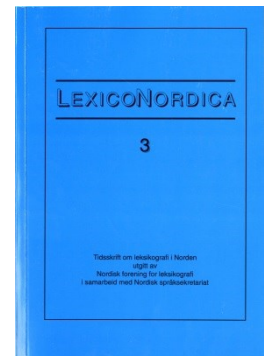


LexicoNordica

Titel: Korpus över äldre finska
Forfatter: Liisa Nuutinen
Kilde: LexicoNordica 3, 1996, s. 129-139
URL: <http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive>



© LexicoNordica og forfatterne

Betingelser for brug af denne artikel

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

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

Søgbarhed

Artiklerne i de ældre LexicoNordica (1-16) 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.

Liisa Nuutinen

Korpus över äldre finska

The articles in the Dictionary of Old Finnish are based on the collection of about half a million entries of words and a corpus in machine-readable form, the compilation of which was started in 1992. The corpus contains the essential documents of old Finnish and additionally other texts from different fields. In choosing the material, the goal has been a coverage as complete as possible from the lexicographic point of view. At present, the corpus contains over 2 million word forms in running texts, and the editors of the dictionary can pick the example sentences directly from it. The search programs always go sentence by sentence in looking for words, word forms or sequences or signs. It is also possible to look for alternative words or sign sequences, either exact or approximate ones, or limit the search to morphological elements. When necessary, word lists can be made using KWIC program.

Om Vanhan kirjasuomen sanakirja

Med *äldre finska* avses finskan i de första skriftliga dokumenten från mitten av 1500-talet fram till år 1810. Denna finska beskrivs i *Vanhan kirjasuomen sanakirja* ('Ordbok över den äldre finskan'), som håller på att utarbetas vid Forskningscentralen för de inhemska språken. Tryckta verk och enhetliga manuskript på finska från denna tid utgör basen, så att alla finska ord som förekommer i den här tidens litteratur tas med i ordboken som uppslagsord. Uppslagsordsformen är för det mesta normaliserad till modern finska, och efter det ges de olika varianterna. Sedan följer ordklassangivelse och efter det kommer uppgifter som hämtats ur gamla ordböcker, sedan ordets definition och sist exemplen. Ordboken över äldre finska är en citatordbok. Det betyder, att exemplen får tala för sig själva genom att de presenteras i betydelsegrupper. Citaten har valts så att de förutom betydelse också visar ordets rektion och typiska kontexter. Dessutom skall de ange uppslagsordets utbredning och skriftliga ålder.

I ordbokens källförteckning ingår litet över tusen verk, av vilka ungefär hälften är ganska korta författningstexter. Utgångspunkten för vår ordbok är ordförrådet hos Mikael Agricola, som skapade det finska skriftspråket. En stor del av vårt material består av översättningar, och därför har originaltexterna på svenska, latin och tyska stor betydelse för ordboksarbetet. En excerptsamling på en halv miljon ord utgör tillsammans med korpusen i maskinläsbar form basen vid redigeringen. Korpusen innehåller i dag drygt 2 miljoner ord.

Ordboken över äldre finska är ett av forskningscentralens stora ordboksprojekt. När den är färdig kommer den att bestå av 6 delar och cirka 5.500 sidor. Hittills har två delar utkommit. De omfattar bokstavsintervallet A till K. De följande delarna planeras utkomma med fyra till fem års mellanrum.

Om korpusens innehåll

Grunden till korpusen över äldre finska lades 1992 då forskningscentralen anslöt sig till Internet och direktionen fattade ett principbeslut om att det skall skapas texter i maskinläsbar form. Då uppstod tanken att vi skulle utnyttja de perforeringsremсор som vi hade fått från förlagen i samband med att några viktiga verk på äldre finska hade satts med moderna bokstäver och getts ut under sina respektive jubileumsår. Verken var Mikael Agricolas samlade verk från åren 1543–1552, den första finska psalmboken av Jaakko Finno från slutet av 1500-talet, översättningen av kyrkolagen av år 1686 samt översättningen av Sveriges rikets lag från år 1734. Dessutom kunde forskningscentralen tack vare ett ekonomiskt bidrag få den första finska Bibeln från år 1642 i elektronisk form, så att den fanns tillgänglig 1992 då den finska Bibeln fyllde 450 år. Efter det här har korpusen utökats så att den i dag består av drygt 2 miljoner ord. Största delen av det nytillkomna materialet har renskrivits på dator, eftersom den ursprungliga texten är skriven i frakturstil. De delar av materialet som inte har varit skrivna i fraktur har vi kunnat skanna. En förteckning över det material som ingår i korpusen ges i ett appendix.

