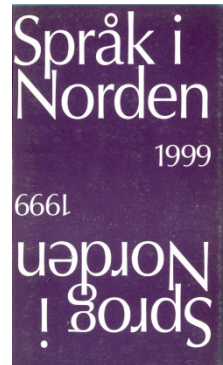


Sprog i Norden

Titel: CD-Perussanakirja som ett hjälpmedel i språkvården
Forfatter: Marja Lehtinen
Kilde: Sprog i Norden, 1999, s. 44-50
URL: <http://ojs.statsbiblioteket.dk/index.php/sin/issue/archive>



© Nordisk språkråd

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 numre af Sprog i Norden (1970-2004) 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.

CD-Perussanakirja som ett hjälpmedel i språkvården

Marja Lehtinen

Översättning: *Eva Orava*

Den här artikeln behandlar cd-romversionen av "Suomen kielen perussanakirja" (Finsk basordbok) och på vilket sätt språkvårdaren kan använda den som hjälpmedel. Vid utarbetandet av CD-Perussanakirja har man använt sig av den allra nyaste språkteknologin, vilket gör att den avviker från många andra elektroniska ordböcker.

Först några ord om själva ordboken: CD-Perussanakirja baserar sig på den tryckta, tredelade "Suomen kielen perussanakirja", som innehåller ungefär 100 000 uppslagsord. Den tryckta boken utkom åren 1990–1994 och är den mest omfattande aktuella ordboken som presenterar det nufinska ordförrådet. Ordboken är också normativ; den skildrar den betydelse som orden har i dag, berättar om hur orden skall böjas, och så vidare. Ordboken är utgiven av Forskningscentralen för de inhemska språken och förlaget Oy Edita Ab. Programmeringen har utförts av det finska språkteknologiföretaget Lingsoft Oy, som är specialiserat på språkprogram. Den elektroniska versionen utkom i februari 1997.

I nordiska kretsar är det troligen känt att finska språkbyrån vid Forskningscentralen för de inhemska språken upprätthåller en flitigt utnyttjad telefontjänst. Dagligen besvaras ca 70–80 samtal, och antalet samtal växer hela tiden. Rådgivarna antar att en orsak till att språkbyrån kan besvara allt fler samtal är att CD-Perussanakirja har gjort deras arbete snabbare: programmet ger snabbare än den tryckta versionen svar på många frågor som ställs.

I det följande ges några exempel på vilka slags frågor telefonrådgivningen har fått den senaste tiden, och hur CD-Perussanakirja kan ge svar på dem.

1 Vad betyder ordet?

Det här är kanske det mest typiska sättet att använda en enspråkig ordbok på, och den elektroniska ordboken utgör i det här fallet inget undantag från hur den tryckta versionen används. Vad betyder egentligen *infrastruktuuri* (infrastruktur)? Eller nyordet *karnevalisoida*¹ (det vill säga att försvaga positionen eller värdigheten hos något, genom att till exempel parodiera eller utsätta det för ytlig publicitet)? Lånord orsakar alltså ofta problem även för personer som talar finska som sitt modersmål, men också sällsynta ord av finskt ursprung och ord som blir allt ovanligare kan väcka frågor. Telefontjänsten har bland annat fått frågan vad adverbet *semminkin* betyder (i modern finska används i stället orden *varsinkin*, *etenkin*, och *nimenomaan*, 'i all synnerhet, framför allt, särskilt').

2 När kan man använda pronomenet *se* (det) om en människa?

Ordartikeln *se* är ganska lång, vilket ordartiklarna ofta är när det gäller pronomen. Det framgår följaktligen inte direkt under vilket moment informationen kan hittas. Om man klickar på knappen Förklaringar radas ordartikelns huvudbetydelser upp, och då ser man lätt att moment tre avser människor. Man kan sedan återgå till hela texten och läsa igenom momentet med alla dess exempelmeningar. Av texten framgår att pronomenet *se* kan syfta på en människa i vissa determinativa sammanhang.

Ordartikeln kan alltså ge information på olika sätt: i CD-Perussanakirja kan man titta på endast uppslagsord och förklaringar eller uppslagsord och exempelsatser. Denna egenskap är användbar när man skall undersöka långa ordartiklar, och det är endast i den elektroniska versionen som man kan läsa ordartiklar på detta sätt, aldrig i den tryckta produkten. – Här är det trots allt inte ännu fråga om språkteknologi, utan om att

¹ Ordagrant *karnevalisera*.

Marja Lehtinen

ordboksmaterialet är kodat med kodspråket SGML. Det innebär att materialet innehåller information om ordartikelns olika beståndsdelar, såsom uppslagsord, förklaringar och exempelsatser.

