

Kvensk ordbok på nett – hvem har nytte av den?

Mervi Haavisto¹, Kaisa Maliniemi¹, Leena Niiranen², Pirjo Paavalniemi¹, Tove Reibo³ og Trond Trosterud².

¹ Kainun instituuti – Kvensk institutt, ² Universitet i Tromsø, ³ Halti kvenkultursenter

Hva er kvensk språk?

- Kvensk språk fikk status som et nasjonalt minoritetsspråk i 2005.
- Kvensk er et agglutinerende, morfologirikt språk som er nært beslektet med nordfinske dialekter og med meänkieli i Sverige.
- I dag regner man at det finnes ca 4 000–8 000 brukere av kvensk talespråk.



Kvenske dialekter. Dagens ordbok og analysator bygger på de vestlige dialektene, vi vil ta hensyn til østlige varianter senere.

Korpusplanlegging - utvikling av kvensk skriftspråk

Kvensk standardiseres av Kvensk språkting. En kvensk grammatikk (Söderholm 2012) basert på Språktingets anbefalinger utgjør et viktig grunnlag for utviklingen av digitale verktøy for kvensk.

Tradisjonell tekst (eventyr)	680	80,9 %	67,2 %	36,6 %
Moderne tekst (Ruijan Kaiku)	5 601	70,8 %	54,3 %	31,6 %
Dekningsgrad for ulike tekstsjangre	Totalt	I analysator	I ordbok med analysator	I ordbok uten analysator

Tabell 1

Utvikling av digitale verktøy for kvensk

Kvensk ordbok (Aronsen 2010), men som de fleste ordbøker inneholder den bare hovedformen av ordene. Når kvenske ord blir bøyd blir det i 90% av tilfellene blir det nødvendig å sette finne grunnforma ved hjelp av et analyseprogram. Vi setter ordboka og analyseprogrammet sammen i grensesnittet Neahttagisánit (Johnson et al 2013). Digitaliseringen er også et bidrag til kvensk korpusplanlegging fordi finjustering av språkregler ofte er nødvendig.

Kvensk morfologisk analysator

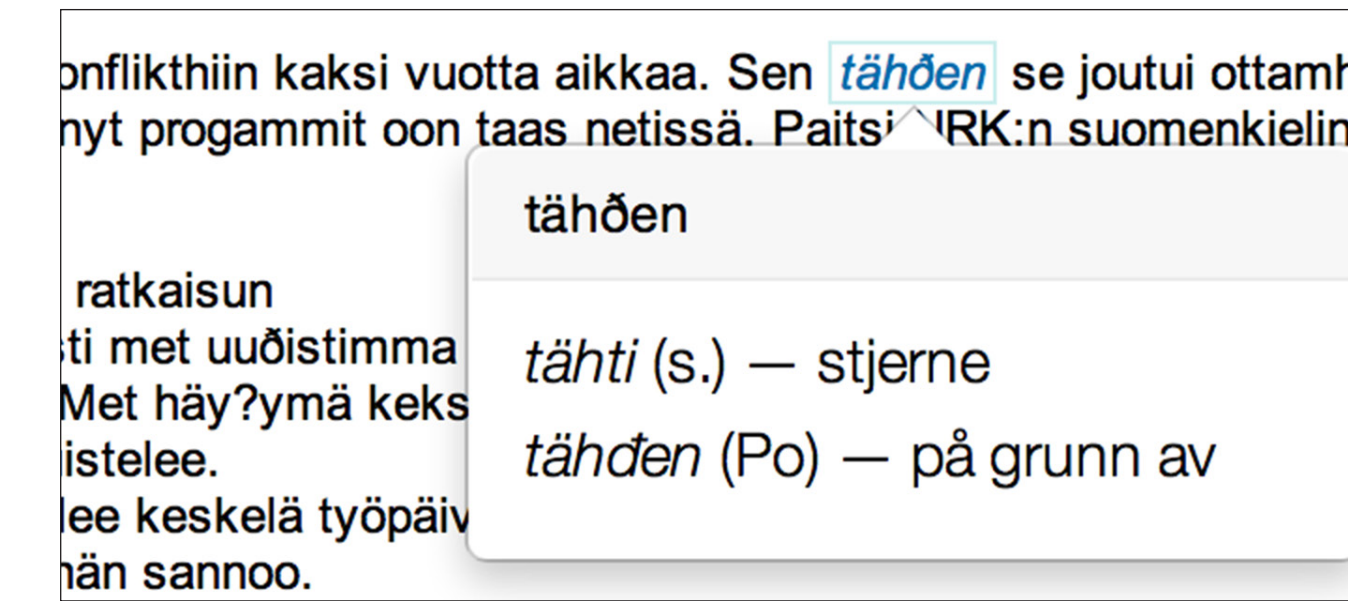
- Per i dag inneholder analysatoren 2 800 substantiv og 1 000 verb.
- Pronomen er i ferd med å bli lagt inn, deretter følger adjektiver.

kala+N+Sg+Par	Grammatisk representasjon.
kala>^A	Stamme + suffiksgrense + fellesvokal for a/ä.
kala>^A	Output fra leksikalsk automat er input til den fonologiske.
kallaa	Vokalharmoni: ^A -> a etter a. Geminering: Dobbel ll foran lang aa.

