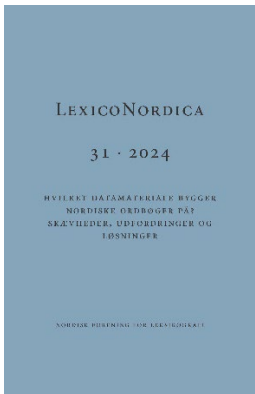


LexicoNordica

Titel:	Skevheter och utmaningar i ordboksbasen för fyra enspråkiga finländska ordböcker	
Forfatter:	Tarja Riitta Heinonen & Caroline Sandström	
Kilde:	LexicoNordica 31, 2024, s. 81-101	
URL:	https://tidsskrift.dk/lexn/issue/archive	

© 2024 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.

Skevheter och utmaningar i ordboksbasen för fyra enspråkiga finländska ordböcker

Tarja Riitta Heinonen & Caroline Sandström

This article gives an overview of what kind of data the four monolingual dictionaries compiled at the Institute for the Languages of Finland are based on and what kind of data-related challenges and biases we have encountered. Major issues that generally affect all dictionaries include the reliability and representativity of the source material. A further concern for us is how to react to the passage of time, since all these dictionary projects trace back over a hundred years. We present a few examples on how we update the dictionary of contemporary language, taking into account that it is partially based on older dictionaries. We also show that current language usage impacts choices made in historical and dialect dictionaries.

1. Inledning

På Institutet för de inhemska språken i Finland redigeras fyra stora enspråkiga ordböcker: en finsk dialektordbok *Suomen murteiden sanakirja* (SMS), en svensk dialektordbok *Ordbok över Finlands svenska folkmål* (FO), en ordbok över äldre finskt skriftspråk *Vanhän kirjasuomen sanakirja* (VKS) och en nufinsk deskriptiv och preskriptiv ordbok *Kielitoimiston sanakirja* (KS). De tre dokumenterande, historiska ordböckerna SMS, FO och VKS bygger på omfattande arkivmaterial. De är alla typiska flergenerationsordböcker (Vikør 1999). Redigeringen inleddes på 1950–1960-talen och pågår fortfarande. Med tiden har redaktörer avlöst varandra och redaktionella principer och praxis har justerats. Teknisk utveckling och digitalisering har gett nya redigeringsredskap, men också medfört vissa utmaningar, i synnerhet vid övergången till digital publicering under 2010-talet. Den nufinska ordboken KS grun-

dar sig i sin tur delvis på tidigare materialbaserade samtidsordböcker.

Samtliga har karaktären av nationella storprojekt och deras tillkomsthistoria sträcker sig långt tillbaka i tiden. De finska ordböckerna är ett resultat av det ordboksprogram för tre stora vetenskapliga finska ordböcker som språkforskaren och politikern E.N. Setälä introducerade 1896. Programmet skulle omfatta en allmänfinsk ordbok, en ordbok över äldre finskt skriftspråk och en ordbok över folkmålen (Setälä 1896; Vilppula 1999:407f.; Onikki-Rantajääskö 2011:544–547; Ruppel & Sandström 2014:148). När det gäller de svenska dialekterna har lexikografin likaså av hävd varit en del av det nationella projektet för svenskan i Finland (Ahlbäck 1946:1; Sandström 2013:92f. och 2016:80f.).

En allmängiltig målsättning för vetenskapliga ordböcker är att ordboksbasen för beskrivningen i en ordbok bör bestå av ett så representativt, heltäckande och pålitligt materialurval som möjligt. Inledningsvis kan vi också konstatera att redigeringen av nuspråkliga ordböcker i högre grad påverkas av språkliga och samhällsliga förändringar än ordböcker som dokumenterar äldre språkskeden.

I denna artikel gör vi jämförelser och fokuserar på utmaningar i ordboksbaserna för de fyra ordböckerna. I avsnitt 2 ges först en kort bakgrund till ordboksbas som begrepp. I avsnitt 3 presenteras ordböckerna, deras omfång och typen av material som de bygger på. I avsnitt 4 diskuterar vi utmaningar som framträder vid redigeringen av dessa ordböcker. Vi tar fram skillnader och likheter i de problem som redaktörerna möter, men fokuserar också på hur redaktionerna har löst och hanterat skevheter och utmaningar i materialen, t.ex. när det gäller representativitet och pålitlighet i förhållande till språkliga och samhällsliga förändringar. Huvudpunkterna i vår artikel sammanfattas i avsnitt 5.

2. Ordboksbas och källor för ordböcker

I *Nordisk leksikografisk ordbok* (NLO 1997:199) definieras *ordboksbas* som de källor som valts ut för en ordbok och enligt Svensén (2004:50) som ”[d]e källor som används för utarbetandet av en ordbok”. Både Svensén och NLO gör en indelning i primärkällor och sekundärkällor. Primärkällorna består av autentiskt språkligt material, antingen muntligt eller skriftligt. De kan vara av olika slag: beläggsamlingar, korpusar eller material som man kommer åt genom introspektion (lexikografens egen kompetens eller informantdata). Sekundärkällorna utgörs av existerande beskrivningar av språket, såsom andra ordböcker, grammatikor eller specialundersökningar.

