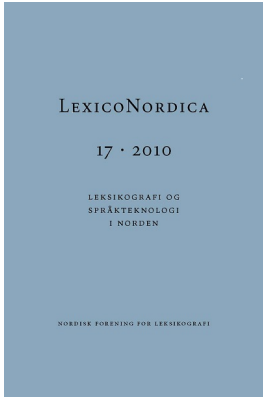


# LexicoNordica

Titel:	Ordbok över karelskan på Internet	
Forfatter:	Ilse Cantell	
Kilde:	LexicoNordica 17, 2010, s. 277-288	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive">http://ojs.statsbiblioteket.dk/index.php/lexn/issue/archive</a>	

© 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.

# Ordbok över karelskan på Internet

*Ilse Cantell*

## 1. Ordboken

Webbversionen av *Karjalan kielen sanakirja* (förk. KKS, sv. Ordbok över karelskan) lades ut på Internet den 15.10.2009. Personerna bakom webbordboken är Marja Torikka som redaktör<sup>1</sup> och Jari Vihtari som planerare av webbgränssnittet. Ordboken ligger på adressen [http://kaino.kotus.fi/cgi-bin/kks/kks\\_etusivu.cgi](http://kaino.kotus.fi/cgi-bin/kks/kks_etusivu.cgi).

Den tryckta ordboken består av sex band och gavs ut 1968–2005. Satu Tanners presentation *En dokumentation av det karelska språket* finns i *LexicoNordica* 13 (2006:169–183). Här tar jag därför bara fram sådant som är specifikt just för webbordboken.

I en intervju som är gjord den 13.5.2005, då den tryckta ordboken kom ut, uttrycker Marja Torikka sin önskan om att ordboken ska digitaliseras. ”Det blir lättare att forska utgående från den. Om man t.ex. har behov av att få tillgång till dialekten i en viss kommun, kan den inte lätt extraheras ur boken, men maskinellt skulle det gå lätt och bekvämt.” (Miikkulainen 2005). Nu har Torikkas önskan gått i uppfyllelse.

Webbversionen av KKS (nedan kallad *Webb-KKS*) är en del av en omfattande samling texter i gratistjänsten *Kaino* på Internet (<http://kaino.kotus.fi/>) som Forskningscentralen för de inhemska språken tillhandahåller för alla som t.ex. i forskningssyfte är intresserade av det finska språket och dess släktspråk, såsom karelskan och samiskan.

---

1 Marja Torikka var redaktionssekreterare för den tryckta KKS fr.o.m. 1988.

## 2. Processen från handsatta boksidor till webbordbok

Överföringen av den karelska ordboken till elektroniskt format har skett något olika beroende på hur gammalt ordboksmaterial det är fråga om.

De tre första banden som utgavs 1968–1983, handsattes i tiden på traditionellt sätt. Materialet i dessa tre band, dvs. bokstavsintervallet A–N har först skannats från de tryckta sidorna. De skannade filerna har sedan gått igenom en textigenkänningsprocess (OCR), dvs. överföring från bild till text med typografisk formatering.

De tre sista banden, bokstavsintervallet O–Ö kom ut under åren 1993–2005 och för deras del var manuskripten skapade och sparade i elektronisk form som formaterade WordPerfect-filer. Efter textigenkänningen förelåg de äldre och de nyare manuskripten i samma form, och resten av processen var identisk för dem alla. Den fortsatta processen bestod i en konvertering till XML-kodad och strukturerad data och sedan en rent teknisk konvertering så att ordbokens lexikografiska kategorier presenteras som de ska i det för ordboken planerade användargränssnittet. Användargränssnittet och de olika sökfunktionerna presenteras i avsnitt 4 nedan.

Innehållet i de sex banden är så gott som oförändrat med undantag av några korrigerade felaktigheter och, för det första bandets del, ett visst förenhetligande som var nödvändigt för att artiklarna ska motsvara redigeringsprinciperna i de senare banden.

## 3. Hur *Webb-KKS* ser ut och fungerar

### 3.1. Tekniska krav på utrustningen

Att olika operativsystem fungerar olika är naturligtvis en utma-

ning för planeringen av en webbordbok och det är troligen omöjligt att gardera sig mot all variation i fråga om system och inställningar som användarna kan ha.

Materialen innehåller en mängd specialtecken och för Kaino är alla tecken kodade enligt standarden Unicode. De många specialtecknen i Webb-KKS återges enligt användarinstruktionerna bäst om användaren har gratisfonten Charis SIL installerad. Fonten finns för både Windows, MacOSX och Linux.

