuperäinen asiakirja (esim. tilitosite, lasku, palkanmaksuun liittyvät paperit)

- **muu originaaldokument** /SBjorklof/
alkuperäinen asiakirja [> **aito** [1*F]: algdokument <],
alkuperäis+asiakirja

Figur 7: Inte tillräckligt fin uppdelning i betydelsegrupper och uppdelningen efter bearbetning (skärmbild från redigeringsprogrammet EELex).

7.2.2. Exempel inte valda för en tvåspråkig ordbok

Det framgår av både innehållet och mängden av exempel i den enspråkiga ordboken att de inte är valda för en tvåspråkig ordbok. I vissa fall finns det för många exempel, som inte ger någon ny information och som inte är kontrastivt intressanta (se figur 8). Sådana har vi lämnat utan översättning och de syns inte heller i ordboken. I artikeln **horoskoop** 'horoskop' är det enda exemplet som översatts "enligt horoskopet i tidskriften ligger det svartsjuka i luften".

F2 horoskoop <s>
 astroloogiline tabel või skeem saatuse või sündmuste
 ennustamiseks taevakehade seisu järgi kellegi sünnihetkel;
 ennustus selle alusel
horoskoop [> horoskoop <]
ab• *horoskoobi järgi on tema salapahe laiskus ##*
 ▪ *ajakirja horoskoop ütleb, et õhus on armukadedust lehden*
horoskoopin mukaan ilmaassa on mustasukkaisuutta
ab• *sünnihoroskoop (ls) ##*
ab• *isikuhoroskoop (ls) ##*

Figur 8: För många exempel för en tvåspråkig ordbok (skärmbild från redigeringsprogrammet EELex).

I andra fall är exemplen för få. Till exempel beträffande uppslagsordet **poleks** 'skulle inte vara (eller ngt annat verb)' räcker det med endast två exempel i definitionsordboken (se figur 9). Ordet har dock många finska ekvivalenter p.g.a. att det i finskan finns ett böjligt nekningsverb som saknas i estniskan. Av den anledningen behövs det också flera exempel, helst ett i varje person. Därför sökte vi lämpliga exempel i ordbokens databas och kompletterade artikeln med dem (se figur 10).

poleks

VERB_ABSTR • VERB_GRAMMATILINE ei oleks

- *Kellel siis muresid poleks.*
- *Poleks kunagi uskunud!*

Figur 9: Artikeln **poleks** i den enspråkiga ordboken: för få exempel för en tvåspråkig ordbok (skärmbild från redigeringsprogrammet EELex).

poleks

ei oleks

en olisi yks.1.p,

et olisi yks.2.p,

ei olisi yks.3.p;

emme olisi mon.1.p,

ette olisi mon.2.p,

eivät olisi mon.3.p

- *eladeski poleks arvanud, et sellist tööd tegema hakkas en olisi ikinä uskonut, että alan tehdä tällaista työtä*
- *siit ma oleks alla sadanud nagu purujutt, kui sa poleks taibanud mind kinni hoida olisin pudonnut tästä päästä pahkaa, ellet olisi huomannut pitää minusta kiinni*
- *kellel siis muresid poleks kenelläpä ei olisi huolia*
- *matkajad asusid teele varahommikul, sest päevakuumuses poleks nad seda suutnud retkeilijät lähtivät matkaan aamuhäälisellä, sillä päivän kuumuudessa he eivät olisi siihen kyenneet*
- *küllap oleksid Balti riigid tahtnud saada iseseisvaks ka juhul, kui Nõukogude Liit poleks olnud diktatuur eivätköhän Baltian maat olisi halunneet itsenäistyä myös siinä tapauksessa, että Neuvostoliitto ei olisi ollut diktatuuri*
- *Poleks kunagi uskunud! Ei olisi uskonut!; En olisi ikinä uskonut!*
- *kui tark ta ka poleks, kõike ta ikka ei tea niin viisas kuin hän onkin, kaikkea hän ei kuitenkaan tiedä*

Figur 10: Artikeln **poleks** i den tvåspråkiga ordboken: exempel har lagts till (skärmbild från redigeringsprogrammet EELex).

8. Sammanfattning

I denna artikel har vi behandlat några exempel som illustrerar de fördelar och problem som har uppstått när en korpus och en enspråkig ordbok använts som bas för den nya estnisk-finska ordboken. Till korpusens fördelar hör att den är omfattande, innehåller mycket nutidsspråk och gör det möjligt att dela orden i frekvensgrupper. Det som vi anser är problematiskt med korpusen är att betydelsegrupperna har ordnats enligt betydelsernas frekvens i korpusen. Dessutom har betydelserna delats i grupper med tanke på den enspråkiga ordboken. Att exemplen härstammar från korpusen gör arbetet snabbare men skapar samtidigt svårigheter eftersom exemplen är valda enligt den enspråkiga ordbokens behov. Att arbetet blir snabbare är överlag en av de största fördelarna med att använda korpus och enspråkig ordbok som bas. Å andra sidan måste redaktörerna göra en hel del förändringar som de hela tiden ska vara medvetna om. Bearbetningen tar tid, och vissa förändringar förblir tyvärr ogjorda p.g.a. tidsbrist.