Svensén (2004:51) framhåller att då data tas fram genom introspektion bör denna metod användas med försiktighet. Vid introspektion kan det egna språkbruket och bedömningen av språkliga data påverkas av mer eller mindre medvetna föreställningar om språkliga normer. Om en lexikograf producerar egna språkprov är det därför viktigt att kontrollera språkprovet mot autentiska språkprov från andra källor. I diskussionen om datainsamling och dataurval skärskådar Svensén (2004:52–57) olika metoder att samla in material till en ordbok från primärkällor och konstaterar att beläggsamlingar tidigare var det enda säkra sättet, men numera är korpusmetoden den vanligaste. Svensén betonar att det finns för- och nackdelar med bägge metoderna, och att korpusmetoden och beläggmetoden kan komplettera varandra. En korpus ger information om det typiska språkbruket och den ger även möjlighet till kvantifiering. I en beläggsamling, som tillkommit genom urval, betonas ofta lågfrekventa och ovanliga språkliga företeelser.

När det gäller de fyra finländska ordböckerna finns det skillnader mellan ordboksbaserna för en dialektordbok eller en historisk ordbok som VKS, och en modern allmänspråklig ordbok. De två dialektordböckerna bygger på traditionella beläggsamlingar som

sammanställts för ordböckerna. Ordboksbasen för den historiska ordboken VKS utgör en kombination av en äldre beläggsamling och en korpus. I flera omgångar har redaktionen för VKS breddat sin ordboksbas och utökat materialet genom en textkorpus där texterna gjorts sökbara. Den allmänfinska ordboken KS bygger dels på två äldre samtidsordböcker, dels utnyttjas såväl introspektion som kontroller i tillgängliga korpusar och varierande primär- och sekundärkällor.

3. Översikt över de fyra ordböckerna SMS, FO, VKS och KS: bakgrund, material, publicering

Det finns många likheter i bakgrund och villkor för de enspråkiga finländska ordböckerna på Institutet för de inhemska språken. Gemensamt för dem alla är att materialet huvudsakligen var färdigt insamlat och att redigeringen hade påbörjats, innan föregångaren Forskningscentralen för de inhemska språken (1976–), senare Institutet för de inhemska språken (2012–), existerade. Likheter framträder speciellt i fråga om de två finländska dialektordböckerna, som är monumentala nationella projekt med uppgift att dokumentera, bevara och publicera det språkliga kulturarvet för nationalspråken finska och svenska.

3.1. Den finska dialektordboken SMS

Den finska dialektordboken SMS (*Suomen murteiden sanakirja*) grundar sig på en enorm beläggsamling som sammanställts i ordbokens arkiv (*Suomen murteiden sana-arkisto*). Arkivet består av originalsamlingar och omfattar ca 8 miljoner ordsedlar med uppgifter om ca 400 000 ord. Samlingarna utgörs dels av svar på frågelistor, dels av uppteckningar gjorda av både vetenskapligt utbildade stipendiater och dialektkunniga lekmän. Materialet

samlades in 1890–1973 i Finland, men också i finsktalande områden utanför Finlands gränser, bl.a. i Värmland, Tornedalen och Finnmarken i Norge (Vilppula 1999:407). Det sammanställdes för redigering 1924–1974. Redigeringen av SMS inleddes på 1960-talet och ordboken publicerades i tryck 1985–2008 i åtta band (av planerade 20 band). Nätpublicering av SMS påbörjades 2012, och 2024 har avsnittet *a-noukkia* utgivits. Omkring 187 500 av planerade 350 000 artiklar och dessutom över 74 000 utbredningskartor har blivit publicerade.

3.2. Den svenska dialektordboken FO

Ordbok över Finlands svenska folkmål (FO) är en dialektordbok som behandlar dialektmaterial insamlat under perioden 1860–1970 i regionerna Österbotten, Satakunta, Åland, Åboland och Nyland inom nuvarande Finlands gränser. Ordboken redigeras utgående från en beläggsamling som omfattar omkring 1 miljon ordsedlar, som excerperats ur dialektologiska och etnologiska uppteckningar, bandinspelningar, dialektologiska avhandlingar och också ur texter av dialektkunniga lekmän. I materialet ingår två äldre dialektordböcker utgivna av Herman Vendell (1904–1907) och V.E.V. Wessman (1925–1932). Beläggsamlingen FOreg. (Register för FO) sammanställdes 1945–1980. Redigeringen inleddes på 1960-talet och ordboken utgavs 1976–2007 i fyra band (av planerade sju band). Nätpublicering av FO påbörjades 2013, och 2024 har avsnittet *a-rättvisa*, ca 79 500 av planerade 120 000 artiklar publicerats.

3.3. Ordbok över äldre finskt skriftspråk VKS

Ordboken över äldre finskt skriftspråk VKS (*Vanhan kirjasuomen sanakirja*) är en historisk ordbok som behandlar över 1 500 skriftliga finska källor utgivna 1543–1810. Ordboksbasen består av en

beläggsamling på ca 0,5 miljoner ordsedlar och en korpus på ca 3,5 miljoner ord. Beläggsamlingen sammanställdes under perioden 1930–1980 och korpusen under 1990-talet och början av 2000-talet. Redigeringen inleddes under 1950-talet. VKS utkom i tryck 1985–1994 i två band (*a–k*). Från 2014 publiceras den som nätordbok. Våren 2024 hade avsnittet *a–puhdistaminen*, 46 400 av planerade 80 000 artiklar, blivit publicerade på nätet.