## 3.2. Användargränssnittet

### 3.2.1. Megastruktur och design

*Webb-KKS* är till sin megastruktur en hel webbplats, vars delar dels innehåller material från omtexterna i den tryckta ordboken, dels sådant material som är specifikt för webbordboken.

Förutom den egentliga ordboksdelen ingår i *Webb-KKS* en presentation av den tryckta ordboken och dess relation till webbordboken, instruktioner för sökningar och ett pdf-dokument som innehåller den tryckta ordbokens inledning på ca 100 sidor. Författare till inledningen är ordbokens första huvudredaktör Pertti Virtaranta. I den tryckta ordbokens inledning ingår en hel del foton som inte ännu finns med i inledningen till *Webb-KKS*. Det är synd men troligen kan de skannas in inom en snar framtid. I fotona ingår mycket kulturhistoria som via publicering kunde få spridning via Internet.

Bland materialet på webbplatsen finns en karta över det språkområde ordboken täcker och en förteckning av dialekterna (som förkortningar och utskrivna) grupperade och markerade med olika färger<sup>2</sup>. De nordkarelska dialekterna är markerade med mörkblått, de sydkarelska med rött och de olonetsiska dialekterna med grönt. Dessa färger återkommer i de geografiska markörerna i ordartiklarna.

---

2 Kartan kan ses på adressen <http://kaino.kotus.fi/kks/kkskartta.html>.

Som ett interaktivt element finns i webbordboken en länk till ett formulär för respons från användarna. Arbetet med att utveckla sökmöjligheterna i webbordboken pågår fortfarande på Forskningscentralen för de inhemska språken. Avsikten är att bl.a. göra det lättare att skriva in dialektala sökord med specialtecken eller diakritiska tecken.

*Webb-KKS* har en grafisk design som anknyter till den tryckta ordboken genom att varje webbsida (med undantag av den kareliska dialektkartan) till vänster i ordboks-fönstret har ett fält som ser ut som ryggen på den tryckta ordboken.

I användarinstruktionerna har man varit mycket noggrann och sett till att användaren får all möjlig hjälp. Man varnar också för vissa brister som ännu inte åtgärdats. Bl.a. påpekar man att (exempel)sökningarna kanske inte alltid ger ett fullständigt data-underlag. I sådana fall instrueras användarna att klicka sig fram till de enskilda artiklarna där de kan komplettera resultatet. En förbättring av detta är under arbete och motiverad.

### 3.2.2. Ordboks-fönstret

Ordboks-fönstrets högra två tredjedelar upptas i själva ordboken av ett sökformulär med var sin flik för de två olika sökmöjligheterna artikelsökning och exempelsökning. Under formuläret ses sökresultaten.

Ett röd-svart-guldfärgat fält återkommer på webbplatsens andra sidor och länkar till webbplatsens andra delar är placerade i fältet. Samma länkar finns också överst i ordboks-fönstret. Eftersom ordboken är en vetenskaplig ordbok och i mitt fall en ordbok över ett språk jag inte behärskar är behovet av användarinstruktioner åtminstone i början rätt stort. Dessutom går det att konsultera en snabbhjälp som öppnas som ett extra-fönster och finns bakom en länk på sökformuläret.

En artikel i *Webb-KKS* är indelad i avsnitt med radbyten. Detta strukturerar artikeln tydligare än vad fallet är med den tryckta ord-

boken där det inte har varit möjligt på grund av det naturliga kravet att spara utrymme. Uppslagsordet med ordklassmarkör finns på en egen rad, och därefter står betydelsenumren på egen rad om de åtföljs av ett exempel eller en dialektvariant av uppslagsordet.

De dialektala varianterna av uppslagsordet och exempelfraser-na finns sedan i ett eget, ofta ganska stort block som det också gärna kunde ha varit indelat i kortare avsnitt. Men någonstans måste ju gränsen dras och risken finns att man i stället för strukturering får splittring.

Exemplen är i stället separerade med lodstreck och efter ett exempel finns en geografisk markör för socknen där exemplet är upptecknat. Markörerna bär färgkoder och är i de flesta fall förkortade. Då man för kursorn till en sockennamnsförkortning, öppnas med hjälp av mouseover-funktionen ett mycket litet extrafönster med det utskrivna sockennamnet. Att också de sockennamnen som inte alls är förkortade är försedda med ett extrafönster med exakt samma text är onödigt, men hellre så än inga extrafönster alls. I stället kunde också andra förkortade markörer ha försetts med utskrivna form i ett extrafönster.