3 Rättstavning: Hur många g finns det i ordet *aggressiivinen* (aggressiv)?

Ibland vill frågeställaren endast försäkra sig om det korrekta sättet att stava ett ord. I sådana fall kan man inte alltid vara säker på att den form som enligt frågeställaren borde vara den rätta finns som uppslagsord i ordboken. CD-Perussanakirja har emellertid sökegenskaper som gör att informationen lätt kan hittas.

Om man inte vet hur ordet *aggressiivinen* (aggressiv) skall stavas, kan man som sökord skriva *ag*ressiivinen*. Det resulterar bara i en träff: det rätta sättet att stava ordet är att skriva det med två g.

4 Hur lyder språkvårdens rekommendation i fråga om adjektiv med efterleden *-rikas*? (-rik)

I CD-Perussanakirja finns det också ord och uttryck som hänvisas till bättre alternativ med orden ”pitää olla” (skall heta) och ”paremmin” (hellre). Ofta får telefonrådgivningen till exempel förfrågningar om vad som kunde utgöra en bättre motsvarighet till adjektiv som slutar på *-rikas* (-rik). Dessa ord har alltid ansetts vara så kallade svecismer – översättningslån från svenskan – som man har rekommenderat och fortfarande rekommenderar mer finska motsvarigheter till. Adjektiven med efterleden *-rikas* fås i CD-Perussanakirja fram med en enda sökning (**rikas*), och rekommendationerna ges behandligt när sökningen kombineras med tillvalet ”Rekommendation”.

5 Hur böjs ordet?

Ordens böjningsformer är inte självklara ens för finskspråkiga, och därför ger ordboken information om dem. Många frågor som kommer till telefonrådgivningen berör böjningen av ord.

I den tryckta Perussanakirja ges böjningen för uppslagsorden med hjälp av olika modellord med tillhörande koder. Ordboksanvändaren måste själv tillämpa anvisningen på det egna problematiska ordet. I CD-Perussanakirja får användaren där-
emot vid behov fram temaformerna för de uppslagsord som kan böjas, även för sammansatta ord. Finskans komplicerade böjningssystem är alltså inte något problem för det här programmet! Denna egenskap hos programmet är förverkligad med ett lingvistiskt program, det vill säga med språkteknologins hjälp. – I programmet finns knappen Böjning som är avsedd för denna funktion.

5.1 Vilken är genitivformen av nomenet palvelu (tjänst) i plural?

Vissa typer av ord kan ha flera genitivformer i plural. Det väcker förvåning hos en vanlig språkbrukare, och han eller hon vill därför få det bekräftat. Det är ofta fråga om att några personer på en arbetsplats har slagit vad om vilken form som är den rätta, och till slut hör de sig för med språkbyråns telefonrådgivning. Många gånger är frågeställarna inte nöjda när de får svaret att alla har haft rätt, eftersom de vill höra att endast en form är korrekt. CD-Perussanakirja åskådliggör att t.ex. ordet *palvelu* har tre genitivformer i plural: *palvelujen*, *palveluiden* och *palveluitten*.

5.2 Hur uttalar man partitivformen av ordet invaliditeetti (invaliditet) i singular?

Främmande ord är ibland problematiska när det gäller att välja rätt vokal till böjningsändelsen, det vill säga att bestämma om ändelsen skall innehålla ett a- eller ett ä-ljud. I princip finns det regler som man kan följa, men de är inte alltid självklara för vanliga språkbrukare. I ord av den här typen bestämmer de vo-

Marja Lehtinen

kaler som finns i början av ordet vilken slutvokal ordet skall få. Eftersom en av ordets första stavelser innehåller den bakre vokalen a, skall också ändelsen bestå av en bakre vokal. Den korrekta formen är alltså *invaliditeettia*.

5.3 Vilken är imperfektformen av ordet *soutaa* (att ro)?

För vissa tvåstaviga finska verb är både ändelsen *-ti* och *-si* möjliga i imperfekt. Till exempel *soutaa* är ett sådant ord, imperfektformen kan antingen vara *souti* eller *sousi*. I ett sådant fall kan ofta också personer med finska som modersmål behöva få bekräftelse.

Denna egenskap, det vill säga att böjningsformerna för ord anges, hade inte kunnat genomföras utan språkteknologins hjälp. Det är inte en lexikograf som för varje enskilt ord har skrivit ner böjningsformerna i textform, utan detta har genomförts med det finska språkteknologiföretaget Lingsoft Oy:s program. Jag kan trots det intyga att lexikografer verkligen har behövts när det har gällt att testa den här egenskapen i ordboksprogrammet. Resultatet av samarbetet med programmeringsföretaget är att böjningsformerna för uppslagsorden på skärmen är korrekta, de följer alltså språkvårdens rekommendationer.