3.4. Den nufinska ordboken KS

Ordboken KS (*Kielitoimiston sanakirja* 'Språkbyråns ordbok') är en deskriptiv och preskriptiv allmänspråklig ordbok. Den är utgiven i sin helhet *a–ö*, och omfattar drygt 100 000 lemman. KS uppdateras kontinuerligt och den har tidigare utkommit i olika utgåvor. KS grundar sig via mellansteget *Suomen kielen perussanakirja* 'Finsk basordbok' (PS) på den äldre, betydligt mer omfattande, *Nykysuomen sanakirja* 'Nufinsk ordbok' (NS) med drygt 200 000 lemman.

KS är alltså en direkt fortsättning av PS, som publicerades i tryck under åren 1990–1994, och i slutet av 1990-talet som CD-rom. Ordboken utgavs första gången under det nuvarande namnet *Kielitoimiston sanakirja* som betald nätordbok och CD-rom 2004. Den utkom också i tryck i två upplagor 2006 och 2012. Från 2014 publiceras KS som en öppet tillgänglig nätordbok.

Arbetet med NS hade påbörjats 1929 och ordboken utkom i sex delar under perioden 1951–1961. NS grundar sig på ett omfattande arkivmaterial (ca 4,5 miljoner ordsedlar med uppgifter om ca 850 000 ord). Materialet, som sammanställdes på 1930–1950-talen, excerperades ur skönlitteratur och facklitteratur utgiven under perioden 1880–1950, men också ur äldre språkligt och kulturellt epokgörande verk som Kalevala och Aleksis Kivis produktion (Haarala 1999:151–155; Ruppel & Sandström 2014:151–153).

Då arbetet med PS inleddes utnyttjades den äldre ordboken NS

som bas. Syftet med PS var att den skulle vara en aktuell, kompaktare version av NS (Haarala 1994). Vid revideringen rensades över hälften av de gamla uppslagsorden ut och i stället togs nya lemman in i PS. Också själva ordartiklarna förkortades och omarbetades.

Ordboksbasen för KS består alltså till stor del av samma ordboksbas som föregångarna, vilket fortfarande skiner igenom i dagens KS. Över 70 procent av uppslagsorden är desamma som i NS, även om deras relativa andel har minskat då nya lemman tagits in.

Det nya materialet för KS kan beskrivas som öppet; underlaget består av en elektronisk beläggsamling på 260 000 ord, som redaktörerna vid sidan av det övriga redaktionella arbetet fortlöpande kompletterar med nya ord och uttryck ur varierande källor. Att ordboksbasen är öppen innebär bl.a. att redaktionen vid valet av nya lemman måste ta ställning till vilka ord som är allmänspråkliga respektive termer inom ett visst specialområde. Eftersom KS också har rollen av en preskriptiv ordbok ingår drygt 1 000 språkvårdande rekommendationer. Vid uppdateringar ses dessa språkliga rekommendationer över i samarbete med de finska språkvårdarna på Språkinstitutet.

4. Skevheter och problem i ordboksbaserna

Som framgått ovan har de fyra ordböckerna en lång historia. Dialektordböckerna beskriver huvudsakligen de traditionella finska respektive svenska dialekterna som talades på landsbygden under 1900-talets första decennier. De är inte inriktade på dagens regionala, utjämnade dialekter, vilket skulle kräva insamling av ett nytt uppdaterat material. För VKS har insamlingen av källmaterial genomgått många olika skeden, men eftersom ordboken beskriver det äldre finska skriftspråket har det bevarade källmaterialet inte förändrats över tid. Däremot har sättet att tillgodogöra sig materialet inom redaktionen utvecklats under årens lopp.

Till skillnad från de historiska, dokumenterande ordböckerna förändras föremålet för beskrivningen från dag till dag i en modern allmänspråklig ordbok med syfte att beskriva och normera det aktuella språkbruket. Det är därför en stor utmaning för KS att hålla sig i takt med tiden; se vidare diskussion om kravet på ständig uppdatering av KS i avsnitt 4.2.

4.1. Representativitet och materialets pålitlighet i äldre ordböcker

Syftet med en dialektordbok eller en historisk ordbok är att beskriva och dokumentera dialekterna eller ett äldre språkskede och därför ges ofta rikligt med autentiska exempel i ordboksartiklarna. Målsättningen för en allmänspråklig ordbok är däremot att beskriva och exemplifiera det samtida språket för språkbrukarna. KS bygger på sina föregångare, och den äldre av dem, NS, grundar sig på en stor beläggsamling. I NS förekommer ofta dels autentiska exempel, dels förkortade exempel som belyser rektion och typiska kollokationer. I KS har exemplen för det mesta kraftigt nedkortats och autentiska exempel är sällsynta. I de få fall autentiska exempel ges är de i regel hämtade ur vissa unika källor såsom Bibeln.

Ett ansevärt antal lemman och exempel i NS är excerperade ur skönlitterära texter. I källorna för ordboken finns många avledningar och parallellformer och för finskan typiska s.k. deskriptiva eller onomatopoetiska ord. I (1) återges ur NS några prov på lemman som vid redigeringen av PS rensades ut, därför att de blivit gammaldags, ovanliga eller verkar konstiga i dagens språkbruk. Exempelorden i (1) har alla ett skönlitterärt sammanhang.