Korpusen kan delas upp i tre ämnesområden: religiös text, varav en del utgörs av psalmer, lagtext och annan profan text. Religiöst material utgör den största andelen av den äldre finskan och därmed också av korpusen. Vi har med de båda viktigaste andliga texterna, alltså Agricolas samlade verk och 1642 års Bibel. Vi har också försökt få med de viktigaste lagtexterna. Nu innehåller korpusen Herr Mårtens översättning av kung Kristofers landslag, som var den första lagtexten som översattes till finska i slutet av 1500-talet, kyrkolagen från år 1686 och Sveriges rikets lag från år 1734. Från verket *Acta Historica Fennica* har vi skannat en samling olika korta lagtexter. Vi hoppas att vi i år skall få in den första finska stadslagen från år 1601 och den andra översättningen av landslagen från år 1609. En samling författningstexter från 1600- och 1700-talet representerar olika livsområden, allt från brännvinsframställning till forsrensning, och eftersom det i allmänhet finns ganska litet världslig text från den här tiden, tänker vi utöka antalet författningstexter i korpusen. Den profana litteraturen

representeras också av 33 stycken almanackstexter från 1700-talet ända från det att almanackor på finska började utkomma i början av 1700-talet. Facklitteraturen representeras av Lizelius *Tietosanomat* som var den första dagstidningen på finska, samt två små böcker av Christfrid Ganander, som representerar dåtidens medicinska uppfattningar om människans och djurens sjukdomar och deras behandling. De två andra böckerna av Ganander består av finska gåtor och en samling fabler. Frosterus *Hyödyllinen Huwitus Luomisen töistä* är en naturvetenskaplig skrift om Guds skapelseverk.

Materialet är sparad i operativsystemet VMS med DEC:s 8-bitars multinationella teckensystem, som är nästan identiskt med ISO:s Latin-1-teckensystem. Forskningscentralen ingår i telekommunikationsnätet FUNET tillsammans med högskolorna och andra forskningscentraler. Genom FUNET är forskningscentralen kopplad till Internet, och via Internet har också utomstående personer tillgång till korpusen. Det kräver dock att man ansöker om en användarlicens av forskningscentralen. Sådana beviljas för forskningsändamål.

Arbetet med korpusen har till stor del utförts vid sidan av det egentliga ordboksarbetet och redaktionen har försökt välja ut de texter som varit mest centrala för redigeringen av ordboken. Vi har också försökt täcka en mängd olika ämnesområden för en så allsidig terminologi som möjligt. Vår strävan är att få med alla texter från 1500-talet, eftersom vi i ordboken alltid försöker ge det första belägget på ett ord.

Kodningen

För att korpusen skall vara så användarvänlig som möjligt har vi i de olika korpusmaterialen strävat efter att förenhetliga såväl formen som kodningen av materialet. Nästan alla verk som finns medtagna är löpande text. Det här innebär att en mening som avslutas med punkt eller liknande bildar en rad i korpusen. I början av varje rad finns en källangivelse där vi kodat in vilket verk det är fråga om. Koderna är desamma som i ordboken.

RWL1759-80-Rak-5:1 1 . § . Weräjä , josta kylän tie käypi , pitää caikilda nabureilda rakettaman ja ylöspidettämän .

RWL1759-80-Rak-5:1 Weräjän omalle pellolle ja niytulle mahta jocainen itze raketa ja sijtä waarin otta .

1Ms-1:0 I. Lucu . MAilman luomises teke Jumala Sanalla ja Hengellä I. Päiwänä / walkeuden .

1Ms-1:1 ALgusa loi Jumala Taiwan ja Maan .

1Ms-1:2 Ja Maa oli autia ja tyhjä / ja pimeys oli sywyden päällä / ja Jumalan Hengi lijckui weden päällä .

1Ms-1:00 Vers . 2. Jumalan Hengi) Silloin ei ollut wielä tuulda / sentähden tällä sanalla (Jumalan Hengi) ymmärretän pyhä Hengi .

JAFro1791-75 Korwalehdet ulkopuolella päätä kokoawat äänen , ja mutkaisen matkan kautta laskewat sen korwan sisälle .