Bilde 1: Hvordan analysatoren fungerer.

Framtidige muligheter for analysatoren

- Interaktiv språklæring
- Retteprogram
- Terminologiutvikling
- Språkforskning



Bilde 2: Løpende tekst med ordklikk og popup-vindu fra ordboka.

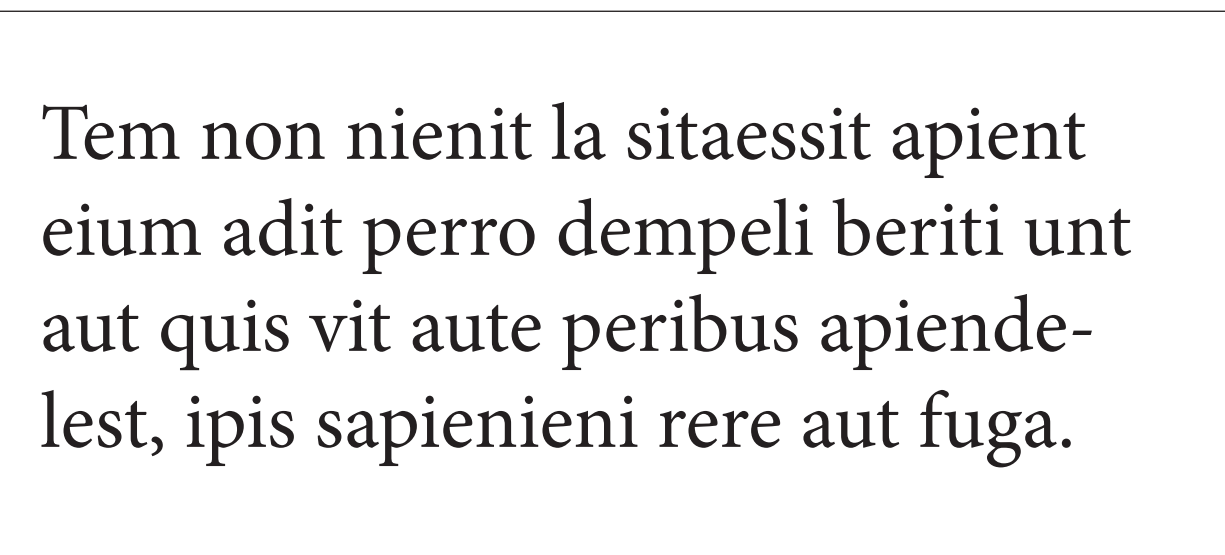
Hvordan kan en digital ordbok gjøres effektiv?

Selv om analyseprogrammet dekker bøyningen av få ord, kan den allerede hjelpe lesere til å forstå ord i løpende tekst. Ordboka håndterer også dynamiske sammensetninger, dvs. sammensetninger ordboka ikke har sett før.