(1) **häkättää** (onomatopoetiskt) skratta på ett visst sätt (Ilmari Kianto)

hälläkkä (deskriptivt) om flicka: 'toka, fjolla' (Aleksis Kivi)

hämäröittä (sällsynt avledning av finska *hämärä* 'dunkel, skum; suddig') 'göra dunkel' (Otto Manninen)

himerä, himeä (sällsynt) jfr finska *himmeä* och *hämärä* 'dunkel, skum; dämpad; oklar' (F.E. Sillanpää, Aleksis Kivi) (våra egna övers.)

Inom äldre finsk lexikografi har litterära källor haft en stark ställning och skönlitterära exempel har fått stort utrymme. I NS återges de som citat med en namngiven författare som källa. En orsak till att skönlitteraturen utnyttjades var att de finska ordböckerna var knutna till det nationella projektet att höja det finska språkets status, därför sågs skönlitterära författare som förebilder. Från 1960-talet framåt har sakprosa och framför allt korpusar bestående av presstexter fått större roll vid uppbyggnaden av ordboksbaser för allmänspråkliga ordböcker (jfr Svensén 2004:54–59; Forsberg & Holmer 2024).

I det följande vill vi lyfta fram representativiteten när det gäller olika områden inom språkbruket. Vilka skevheter och lemmaluckor påträffas i ordböckerna? För dialektordböckernas del har det sedan gammalt varit mycket vanligt att förutom att beskriva själva ordförrådet också ge etnologiska, s.k. sakliga, materiella, uppgifter om orden. Detta kan ses som ett naturligt element, typiskt för dialektordböckerna, men det har också lett till en slagsida, där ingående beskrivningen av bondekulturen kraftigt dominerat på bekostnad av andra levnadssätt och kulturer. I Lilja (1996) lyfts denna snedvridning till förmån för bondekulturen fram med utgångspunkt i en undersökning av traditionssamlade arkiv i Uppsala 1914–1945. Lilja (1996:26) konstaterar att "[d]en folkkultur som undersöktes [...] nästan uteslutande [har] varit liktydig med landsbygdens och böndernas". Detta gäller även för dialektordböckerna SMS och FO.

Det kan ibland vara utmanande att avgöra vad representativitet innebär när det gäller en dialektordbok. Utgående från arbetet

med FO har redaktörerna lagt märke till att uppgifter om vissa ord eller betydelser kan saknas i ordboksbasen från ett område eller en socken. Orsaken kan vara att insamlaren förbisett eller missat ordet, men det kan också vara så att ordet eller betydelsen inte används i dialekten i detta område. Det är därför viktigt att redaktören tolkar uppgifterna om ordets utbredning noggrant och korrekt. När det gäller materialet för FO har redaktionen också sett en geografisk snedvridning i att det insamlade materialet från stora delar av Österbotten samlats in av vetenskapligt utbildade insamlare i mitten och slutet av 1900-talet. Uppgifterna från Österbotten tenderar att vara mer detaljerade, pålitliga och mångsidiga i fråga om uttal, betydelser och exempel än i materialet från Nyland, som ofta är äldre och insamlat med mindre systematik i slutet av 1800-talet.

En slagsida för dialektordböckerna kan också vara manér hos vissa kreativa upptecknare. Det kan i FO:s material handla om rena skrivbordskonstruktioner av produktiva partikelverb eller flerledade sammansättningar, om ett stort antal pejorativa benämningar på kvinnor eller ovanligt många ord för sex och köns-
umgänge från samma orter. I de fall när det finns återkommande mönster där ord inom vissa områden är överrepresenterade i samlingar från en viss ort är det viktigt att redaktörerna känner sitt material väl, är medvetna och idkar källkritik. Lösningen i de här fallen är helt enkelt att utelämna sådana belägg vid redigeringen av ordboksartiklar.

SMS:s arkiv omfattar som sagt över 8 miljoner ordsedlar, det är därför ingen större överraskning att beläggkorten inte genomgående håller en jämn kvalitet. En del av insamlarna hade språkvetenskaplig utbildning, andra var lekmän. En del av lekmännen hade också flyttat och bott på flera olika orter, vilket gjorde dem till mindre pålitliga informanter för en viss dialekt. Med tiden har erfarenheten att arbeta med materialet inom SMS-redaktion gett upphov till en gemensam insikt om hur pålitliga de olika käl-

lorna är, något som även har beaktats i ordbokens redigeringsmanual.

Utmaningar när det gäller materialets pålitlighet i SMS kan bero på att en del orduppgifter har samlats in med hjälp av frågelistor som behandlar ett begränsat ämnesområde, t.ex. skörd, fiske eller olika jordbruksredskap. I vissa fall har det funnits bakgrundsuppgifter och frågeställaren kan ha gett ett dialektord som bjudord. Informanten kan då ha angett att ordet också används i hens dialekt, även om detta inte får stöd i andra källor från samma ort.

Misstankar när det gäller materialets pålitlighet kan också uppstå när en redaktör möter ett enstaka enda belägg. Har den som besvarat frågan inte haft en korrekt minnesbild eller handlar det om ett ord som bara använts i en mindre krets, inom en familj? Den här sortens ströbelägg förekommer även då informanter besvarat frågor om olika delar på ett bruksföremål. Det kan handla om att informanten i stunden använt ett tillfälligt beskrivande ord för just den delen av bruksföremålet.