Källangivelsen innehåller för det mesta namnet på författaren eller översättaren och verkets utgivningsår samt uppgift om sidnumret i originalkällan. För lagtexter anges också namnet på lagbalken, samt kapitel och paragraf. I källangivelserna för Bibeln anges bok, kapitel och vers, och dessutom anger vi också om det är fråga om text som inte ingår i verstexten utan t.ex. i förordet, i summarierna före verstexten (i exemplet 1Ms-1:0) eller i efterskriften (1Ms-1:00).

Det finns alltid ett mellanslag före och efter textordet, så att man så exakt som möjligt skall kunna söka ett visst textord. Det här gör att materialet också lämpar sig för morfologiska sökningar och dessutom kan tjäna annan språkforskning. Av samma orsak har alla skiljetecken avskilts från textorden med ett mellanslag. Ord som har delats på två rader i det tryckta materialet har i korpusen sammanfogats till ett odelat ord. Vi har kodat sidbyten mitt i en mening och vid behov har vi också kodat främmande ord, rubriker och vissa diakritiska tecken. Däremot har vi inte kodat typografin eller strukturen i den ursprungliga texten.

Mikael Agricolas samlade verk var vårt första korpusmaterial. Som grund för korpusen ligger en datafil som kom till då Agricolas verk utgavs i hålkortsform i slutet av 1970-talet. Materialet innehåller cirka 490.000 textord. Av tekniska orsaker var hela hålkortsmaterialet skrivet med versaler och följde det tryckta materialets sid- och radindelning. Här är ett exempel, början av psalm 23 ur Agricolas översättning, den som börjar "Herren är min herde, mig skall intet fattas".

DATA: [KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.TXT

324121 DAUIDIN PSALMI .

324124 HERRA ON MINUN PAIMENEN / EI MITEN MINULTA

324125 PWTU .
 324126 HEN CAITZEPI MINUA WIHERIEISES NITYS / IA
 324127 WIEPI MINUN WIRGOTTOUAN WEDHEN TYGE .
 324128 MINUN SIELUNI HEN WIRGOTTA / HEN WIEPI MINUN
 324129 OIKIALLE TIELLE / HENEN NIMENS TEDHEN .
 324130 IA WAICKA MINE WAELLAISIN PIMIES LAXOS / EM MINE
 324131 PELKEISI MITEKEN PAHUTTA / ETTES OLET MINUN CANSANI / SINUN
 324201 CAPULAS IA SAUWAS MINUN TUKEEUAT .

Det var väldigt besvärligt att läsa och hantera ett material som var skrivet i versaler, och som hade mycket korta rader och som dessutom hade en sexsiffrig källangivelse i början av varje rad. Det här gjorde att vi lät ett dataprogram ändra alla versaler till gemener, utom den första bokstaven i sådana ord som Gud, Jesus, Kristus, Amen o.s.v. Därefter måste nödvändiga ändringar från gemener till versaler göras manuellt genom att korpustexten jämförs med originalet, och det är både arbetsamt och tidskrävande. Dessutom har vi nu ändrat på radindelningen i texten. Vi ordboksredaktörer plockar ju ut de textavsnitt vi vill ha direkt ur korpusen, och vi har ingen nytta av att textens radindelning följer originalet, tvärtom. En kort rad ger inte tillräcklig kontext. Radindelningen har nu ändrats så att en mening i texten motsvaras av en rad i korpusen. I Agricolas bibelöversättningar motsvaras en vers av en rad i korpusen. På så vis omges textordet alltid av en hel mening, och inte av en textrad, som i allmänhet är kortare än en mening i det tryckta verket. En sådan här rad i korpusen kan gå över flera rader på skärmen eller i utskrift, men för datorn är den alltså bara en rad.

DATA:[KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.SEN

324121 Daidin Psalmi .
 324124 HERRA on minun Paimenen / ei miten minulta pwtu .
 324126 Hen caitzeppi minua wiherieises Nitys / ia wiepi minun wirgottouan Wedhen tyge .
 324128 Minun Sieluni hen wirgotta / Hen wiepi minun oikialle Tielle / henen Nimens tedhen .
 324130 Ja waicka mine waellaisin pimies Laxos / em mine pelkeisi miteken pahutta / Etttes olet minun cansani / Sinun <> Capulas ia Sauwas minun tukeeuat .

Bibeln från år 1642 hade också ursprungligen originalets radindelning. Även den har ändrats så att varje vers bildar en rad i datorn. Verket finns som helhet i korpusen, också förord och register. Och all text som

inte ingår i själva verstexten har kodats (summerier före verstext, efterskrift, rubriker och marginaltexter).