### 3.3. Sökningar

#### 3.3.1. Sökning i ordlistan och sökning med hjälp av sökformulär

Då man går in i ordboken ser man i ordboksfönstrets vänstra del ett röd-svart-guldfärgat fält vars synligaste del är en lista med 21 uppslagsord i alfabetisk ordning. Har man redan gjort en sökning med hjälp av sökformuläret står det sökta ordet mitterst i listan. I det röd-svart-guldfärgade fältet finns nederst länkar till var och en av bokstäverna enligt vilka ordbokens uppslagsord är alfabetiserade för att man snabbt ska komma fram i listan. Allra nederst finns en sökruta för att ta fram önskat ord i listan. Ordlistans ord är länkade till ordartiklarna.

Egentliga sökningar i ordboken kan göras på artikelnivå eller som fritextsökningar på exempelnivå. I sökformuläret på skärmen väljer man en av två flikar, den ena för artikelsökning och den andra för exempelsökning. Det är möjligt att göra sökningar med trunkeringstecken. Webb tjänstens instruktioner för sökningar i *Webb-KKS* är mycket ingående, vilket gör att man till fullo kan lära sig utnyttja alla de möjligheter ordboken erbjuder.

### 3.3.2. Artikelsökning

Vid artikelsökning kan sökelementet vara uppslagsord, ordklassangivelse, förklaring eller finsk översättning, dialektord, socken, källa, bruklighetsmarkör och grammatiska uppgifter. I en del av sökfälten kan sökelementet väljas i en rullgardinsmeny. Detta är en bra hjälp mot onödiga sökmissar på grund av felskrivning.

Min viktigaste invändning gäller det komplicerade i att skriva dialektala sökord och uppslagsord vid artikelsökning. Nu kräver en smidig sökning antingen rätt mycket träning eller förhandskunskaper av användaren. Annars är man för uppslagsordens del hänvisad till sökning i uppslagsordsbalken genom att klicka på begynnelsebokstäverna och sedan skrolla fram till rätt ord.

Uppslagsorden i den karelska ordboken är lemmatiserade i nordkarelsk form och därför är det också motiverat med ett sökfält för dialektord. Att skriva det ord man söker är en utmaning för en icke karelskunnig. Nordkarelskans ortografi innehåller en del tecken som inte går att producera med ett vanligt (skandinaviskt) tangentbord. Därför kan ersättande tecken användas i sökfältet ”uppslagsord”. Bokstäver med diakritiska tecken ersätts med bokstäver utan sådana: i stället för ett *d* med akut accent skrivs i sökfältet *j* eller *t* och i stället för *š* skrivs *ts*. Däremot kan det också i finsk ortografi förekommande tecknet *š* skrivas i sökfältet. Detta tecken finns dock inte på ett finskt tangentbord. Man måste alltså ta till olika knep för att åstadkomma *š*, t.ex. kopiering från en Word-text.

Lemmatiseringsprinciperna<sup>3</sup> finns förklarade i omtexterna, men trots det kan det bli svårt att göra en sökning på uppslagsordet. En användare som kan finska kan ju alltid ”fuska” lite och gissa sig till en finsk översättning, söka på den och sedan bläddra bland sökresultaten. Alla elektroniska ordböcker kan ju användas i motsatt riktning än vad de ursprungligen är tänkta för, men den metoden ger inte optimalt sökresultat.

Avancerad sökning på en söksträng med avseende på t.ex. socken och bruklighet ger till resultat alla de artiklar där sökelementen förekommer. Detta fungerar dock inte optimalt utan på samma sätt som t.ex. en Google-sökning utan citattecken, dvs. de två sökelementen står inte nödvändigtvis nära varandra. Jag sökte med sökelementen sockennamnet *Oulanka* och brukligheten *itkuvirsissä* (‘i gråtkväden’). Sökningen gav nio träffar. En av träffarna, artikeln *tulla*, innehåller ett exempel från *Oulanka*, men det är inte från ett gråtkväde. Däremot innehåller artikeln ett exempel från ett gråtkväde, men detta är upptecknat i socknen *Tulemajärvi*. Det återstår alltså för användaren att gallra i sökresultatet för att eventuellt få fram det han eller hon söker. I detta fall finns det inte i materialet några belegg från gråtkväden som skulle vara upptecknade i *Oulanka*. Det är alltså inte möjligt att söka på två olika sökelement och dessutom lägga in en begränsning på avståndet mellan dem i artikeln. Sådana sökningar kräver att man kommer åt ordbokens databas som text. Ett annat alternativ vore att ordboksprogrammet först plockar ut det man sökt med ett sökelement och sedan möjliggör en sökning till i detta resultat med ett annat sökelement.