I ordboksartiklarna syns detta ofta genom en spretighet i betydelsebeskrivningen. Typiska exempel på detta ur SMS är en del av de vaga betydelser som ges i artikeln *aluspuu* 'nedre el. undre trä'. Artikeln är indelad i fem betydelsemoment som anger olika typer av ribbor eller stänger och i betydelsemoment 5 ges ytterligare "muuta" 'övrigt' enbart i form av exempel. Artikeln *etupuu* 'främre trä' är på liknande sätt indelad i tre betydelsemoment, varav de två första avser främre trä på en släde eller en vävstol, och i det tredje ges enbart exempel. Enligt dagens uppdaterade redigeringsprinciper strävar redaktörerna efter att undvika lösningar där exempel ges utan betydelseförklaring i egna betydelsemoment.

I fråga om dialektordböcker påverkar även samhällliga förändringar, t.ex. när det gäller vad som är politiskt korrekt att publicera i vår tid jämfört med då materialet samlades in för över hundra år sedan. På FO har redaktionen noterat en skevhet när det gäller pejorativa benämningar på folkgrupper. En jämförelse av antalet

sammansättningar som anger pejorativa personbeteckningar under huvudorden *finne* 1 och *ryss*, *rysse* 1 visar att redigeringspraxis justerats på denna punkt; antalet pejorativa sammansättningar med *finn-* som förled (som utgavs på 1980-talet) är märkbart fler än sammansättningarna med *ryss-* som förled (publicerade 2023). Här uppstår lätt ett dilemma då ordboksredaktören ska navigera mellan vad som är politiskt korrekt idag, och i vilken utsträckning vi ger synlighet åt ett äldre samhälle som återspeglas i ordboks-basen.

VKS, som beskriver det tidiga finska skriftspråket utifrån skriftliga källor utgivna 1543–1810, domineras av religiöst språk, såsom Agricolas översättning av Nya testamentet och andra bibelöversättningar samt lagspråk i lagar och förordningar. Förklaringen till att dessa genrer är så starkt företrädda är att lejonparten av de texter som skrevs och utgavs på finska under perioden bestod av religiösa och administrativa texter. Vid redigeringen strävar redaktörerna efter att ge exempel som företräder varierande texttyper med jämn fördelning från olika decennier.

4.2. Krav på ständig uppdatering i en samtidsordbok

Då en äldre ordbok utgör ordboksbas för en efterträdare kan detta ses som ett slags halvfabrikat till en ordbok. Det försnabbar i viss mån redigeringen, men kan göra arbetet mer invecklat och begränsa möjligheterna att skapa nya lösningar. I synnerhet när det gäller exempel finns det en uppenbar risk att en del uttryck och fenomen föråldras oförmärkt. Övergången till KS från NS, som utkom 1951–1961, via PS på 1990-talet, har gått via processer av att sovra, utmönstra och komplettera. Till följd av resursbrist har det inte heller varit möjligt att påbörja arbetet helt från början. KS innehåller därför genom sin bakgrund flera olika tidsskikt där nya och gamla artiklar varvas.

4.2.1. Korpusmetoden

I dagens digitala värld förväntas en allmänspråklig ordbok vara baserad på stora korpusmaterial (Svensén 2004:58). Atkins & Rundell (2008:95–96) framhåller att lexikografin gärna ska vara evidensbaserad med en trovärdig förankring i det allmänna varierande språkbruket, vilket enligt dem kräver en tillräckligt stor och mångsidig korpus:

for lexicography, the best, most useful kind of corpus is one that combines very large volumes of data with diversity in a number of broad categories (like mode, medium, and domain), and a level of linguistic and textual annotation which aspires to high quality but does not seek perfection. [...] the single biggest benefit is the access it gives us to the 'regularities' of the language – the typical and recurrent features and patterns which make up the norms that lexicographers seek to identify and describe. There is no longer any serious argument about whether or not to use corpora in creating dictionaries.

För det finska språket saknas en heltäckande korpus som skulle ge tillräckligt underlag för att mångsidigt belysa olika användningsområden. Allmänt tillgängliga korpusar utnyttjas däremot regelbundet i arbetet med KS. I övrigt följer redaktörerna med vad som händer omkring dem, gör Google-sökningar och frågar experter. Vid uppdateringen av en ordbok behöver korpusmetoden inte heller alltid vara den enda eller den bästa lösningen (jfr Svensén 2004:56–58).

De mest centrala korpusarna finns i *Språkbanken i Finland* med konkordansverktyget Korp, som anpassats till finska språkbanken. Språkbanken innehåller språkresurser inom digitala humaniora och samhällsvetenskap och har ett brett utbud av text- och talkorpusar. Den omfattar ett flertal olika texttyper, men dagspress och ore-

digerade texter från diskussionsforum på internet dominerar för finskans del. En viktig poäng är att de oredigerade texterna delvis innehåller osakliga och alltför talspråkliga formuleringar. Men de kan vara ett hjälpmedel för att belysa bruket av vissa ord och uttryck.