Teknologin gör det lättare att producera tvåspråkiga ordböcker, men är fördelarna med att använda en enspråkig ordbok som utgångspunkt så stora att de väger upp nackdelarna? Med tanke på ovannämnda problem borde man se till att sådana här projekt hade tillräckligt med tid och andra resurser.

Litteratur

Ordböcker

- Eesti keele sõnaraamat* 2019. Margit Langemets m.fl. (red.). Tallinn: Eesti Keele Instituut. <<https://sonaveeb.ee>> (januari 2019).
- Eesti-soome sõnaraamat. Eestiläis-suomalainen sanakirja* (1972). Valdek Pall m.fl. (red.). Tallinn: Valgus.

- Eesti-soome sõnaraamat. Viro–suomi-sanakirja* (2015). Sirje Laidre m.fl. (red.). Tallinn: Valgus.
- Kettunen, Lauri (1917): *Virolais-suomalainen sanakirja. Eesti-soome sõnaraamat*. Helsinki: Otava.
- Suomen kielen perussanakirja* (1990–1994). Risto Haarala m.fl. (red.). Helsinki: Kotimaisten kielten tutkimuskeskus / Valtion painatuskeskus.
- Suomi–viro-suursanakirja. Soome-eesti suursõnaraamat* (2003). Valdek Pall m.fl. (red.). Tallinn/Helsinki: Eesti Keele Instituut / Kotimaisten kielten keskus.
- Suuri suomi–ruotsi-sanakirja. Stora finsk-svenska ordboken* (1997). Birgitta Romppanen m.fl. (red.). Porvoo/Helsinki/Juva: WSOY.
- Viro–suomi-sanakirja. Eesti-soome sõnaraamat* (2019). Verkko-sanakirja [webbordbok]. Heikki Hurttu m.fl. (red.). Tallinn/Helsinki: Eesti Keele Instituut / Kotimaisten kielten keskus. <www.eki.ee/dict/efi/> (januari 2019).

Annan litteratur

- Kamppi, Nina (2014): Metonymiset merkitykset sanakirjoissa [recension av Wojciechowska (2012)]. I: *Virittäjä* 118, 148–153.
- Meer, Geart van der (2000): Core, subsense and the *New Oxford Dictionary of English* (NODE). On how meanings hang together, and not separately. I: *Euralex 2000 Proceedings*. Stuttgart: Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart, 419–431. <www.euralex.org/elx_proceedings/Euralex2000/> (januari 2019).
- Meer, Geart van der (2002): Dictionary entry and access trying to see trees and woods. I: *Euralex 2002 Proceedings*. Copenhagen: Center for Sprogteknologi, 509–519. <www.euralex.org/elx_proceedings/Euralex2002/> (januari 2019).
- Meer, Geart van der (2004): On defining: polysemy, core meanings and ‘great simplicity’. I: *Euralex 2004 Proceedings*. Lorient: Uni-

versité de Bretagne Sud, 807–816. <www.euralex.org/elx_proceedings/Euralex2004/> (januari 2019).

Statistikaamet (2011a): Statistika andmebaas. RLo431: Rahvastik emakeele ja soo järgi, 31. detsember 2011. Tallinn: Statistikaamet. <<http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=RL415>> (mars 2019).

Statistikaamet (2011b): Statistika andmebaas. RLo438: Rahvastik võõrkeelte oskuse, vanuserühma ja elukoha järgi, 31. detsember 2011. Tallinn: Statistikaamet. <<http://andmebaas.stat.ee/Index.aspx?lang=et&DataSetCode=RL415>> (mars 2019).

Statistikcentralen (2019): Finlands officiella statistik (FOS): Befolkning efter språk 31.12.[2018]. Helsingfors: Statistikcentralen. <https://www.tilastokeskus.fi/tup/suoluk/suoluk_vaesto_sv.html#befolkningeftersprak31.12> (mars 2019).

Svensén, Bo (2000): *Handbok i lexikografi*. Stockholm: Norstedts Ordbok.

Wojciechowska, Sylwia (2012): *Conceptual metonymy and lexicographic representation*. Frankfurt am Main: Peter Lang.

Sofia Björklöf
doktorand, FM
Helsingfors universitet
PL 24, Unionsgatan 40
FI-00014 Helsingfors universitet
sofia.bjorklof@helsinki.fi

Saarni Laitinen
forskningsassistent, HuK
Helsingfors universitet
PL 24, Unionsgatan 40
FI-00014 Helsingfors universitet
saarni.laitinen@helsinki.fi