Arbetet med korpusen har lärt oss att lägga in tillräckligt med information i källangivelsen. Men det måste vara rätt sorts information. Vi har exempelvis märkt att radnumreringen för det mesta är överflödig. Däremot har källangivelsen och informationen i den en viktig funktion för jämförelser mellan texterna, t.ex. mellan olika översättningar av en och samma text. Vid redigeringen av ordboken jämför vi alltid Agricolas bibelöversättningar med Bibeln från 1642 och tvärtom. För närvarande är de här jämförelserna tidskrävande. Eftersom det inte finns versindelningar i Bibeln i Agricolas tryckta böcker (och därför inte i den nuvarande korpusen) kan samma ställe i de båda bibeltexterna inte lokaliseras automatiskt. Därför har vi nu börjat koda in verserna i korpusen i enlighet med 1642 års Bibel för att underlätta jämförelsearbetet. Den del av Agricola-korpusen som är färdigkodad ser ut så här:

DATA:[KORPUS.WANHASUOMI.AGRICOLA]AGRI3PS.SEN

A-III-241-[Ps-23:1] Daudin Psalmi . HERRA on minun Paimenen / ei miten minulta pwtu .

A-III-241-[Ps-23:2] Hen caitzepi minua wiherieises Nitys / ia wiepi minun wirgottouan Wedhen tyge .

A-III-241-[Ps-23:3] Minun Sieluni hen wirgotta / Hen wiepi minun oikialle Tielle / henen Nimens tedhen .

A-III-241-[Ps-23:4] Ja waicka mine waellaisin pimies Laxos / em mine pelkeisi miteken pahutta / Ettes olet minun cansani / Sinun <> Capulas ia Sauwas minun tukeeuat .

Det här innebär att vi med ett enda kommando får fram samma vers i 1642 års Bibel och i Agricolas verk. Uppgifterna om bibelbok, kapitel och versnummer står inom parentes, eftersom de inte anges i ordboksartiklarna. Vi har också löst upp nummerkoden i början av varje rad, så att källangivelsen nu har samma form i korpusen som i ordboken. Vi kommer att föra in likadana koder också i Finno-korpusen för att underlätta jämförelser mellan Finnos psalmbokstexter och de två senare utkomna psalmböckerna från 1605 och 1701. I lagtexternas källangivelser har vi fört in uppgifter om både lagbalk, kapitel och

paragraf så att det är lättare att lokalisera ett textord. Den praktiska erfarenheten har alltså lärt oss att utforma korpusen enligt användarbehovet och att korpusinnehållet kan vara detsamma som i originalverken trots att formen är något helt annat. Det är syftet med en korpus som avgör hur den skall se ut.

För att underlätta de tidskrävande textjämförelserna har vi som målsättning att få in alla sådana texter i korpusen vars ordformer vi måste jämföra sinsemellan. För att reda ut ords etymologi vore det många gånger också idealiskt att ha med de svenska originaltexterna. Det gäller först och främst lagtexterna och den första svenska Bibeln.

Utnyttjandet av korpusen vid ordboksredigeringen

Sättet att utnyttja korpusen vid artikelredigeringen varierar antagligen något från redaktör till redaktör, men som regel tror jag att stommen till en artikel kommer till på basis av materialet i excerptsamlingen, eftersom den har en högre täckningsgrad både tids- och materialmässigt, medan korpusen främst utnyttjas för kompletteringar. Det händer dock att det kommer fram så mycket nytt material vid sökningarna i korpusen att artiklarna måste omstruktureras, t.ex. vad gäller indelningen i betydelsegrupper. Sökningarna i de större och centralare korpusdelarna görs alltid som egna körningar, eftersom de ger ett så rikt exempelmaterial. Om en sökning görs på flera olika ord grupperas sökresultatet ordvis, varefter en jämförelse görs med motsvarande ordboksartiklar, som vid behov kompletteras genom att exempel kopieras och klistras in. För det mesta görs sökningarna över ett visst alfabetavschnitt åt gången och inte på enskilda uppslagsord. I samband med körningarna kommer det ofta också fram nya ord, som inte återfinns i excerptsamlingarna. Dessa får självfallet egna artiklar i ordboken.

Sökprogrammen *search*, *agrep* och *kwic*

I de sökprogram vi använder riktar sig sökningen alltid till en rad i datorn. Man kan söka en eller flera sinsemellan olika teckensträngar. Det går också att söka ordpar, och just för de sökningarna är det väsentligt att korpusraderna är tillräckligt långa. Man kan söka samtidigt i flera olika filer.