4.2.2. Äldre ordböcker som ordboksbas

I KS påträffas gamla artiklar sida vid sida med nytillkomna artiklar. Nya ord föds genom nya fenomen och samhällsförändringar. Ett typiskt exempel är den administrativa termen *hyvinvointialue* 'välfärdsområde' som togs in i KS 2022, eftersom redaktionen i förväg kände till att termen skulle införas officiellt 1.1.2023. Också tillströmningen av lånord från engelska såsom verbet *chillata* 'chilla' har ökat. Ordet kom in i KS vid uppdateringen 2016, se (2).

- (2) **chillata** [tš-] slg. *oleilla, hengaiilla, chillailia*. *Nuorisoporukka chillasi nurtsilla*. (KS 2016)
 'chilla' [tš-] slang. 'hålla till, hänga, chilla' *Ungdomsgänget chillade på gräsmattan*. (förf. övers.)

I redigeringsprocessen har gamla artiklar i regel uppdaterats i större utsträckning än nya uppslagsord har kommit in i ordboken. Vid uppdateringarna av KS 2022 och 2024 omarbetades ca 1 000 artiklar men bara omkring 400 nya artiklar infördes. De redaktionella resurserna har varit alltför små för heltäckande systematiska genomgångar. Därför kan man på många ställen i ordboken hitta sådana artiklar som skulle kräva mer omfattande uppdatering. Exempelvis artikeln *kaulin* 'kavel' i (3) står fortfarande i samma form som den hade i PS 1990. Sett med dagens ögon framstår artikeln *kaulin* gammaldags och stereotypisk, eftersom det enda exemplet berättar om ett huskors. Å andra sidan var motsvarande artikel i föregångaren NS från 1950-talet i (4) ännu mer rakt på sak, för där belyser exemplet hur en hustru slår sin man i huvudet med en kavel.

(3) **kaulin** (puu)tela jolla kaulitaan taikinaa. Leik. ”naisval-
lan” symbolina. *Kaulinta heilutteleva pirttihirvu.* (PS 1990,
KS 2004–)

’rulle av trä med vilken man kavlar deg’. Lekfullt symbol för
”kvinnomakt”. *Ett huskors som svingar en kavel.* (förf. övers.)

(4) **kaulin** puutela, jota edestakaisin pyörittämällä taikina
kaulitaan. *Löi miestään kaulimella päähän.* (NS 1950-tal)

’rulle av trä, med vilken man kavlar deg’. (*Hon*) *slog sin man
med en kavel i huvudet.* (förf. övers.)

Redaktionen för KS har med vissa intervall försökt systemati-
sera genomgången av äldre material utifrån användningsområde
och ämne. Under de senaste åren har redaktionen tagit itu med
könsstereotyper genom att systematiskt gå igenom ordboken och
granska hur orden har använts. Bland annat har användningen av
orden *tyttö* ’flicka’ och *poika* ’pojke’ i exemplen granskats och åt-
gärdats. Ordet *poika* har varit överrepresenterat i olika negativa
sammanhang, såsom i exempelfrasen *Pojan ryökäle* ’pojkslyngel’ i
artikeln *ryökäle* i (5) A. Vid uppdateringen 2022 lades det i stället
in ett mer neutralt exempel, se (5) B.

(5) A **ryökäle** – – *Pojan ryökäle.* (KS –2021)

’lymmel, slyngel, skurk’ *Pojkslyngel.*

(5) B **ryökäle** – – *Kuka ryökäle on syönyt keksipurkin ty-
hjäksi?* (KS 2022–)

– – ’Vem sjutton har ätit upp alla kakor i kakburken’ (förf.
övers.)

Ordet *tyttö* ’flicka’ har tidigare används mer generellt i betydelsen
’ung kvinna’. Exempel på denna användning ges i (6), betydelse-
moment 4 b, nedan. Där vi kan följa hur det i KS 2020 lagts in en
kommentar om att detta i synnerhet gäller i äldre språkbruk.

(6) **tyttö** – – 4. b. *tyttöystävä, rakastettu, heila. Puistossa oli meripoikia tyttöineen.* (KS –2018)

flicka – – 4.b. *'flickvän, käraste, flamma'. I parken var (några) unga sjömän med sina flammor.*

tyttö – – 4. b. **vars. aik.** *tyttöystävä, rakastettu, heila. Puistossa oli meripoikia tyttöineen.* (KS 2020–).

flicka – – 4. b. *'i synnerhet i äldre språkbruk' [---] (förf. övers.)*

Lösningen har alltså dels varit att göra ändringar i exempel, såsom i (5) ovan, dels att markera att en viss användning av ett ord är föråldrad, vilket exemplifieras i (6).

4.2.3. Gränsdragning och lemmaselektion

I KS-nätordbok finns det i varje artikel en länk genom vilken användarna per e-post kan sända respons till redaktionen. Många ändringar och uppdateringar i ordboken har gjorts tack vare respons från användarna. Till exempel termer inom fysiken uppdaterades för några år sedan genom tips från en aktiv ordboksanvändare.