---

3 Eftersom uppslagsorden i ordboken ges i sin nordkarelska form, kan ett dialektord man träffat på i en text och vill slå upp skilja sig från lemmaformen i fråga om diakritiska tecken för suprasegmentala drag, i fråga om vissa konsonanter, speciellt klusiler som i nordkarelskan saknar stämbandston och slutligen i fråga om vissa diftonger.



### 3.3.3. Exempelsökning

Vid exempelsökning kan sökelementet vara dialektord, förklaring eller finsk översättning, socken, källa, bruklighetsangivelse och grammatiska uppgifter. Nederst i formuläret finns två luckor som kan användas för avgränsning till en viss artikel enligt uppslagsordet och ordklassen. Den senare avgränsningen kan användas t.ex. om man vill skilja mellan adjektiv och adverb. Vid exempelsökning kan trunkeringstecken användas bara då man söker på uppslagsord.

Vid exempelsökning ges sökresultatet i form av en konkordansliknande lista där raderna börjar med uppslagsordet i den artikel där exemplet i fråga finns. Uppslagsordet står inom hakparenteser och är en länk till hela artikeln. Skillnaden till sökresultat i konkordanser är den att söksträngen här inte centreras, däremot är den markerad med gult. Den gula markeringen försvinner dock om man går till artikeln via länken på radens början. Detta är inte riktigt bra, men man kan ju alltid använda webbläsarens egen sökfunktion då man har tagit fram själva artikeln.

En exempelsökning med avgränsning till uppslagsord kan kännas irrelevant, men i fråga om mycket långa artiklar är en sådan sökning användbar. T.ex. i artikeln *tulla* ('komma') ingår 338 exempel. Det är lättare att få en överblick över exemplen då de är uppställda som i en konkordans. Om man bland dem vill söka de exempel som är upptecknade i Tulemajärvi, gör man en avancerad sökning på sökelementen uppslagsord och socken kombinerade och får fram artikelns alla sex exempel som är upptecknade just i Tulemajärvi. Här går det alltså att söka på det sätt som efterlystes ovan.

## 4. Vad kunde ytterligare önskas?

Då man tar fram elektroniska ordböcker, framför allt webbordböcker kan man praktiskt taget strunta i kravet på att spara utrym-

me. Den elektroniska ordbokens fördelar kommer på många sätt fram då man använder *Webb-KKS*. En egenskap hos elektroniska ordböcker har dock inte utnyttjats och det är möjligheten att visa respektive dölja element i ordboksartiklarna.

Ett sökresultat kan vara en mycket lång artikel som kan vara besvärlig att läsa då man har hela texten på skärmen. Genom att visa respektive dölja element får användaren fram det han eller hon verkligen är intresserad av. Den som bara är intresserad av t.ex. ett ords betydelser och kanske de finska översättningarna, kunde tänkas vilja gömma exemplen och vice versa. Utgående från databasens nuvarande kodning är detta troligen helt möjligt. Kanske en idé för det fortsatta utvecklingsarbetet med användargränssnittet? Så kunde denna webbordbok också bli en ”andra generationens” ordbok. Ett gott exempel på en sådan är *Oxford English Dictionary*s webbversion *OED Online*.

## 5. Avslutning

*Webb-KKS* är en ”första generationens” webbordbok och är som sådan jämförbar med *Svenska Akademiens Ordbok* på Internet. Ett liknande utvecklingssamarbete mellan lexikografer, dataplanerare och programmerare ligger troligen bakom de båda storverken. Samarbetet har burit en vacker frukt i och med att *KKS* nu finns gratis tillgänglig för forskare och den stora allmänheten. Detta ökar demokratin inom kunskapens område i och med att ordboken kan användas också där det inte finns tillgång till den tryckta ordboken. Utgivningen på Internet garanterar också en internationell spridning. Många av den karelska ordbokens användare finns t.ex. i Ryssland.

Informationen i det kodade manuskriptet till ordboken ser ut att vara så gott som maximalt utnyttjad. Arbetet är så lyckat att den i dessa sammanhang än så länge något oortodoxa tanken

infinner sig att den tryckta ordboken förlorar en hel del i jämförelsen med webbordboken. Å andra sidan hade det naturligtvis varit synd att avbryta utgivningen av en bokserie som redan funnit sin väg till universitet, bibliotek och också till ett antal privata bokhyllor. Diskussionen omkring publiceringen av några andra omfattande ordböcker som fortfarande är under arbete vid Forskningscentralen för de inhemska språken pågår dock som bäst. I framtiden kommer den elektroniska utgivningen att vara självklar. Jag ser fram emot att se också andra av forskningscentralens ordböcker på Internet, helst så att de alla kan konsulteras med hjälp av samma sökformulär.