Det sökkommando som lämpar sig bäst för vårt arbete heter *search*. Det är ett kommando som ingår i operativsystemet. I sin enklaste form kan kommandot se ut såhär:

search fil(er) teckensträng

det vill säga "sök en viss teckensträng i en eller flera filer". Om datorn hittar teckensträngen visar den träffarna på skärmen tillsammans med kontexten i den ordning de förekommer i korpusen. Man kan lägga till preciseringar för att exempelvis förlänga kontexten kring textordet eller samla statistiska uppgifter.

Ett annat sökkommando som kan användas är *agrep* som hör till UNIX-systemets program av grep-typ. Agrep används i synnerhet för approximativa sökningar. Det godkänner regelbundna uttryck som sökargument. Med hjälp av så kallade metatecken kan man bilda mer komplicerade söksträngar enligt vissa regler. Man kan meddela hur stor avvikelser får vara från den teckensträng man söker. Man kan också använda optionen *best-match* och då letar algoritmen efter den mest lämpliga träffen och meddelar hur stor avvikelser är. För en användare som inte känner gammal finsk ortografi är *agrep*-kommandot mer användbart än *search*-kommandot.

När man bläddrar i ett stort material är *kwic* ett bra hjälpprogram. Vårt *kwic*-program godkänner också regelbundna uttryck som sökargument. Det listar beläggen i alfabetisk ordning under varandra på skärmen, vilket gör det möjligt att komma åt den äldre finskans mycket varierande skrivformer av ett och samma ord. En av *kwic*-programmets förtjänster är att man kan lägga till en sekundär söknyckel så att beläggraderna i andra hand listas enligt ordet före eller efter själva belägget. På så vis får man lätt fram sammansatta ord som mycket ofta är särskrivna i äldre finska. Exemplet nedan visar en sökning på ordet *tauti* 'sjukdom' med dess böjningsformer där ordet före sökordet är sekundär söknyckel.

```
kwic -r -s-1 "tau[t|d][i|e]" kotiapteekki.sen
```

```
[Gan1788a-40] tu ottaa pois hammas taudin , hauteexi ajoxijn äkämiijn
```

```
[Gan1788a-40] aan , parantaa korwa taudin .
```

```
[Gan1788a-26] 38. Recepti Rewäsin Taudin öljyyn. ( Moder droppar )
```

```
[Gan1788a-8] 16. Jalka taudisa ( podagra ) uloswetää se
```

```
[Gan1788a-23] n alusa , lankeewasa taudisa ; Kelta- jäsenten- ja [Gan1788a-19]
```

```
hewoisille loukkaus taudisa , lusikallinen sisälle
```

```
[Gan1788a-26] 100 tippua , Rewäsin taudisa , raskaisa lapsen
```

```
[Gan1788a-8] n päänskipuin , wattan tauteihin , wiemistymisiin , ja
```

```
[Gan1788a-8] jn , selkä- ja lonkka tautein , hammas tautijn ja pahan
```

```
[Gan1788a-24] hyvä keuhko ja rinta tautein ja raskampaan ja helpompään
```

```
[Gan1788a-18] ain ja saastuttawain Tautein lijkkeisä , warjeelee , ja
```

En annan fördel med programmet är att man får fram också sådana sökord där den sökta teckensträngen finns mitt inne i ordet. Tack vare det kan man söka ihopskrivna sammansättningar på efterledsstammen.

kwic -r "tau[t|d]" kotiapteekki.sen

[Gan1788a-40] ain pureman päälle , jalkatautiijn (podagraan) ja
 [Gan1788a-7] 12. Punataudisa , yxi plaastari lappu
 [Gan1788a-46] 0 pisarta pureisa ja punataudisa , Hammas-taudisa
 [Gan1788a-18] ta tulleisa Räkä- ja Päätaudeisa woidellaan päälaki .
 [Gan1788a-7] 13. Ruusutaudisa päälle pannaan kahden
 [Gan1788a-21] , weren sylkemys ja täwytauti .
 [Gan1788a-34] uin woiman kuluttawa ulkotauti on seurasa .
 [Gan1788a-23] sten kansa tullesa ; Wesitautien alusa , lankeewasa
 [Gan1788a-18] 3. Wilutaudeisa sisälle otetaan 1 eli
 [Gan1788a-40] auttaa 3:nen päiwän wilutaudisa , Lukinwerkko haawan
 [Gan1788a-22] ia , perkaa wattan , wilutaudisa hywä , ja ummella