Eftersom ordboksbasen för KS är öppen och inte har fastställts i förväg i form av en korpus måste alla beslut om gränsdragning tas inom redaktionen. Exempelvis vilka ord inom ett visst specialområde som ingår i allmänspråket. Trots att KS är en allmänspråklig ordbok tas också termer från en mängd olika specialområden med i ordboken. Det kan handla om ord och termer som lekmän stöter på då de läser en dagstidning eller tar del av sin egen eller en familjemedlems patientjournal. Frågan är om vissa specialområden är överrepresenterade och andra specialområden fått för liten uppmärksamhet, vilket också påverkar vid uppdatering av ordboken.

Särskilt medicinska termer förekommer ofta i KS, men också ämnesområden som sport och matlagning är väl företrädda. Då sökningar med hjälp av markörer för användningsområden gjor-

des visade resultatet att det i KS finns sammanlagt: 2 180 sporttermer och 2 126 medicinska, 1 115 inom matlagning, 455 inom datateknik, men bara 173 inom handarbete. I viss mån har redaktörernas egna särintressen och specialkunskaper styrt de här processerna. Redaktörerna kan med hjälp av markörerna söka ord och termer inom ett visst ämnesområde för granskning och genomgång. Hittills har bl.a. språkvetenskapliga och zoologiska termer granskats gruppvis.

En annan fråga inom ramen för användningsområden berör allmänspråkets förhållande till det informella språket, alltså till vardagsspråket, dialekter och slang. I vilken utsträckning kan man säga att allmänspråket innehåller stilistiskt informella drag? Eller handlar det snarare om att vissa texttyper eller genrer i så hög grad förändrats i en mer informell riktning att man inte längre kan tala om ett helt neutralt allmänspråk? Tidningsspråk och talspråket i tv-program tillåter för finskans del numera en tämligen stor variation och innehåller fler och fler anglicismer. Linjedragningen gentemot lånord har alltmer blivit en kompromiss mellan den lavinartade tillströmningen och ett försiktigt godkännande och kodifiering av lånord. Vid valet av nya uppslagsord i KS påverkar även språkvårdens rekommendationer, eftersom ordboken också är normerande för språkbruket. Språkvården har i allmänhet favoriserat förfinskade former, och försökt undvika skrivformer som är främmande för finskan: bland annat har w-bokstaven fram till nyligen i allmänhet ersatts med v. Först i den version av ordboken som publiceras 2024 har det för verbet *vokata* 'tillreda wokmat' i samråd med språkvården förts in en parallellform *wokata* stavat med w. I årets uppdatering av KS ändras också stavningen av ordet *unisex*, till stavning med bokstaven x – fram till 2022 hade KS formen *uniseksi*, en stavning med ett -ks- enligt finska stavningsregler. Formen slog inte igenom eftersom den uppfattades som komisk, bl.a. eftersom det blir mångtydigt och kan tolkas som *uniseksi* 'sömnsex; drömsex'.

5. Sammandrag

Gemensamt för de fyra finländska enspråkiga ordböckerna SMS, FO, VKS och KS som vi har diskuterat är att deras tillkomsthistoria sträcker sig långt tillbaka i tiden. När det gäller ordboksbasen för dessa ordböcker och vissa skevheter och utmaningar som de innehåller är det just tidsaspekten som kan vara en viktig förklaring till utmaningarna.

Materialet för en ordbok är förmodligen aldrig fullkomligt; inom ordboksredaktionerna har det därför utvecklats olika lösningar för att vid redigeringen av ordboken rätta till och komplettera de skevheter som finns i materialet. I vissa fall fylls ordboksbasen på, t.ex. genom att redaktionen för KS exciperar nya ord och nya betydelser som vid uppdatering kan tas in i ordboken. Redaktionen för VKS har i flera omgångar breddat sin ordboksbas. Dessutom har man t.ex. vid redaktionerna för dialektordböckerna SMS och FO lärt sig att handskas med problem genom källkritik; lösningen är då vanligen att utelämna opålitliga belägg. I vissa fall kan det också vara ett medvetet val att låta skevhet och överrepresentation av vissa ord, betydelser och användningsområden framträda i ordboken, därför att syftet med ordboken är att beskriva en dialekt eller ett äldre språkbruk. När det är frågan om långvariga, omfattande ordboksprojekt är det inte heller möjligt att lättvindigt förändra redigeringsprinciper och praxis, eftersom ordboken helst ska vara så logiskt sammanhållen och enhetlig som möjligt från a till ö.

För såväl den allmänspråkliga ordboken KS som dialektordböckerna SMS och FO gäller att gränsen mellan vad som idag uppfattar som politiskt korrekt eller partiskt förändrats sedan materialen för ordböckerna samlades in. När det gäller en historisk dialektordbok vars material återspeglar ett äldre språkbruk och ett samhälle för hundra år sedan, innebär det en balansgång. Hur låter man ett äldre samhälle med andra värderingar återspeglas men

undgår att publicera artiklar med betydelser och exempel som i dagens värld kan uppfattas som stötande?

Litteratur

Ordböcker, korpusar och digitala resurser.

FO (2013–) = *Ordbok över Finlands svenska folkmål*. Helsingfors: Institutet för de inhemska språkens webbpublikationer 33. <kaino.kotus.fi/fo/> (juni 2024).

Korp = Språkbankens konkordansverktyg, anpassad till Finska språkbanken. <kielipankki.fi/korp/>.

KS (2004–) = *Kielitoimiston sanakirja*. Helsingfors: Institutet för de inhemska språkens webbpublikationer 35. <kielitoimistonsanakirja.fi> (juni 2024).