6 Sökning efter böjda former

Ordboksprogrammet har ytterligare en egenskap där man har dragit nytta av Lingsofts kunnande inom språkteknologin. Den här egenskapen innebär att programmet kan söka upp även böjda former av uppslagsord i CD-Perussanakirja utgående från ordens grundform. Eftersom finskan är ett språk som i hög grad grundar sig på böjda former, har ordboksanvändaren stor nytta av detta drag i ordboksprogrammet. Det finns såvitt vi vet inga andra kommersiella program på finska som har den här egenskapen genomförd på detta sätt. CD-Perussanakirja är i det här avseendet alltså en banbrytare.

I praktiken betyder detta att hela texten i ordboken är analyserad med Lingsoft Oy:s morfologiprogram, och därmed har

varje ord i texten försetts med en grundform, oberoende av vilken form ordet har i själva texten.

6.1 Sökning efter juosta (att springa) i förklaringarna

Ett exempel på en sådan sökning: man kan söka efter ordet *juosta* (att springa) i förklaringarna. Det här är ett sätt att söka efter synonymer, fastän synonymer egentligen inte finns utmärkt i materialet. Vi får 34 träffar, men antalet träffar ökar till 65 om vi dessutom bland tillvalen väljer alternativet "Även böjda former". Bland svaren påträffas då också sådant som hör till detta begreppsområde, om än inte i form av direkt synonyma uttryck. Vid en sådan sökning finner man även böjda former bland träffarna: *juostessa*, *juoksemisesta*, *juokseminen*, *juoksemaan* och så vidare. För alla dessa ord finns det i programmet dold information om att grundformen är *juosta*. När alternativet "Även böjda former" är förkryssat riktar sökningen sig in på de dolda grundformerna, medan de träffar som kommer upp i rutan består av de böjda formerna.

6.2 Sökning efter fraser

Det är speciellt användbart att kunna söka efter böjda former när man söker en fras och inte vet den exakta utformningen, eller åtminstone inte i vilken form lexikografen har tagit med uttrycket i ordboken. Ett exempel på en sådan fras är *nostaa kissa pöydälle*. Det är en fras som vi inte har hittat någon direkt motsvarighet till på andra språk, men uttrycket betyder ungefär att man tar upp en kvistig fråga till grundlig diskussion, det vill säga att man tar tjuren vid hornen (den ordagranna översättningen är att man "lyfter upp katten på bordet"). I sökruan kan man då med plustecken kombinera ifrågavarande ord i grundform, det vill säga *nostaa+kissa+pöytä*, och eventuellt göra några tillval i fönstret Sökning. I detta fall lönar det sig att söka efter information bland exempelsatserna och dessutom bland de böjda formerna av orden.

Marja Lehtinen

7 Det framtida arbetet med ordboken

Till slut några ord om den situation ordboksredaktionen nu befinner sig i och vilka planer den har för nästa version av ordboken. Redaktionen uppdaterar naturligtvis framför allt materialet i ordboken i och med att den tillför nyord, som den breda allmänheten alltid är intresserad av. Ett färskt material är också ett bra försäljningsargument. Inför nästa version utökas materialet med ungefär 2 000 nyord.

Programmet förbättras en aning genom att det kompletteras med komparativ- och superlativformerna till alla 10 000 adjektiv. För detta ändamål har vi redan fått en preliminär lista på böjningarna från språkteknologiföretaget Lingsoft Oy, som fortfarande är vår samarbetspartner. Vi gör också några tillägg till temaformerna för uppslagsorden: nomen skall förses med nominativform i plural, och verb med potentialisform. – En alldeles ny egenskap i programmet är att 20 000 ortnamn i Finland skall tillsättas i form av så kallade bebyggelsenamn och deras korrekta böjning i lokalkasus. Ett problem med finskspråkiga ortnamn är nämligen att det ofta är svårt att veta om de skall böjas i inre eller yttre lokalkasus (det heter till exempel *Helsingissä* 'i Helsingfors', men *Tampereella* 'i Tammerfors').

Den breda allmänheten önskar också att CD-Perussanakirja kunde ge svar på mycket annat, men mer än detta hinner ordboksredaktionen inte genomföra i nästa version. Många skulle till exempel vilja ha lite bakgrundsfakta om normer inom språkvården. Tekniskt sett kunde det lätt genomföras i form av ett separat fönster som öppnades bredvid ordartikeln i de fall där sådan bakgrundsfakta kunde finnas. Men det skulle naturligtvis också kräva att någon skrev texten.

På basis av den respons som användarna har gett blir en tydlig trend synlig: människor som producerar texter skulle vilja att CD-Perussanakirja blev en slags all round-program, en verktygslåda för skribenter. Tekniken utgör inget hinder; problemet ligger i att sådana egenskaper hos programmet krävde omfattande resurser av ordboksredaktionen, det vill säga av dem som ansvarar för produktens innehåll.