Problemet med de svårskrivna sökorden står på arbetslistan då man vid Forskningscentralen för de inhemska språken utvecklar användargränssnittet för *Webb-KKS*. En del förbättringar har redan gjorts under våren 2010 medan detta skrivs. Vid exemplsökning på dialektord kan man nu i stället för en del tecken i den fonetiska skrift som används i ordboken skriva tecken som kan åstadkommas på ett vanligt tangentbord. T.ex. tecknet för ett muljerat *d* (*d* + akut accent) kan i sökningarna skrivas *d'* (*d* + vanligt accenttecken). Avsikten är att åstadkomma ett hjälpfönster där man kunde klicka fram specialtecken och bokstäver försedda med diakritiska tecken. Det kunde dessutom vara bra med ett extrafönster med en lista över omskrivningarna av vissa bokstäver och bokstavskombinationer vid lemmatiseringen. Extrafönstret kunde t.o.m. vara optionellt så att man kan klicka bort det om man tycker att man inte behöver det.

Men för *Webb-KKS* (och för *SAOB*) återstår ännu att ta steget fullt ut till att vara en ”andra generationens” webbordbok där delar av omfattande ordboksartiklar kan lyftas fram beroende på vad man är intresserad av.

Till sist vill jag kommentera webbordbokens namn. Varför kalla webbordboken för en version? Ordböcker utarbetas numera som databaser som sedan kan ges ut på olika plattformar beroende

bl.a. på hur man vill presentera databasens innehåll och vilka möjligheter till konsultering av ordboken (slå upp vs. söka) man vill erbjuda sina användare. Arbetsgången med *Webb-KKS* har varit något mera komplicerad än detta och visst har den tryckta ordboken varit primär eftersom elektroniska ordböcker knappast var i tankarna då ordbokens första band kom ut 1968. Men i dagens läge är både den tryckta och den elektroniska ordboken ”versioner” och då kanske vore förslagsvis namnet *Karjalan kielen verkkosanakirja* (sv. Webbordbok över karelskan) bättre motiverad.<sup>4</sup> På detta sätt tar man inte ställning till huruvida någondera av de två utgivningsformerna är primär eller sekundär. *Webb-KKS* förtjänar i alla fall på inget sätt att bli betraktad som en sekundär version.

## Litteratur

### Ordboken

*Karjalan kielen sanakirja* 1–6. 1968–2005. Utgivare: Finsk-Ugriska Sällskapet; Lexica Societatis Fenno-Ugricae 16 band 1–6, Forskningscentralen för de inhemska språken: Kotimaisten kielten tutkimuskeskuksen julkaisuja 25. Helsinki 1–3, Vammala 4–6.

*Karjalan kielen sanakirjan verkkoversio* (Webbversionen av Ordbok över karelskan). [http://kaino.kotus.fi/cgi-bin/kks/kks\\_etusivu.sgi](http://kaino.kotus.fi/cgi-bin/kks/kks_etusivu.sgi) (mars, april 2010)

---

4 Efter att detta skrivits har *Webb-KKS* uppdaterats den 7.9.2010. Redaktionen har fått möjlighet att ta del av skribentens kommentarer och en del av dem har beaktats. Den mest synliga ändringen är den att webbordboken nu faktiskt heter *Karjalan kielen verkkosanakirja*.

## Övrig litteratur

Tanner, Satu 2006: En dokumentation av det karelska språket. I:  
*LexicoNordica* 13, 169–183.

## Internethänvisningar

*Kaino* <http://kaino.kotus.fi/> (mars, april 2010)

Miikkulainen, Raisa 2005: *Karjalan kieli innostaa* (Det karelska språket inspirerar). Intervju på Forskningscentralens för de inhemska språken webbplats <http://www.kotus.fi/index.phtml?s=291> (mars 2010)

*Oxford English Dictionary (OED) Online* <http://dictionary.oed.com/>

*Svenska Akademiens ordbok (SAOB)* <http://g3.spraakdata.gu.se/saob/>

Ilse Cantell  
forskare, fil. lic.  
Rödbergsgatan 23 I 197  
FI-00150 Helsingfors  
[il.cantell@pp.inet.fi](mailto:il.cantell@pp.inet.fi)