NLO (1997). = *Nordisk leksikografisk ordbok*. Henning Bergholtz, Ilse Cantell, Ruth Vatvedt Fjeld, Dag Gundersen, Jón Hilmar Jónsson, Bo Svensén (red.). Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 4. Universitetsforlaget.

NS (1951–1961) = *Nykysuomen sanakirja*. Matti Sadeniemi, Jouko Vesikansa m.fl. (red.). Suomalaisen Kirjallisuuden Seura, WSOY.

PS (1990–1994) = *Suomen kielen perussanakirja*. Kotimaisten kielten tutkimuskeskus, Painatuskeskus.

PS (1997) = *CD-Perussanakirja*. Kotimaisten kielten tutkimuskeskus, Oy Edita Ab.

SMS (2012–) = *Suomen murteiden sanakirja*. Helsingfors: Institutet för de inhemska språkens webbpublikationer 30. <kaino.kotus.fi/sms/> (juni 2024).

Språkbanken. <kielipankki.fi/sprakbanken/> (juni 2024).

Vendell, Herman (1904–1907): *Ordbok över de östsvenska dialekterna*. Helsingfors: Svenska litteratursällskapet i Finland.

- Wessman, V. E. V. (1925–1932): *Samling av ord ur östsvenska folk-mål*. Tillägg till H. Vendells ordbok över de östsvenska dialekterna. Helsingfors: Svenska litteratursällskapet i Finland.
- VKS (2014–) = *Vanhan kirjasuomen sanakirja*. Helsingfors: Institutet för de inhemska språkens webbpublikationer 38. <kaino.kotus.fi/vks/> (juni 2024).

Annan litteratur

- Ahlbäck, Olav (1946): *Studier över substantivböjningen i Finlands svenska folkmål*. Studier i Nordisk Filologi 33–34. Helsingfors: Svenska litteratursällskapet i Finland.
- Atkins, Sue & Michael Rundell (2008): *The Oxford Guide to Practical Lexicography*. Oxford University Press.
- Forsberg, Markus & Louise Holmer (2024): Datatillgång, metodutveckling och lexikografiskt arbete vid Språkbanken Text. I: *LexicoNordica* 31 (denna volym).
- Haarala, Risto (1994): De viktigaste ordböckerna över finskt allmänspråk. I: *LexicoNordica* 1, 53–61.
- Haarala, Risto (1999): Vem är de finska ordböckerna gjorda för? I: Peter Slotte, Pia Westerberg & Eva Orava (red.): *Nordiska studier i lexikografi* 4. Rapport från Konferensen om lexikografi i Norden Esbo 21–24 maj 1997. (Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 5.). 149–158.
- Lilja, Agneta (1996): *Föreställningen om den ideala upppteckningen. En studie i idé och praktik vid traditionssamlade arkiv – ett exempel från Uppsala 1914–1945*. Skrifter utgivna genom Dialekt- och folkminnesarkivet i Uppsala, Ser. B: 22. Uppsala.
- Onikki-Rantajääskö, Tiina (2011): Mestari–kisälli–mallin soveltuvuus fennistiikan historiaan. I: *Virittäjä* 2011, 542–574.
- Ruppel, Klaas & Caroline Sandström (2014): Stora finska ordböcker i ett historiskt perspektiv. I: *LexicoNordica* 21, 141–160.

- Sandström, Caroline (2013): Från folkmålsstudier till interaktionell dialektologi. I: *Folkmålsstudier* 51, 87–113.
- Sandström, Caroline (2016): Perspektiv på svensk lexikografi i Finland med Ordbok över Finlands svenska folkmål och Finlandssvensk ordbok som exempel. I: Anna Helga Hannesdóttir (red.): *Framtidens lexikografi. Rapport från ett symposium*. Mejerbergs arkiv för svensk ordforskning 42. Göteborg. 75–109.
- Setälä, E. N. (1896): Suomalaisen sanakirjatyön ohjelmasta. I: *Suomi* 3:13, 89–95.
- Svensén, Bo (2004): *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Andra omarbetade och utökade upplagan. Stockholm: Norstedts Akademiska Förlag.
- Vikør, Lars S. (1999): Fleirgenerasjonsordbøker og tida. I: Peter Slotte, Pia Westerberg & Eva Orava (red.): *Nordiska studier i lexikografi* 4. Rapport från Konferensen om lexikografi i Norden Esbo 21–24 maj 1997. (Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 5.). 395–405.
- Vilppula, Matti (1999): Ordbok över finska dialekter och evigheten. I: Peter Slotte, Pia Westerberg & Eva Orava (red.): *Nordiska studier i lexikografi* 4. Rapport från Konferensen om lexikografi i Norden Esbo 21–24 maj 1997. (Skrifter utgivna av Nordiska föreningen för lexikografi. Skrift nr 5.). 407–411.

Tarja Riitta Heinonen
 ordboksredaktör, fil. dr
 Institutet för de inhemska språken
 Hagnäskajen 6
 FI-00530 Helsingfors
 tarja.heinonen@kotus.fi

Caroline Sandström
 ordboksredaktör, fil. dr
 Institutet för de inhemska språken
 Hagnäskajen 6
 FI-00530 Helsingfors
 caroline.sandstrom@sprakinstitutet.fi