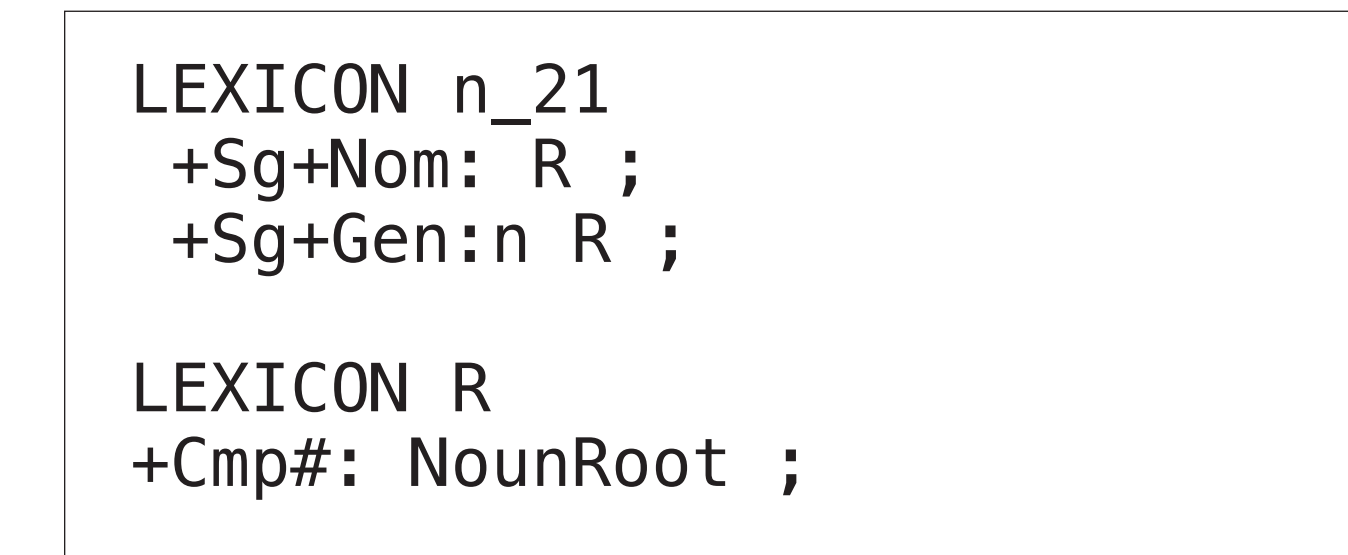
En morfologisk analysator gjør det mulig å behandle også ortografisk variasjon: Den kvenske analysatoren aksepterer d for đ (saada/saada ‘å få’) og æ/ø for ä/ö (mäki/mäki ‘bakke’).

Bare grunnformer	32%
Med fst, uten dynamisk sammensetning	58% (50?)
Med fst, med dynamisk sammensetning	58% (58?)
Spellrelax	(nn %)

Tabell 2



Bilde 3. Dynamisk sammensetning i praksis.



Bilde 4. Dynamisk sammensetning i analysatoren.

Statusplanlegging - utvidet bruk av kvensk

Ordboka vil være nyttig for språkbrukere som kan snakke kvensk, men som ikke er vant til å skrive.

De som vil ha størst utbytte av en digital kvensk ordbok er studenter som vil lære seg kvensk og ta språket i bruk. Denne gruppa mangler både ordforråd og grammatisk kunnskap om hvordan ordene forandrer seg. Ordboka vil derfor gi dem verdifull hjelp til å kunne forstå kvenske skrifter og ikke minst til å kunne produsere egne skriftlige tekster. Ordboka bidrar til kvensk statusplanlegging fordi ordboka er med på å utvide bruken av kvensk i samfunnet.

Brukerlogg

Brukerlogg (N=691) gir informasjon om søkeord, søkeretning og treff.

Kvensk-norsk	658	545	112	82,8%
Norsk-kvensk	33	19	14	57,6 %
Oppslag	Total	Ja	Nei	Suksess

Tabell 3: Oppslag i ordboka og suksessrate.

Ord i ordboka	Ord ikke i ordboka
Varpu, tulla, tamppu, saada, jänkkä, astii, vikaatet, veitti, vantut, traasu, tervet, apu, väki, varsi, valas, usko-nu, tyyris, turvet, savi, saita, ikke, jeg, tro, kvänsk, jente, jakt, hus, hagl, går, gikk, fjell,fangst, dåp, bøker, båt, bjørn.	Vikatet, valokuvvii, stre, sisältää, rautio, myö, kampela, asemasta, ä/Øti, äidinkielenopetusta, yllättynny, yksinäinen, vuoli, vikateinen, vikate, verto, veittinen, vasen, vanhoita, vaijeri, rae, raato, simppa, rauttii, olla, litteratur, kvensk, kluva, kakt, dåpe, dope, båten.
Kvensk og norsk: Grunnord-forrådet og ord for tradisjonell kultur.	Kvensk: Skrivefeil, sammensatte ord, lakuner. Norsk: Kvenske ord, lakuner, skrivefeil, tekniske feil i ordboka (båten).

Tabell 4. De 20 vanligste kvenske og norske oppslagsorda i og utenfor ordboka.

Konklusjon

En digital ordbok er ikke bare en del av korpusplanlegginga, men også en del av statusplanlegginga av kvensk. En grammatisk analysator, selv med dårlig dekningsgrad, bidrar til ei god elektronisk resepsjonsordbok.

Litteratur

Aronsen, Terje 2010: Kvensk-norsk-kvensk elektronisk ordbok. Redigert av Verena Schall og Trond Trosterud. Universitetet i Tromsø.
Johnson, Ryan, Lene Antonsen, Trond Trosterud 2013: Using finite state transducers for making efficient reading comprehension dictionaries. Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013), May 22–24, 2013, Oslo University, Norway.
NEALT Proceedings Series 16: 59–71.
Söderholm 2012: En liten kvensk grammatikk. Vesta – Porsanger – Nordreisa -varianten. Ms, UiT.