Denna funktion möjliggör också morfologiska sökningar.

kwic -r "ium" kirkkolaki.sen

[Flor1688-126-XXV:1] ispana canssa Auditoriumin / eli sijhen paikkaan
 [Flor1688-146-IV:0] / nautita sitä beneficiumita / cuin täsä processis
 [Flor1688-104-XXII:2] cuin Upsalan Conciliumis , Anno 1593 . caikilda
 [Flor1688-1-I:1] Upsalan Conciliumis wastanotettu / ja
 [Flor1688-Aa4b] / ilman Pispan ja Consistorialium saldimat .
 [Flor1688-144-I:0] Nijsä Consistoriumeis joisa
 [Flor1688-123-XXIV:29] Pispalda ja Consistoriumilda anoman Pappia
 [Flor1688-91-XIX:22] n Pispalda ja Consistoriumilda täydel woimalla
 [Flor1688-11-II:5] itä Epistola ja Evangeliumi weisattaman Altarin edes
 [Flor1688-150-XI:0] ja hänen pyhän Evangeliumins päälle / että minä
 [Flor1688-81-XIX:6] / Schouluis / Gymnasiumeis eli Academiois / jolla
 [Flor1688-144-I:0] Meidän Hof-Consistoriumis on Meidän
 [Flor1688-88-XIX:17] ten ja Caluin Inventarium ynnä Kircon Kirjain
 [Flor1688-Bb4b] lla on erinomainen Privilegium Papin walitzemises /
 [Flor1688-85-XIX:11] n erinomaisen Privilegiumin Papin walitzemises /

Avslutning

Jag har gjort vissa uträkningar över hur stor del av våra ordboks-exempel som i dag kan tas direkt ur korpusen. I de ordboksartiklar jag baserade uträkningen på fanns det 25 till 115 exempel per artikel, och jag kom fram till att 32 till 50 procent av ordboksexemplen fanns i korpusen. Extremvärdena var 24% och 64%. Jag tror att de här siffrorna är representativa för exemplen som helhet, eftersom vår korpus täcker de centralaste och samtidigt största verken över äldre finska. Det här innebär att ungefär hälften av exemplen som kommer att ingå i ordboken inte behöver skrivas av ordboksredaktörerna. Dessutom slipper vi granskningen. Vi kan också vara säkra på att få med alla ord i de centralaste verken. Det här hör till fördelarna med korpusarbetet. Det negativa med det är de stora textmängder som är resultatet av vissa sökningar och som redaktörerna blir tvungna att gå igenom manuellt. Men tack vare korpusen kan vi i alla fall vara förvissade om att vi i vår ordbok får med ett heltäckande och representativt exempelmaterial över det centralaste ordförrådet i äldre finska.

Appendix: korpus över äldre finska

Agricolas verk I–III (1543–1555) (492 000 ord)

Jaakko Finnos Psalmbok c. 1583 (21 000)

Piae Cantiones 1616 (14 350)

Bibeln 1642 (c. 1 milj. ord)

Likpredikningar (1600- och 1700-talet) (176 400)

Herr Mårtens landslag (1500-talet) (52 400)

Kircko-Laki ja Ordningi 1688 'Kyrckio-Lag och Ordning' (42 800)

Ruotzin Waldacunnan Laki 1759 'Sveriges Rikes Lag' (102 400)

Acta Historica Fennica (1500- och 1600-talet) (22 900)

Författningstexter 1600- och 1700-talet (44 500)

Almanackstexter från 1700-talet (53 000)

Lizelius:

Suomenkieliset Tieto-Sanomat 1775–1776 'Finskt Tidningsblad' (24 800)

Ganander:

Suomalaiset Arwotuxet, Wastausten kanssa 1783 'Finska gåtor med svar' (4 450)

Uudempia Uloswalituita Satuja 1784 'Nya utvalda fabler' (28 200)
Maan-Miehen Huone- ja Koti-Aptheeki 1788 'Landtmannens hus-
apotek' (8 300)
Eläinden Tauti-Kirja 1788 'Om husdjurens sjukdomar' (13 300)
Frosterus:
Hyödyllinen Huwitus Luomisen Töistä 1791 'Nyttig förströelse om
Guds skapelseverk' (23 400)
Varia: tryckta verk från 1700-talet (5 500)

(Övers. från finska av Pia Westerberg, bearbetning av Nina Martola)