

Kvensk ordbok på nett – hvem har nytte av den?

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

¹Kainun institutti – Kvensk institutt, ²UiT Norges arktiske universitet, ³Halti kvenkultursenter

Hva er kvensk språk?

- Kvensk er et agglutinerende, morfologirikt språk som er nært beslektet med nordfinske dialekter og med meänkieli i Sverige.
- Kvensk språk fikk status som et nasjonalt minoritetsspråk i 2005.
- 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 språktning. En kvensk grammatikk (Söderholm 2012) basert på Språktingets anbefalinger utgjør et viktig grunnlag for utviklingen av digitale verktøy for kvensk.

Tradisjonell tekst (eventyr)	680	80,6 %	64,4 %	25,7 %
Moderne tekst (Ruijan Kaiku)	5 633	70,0 %	52,9 %	33,4 %
Dekningsgrad for ulike tekstsjangre	Totalt	I analysator	I ordbok med analysator	I ordbok uten analysator

Tabell 1. Ordbokas dekningsgrad for to teksttyper.

Digitale verktøy for kvensk

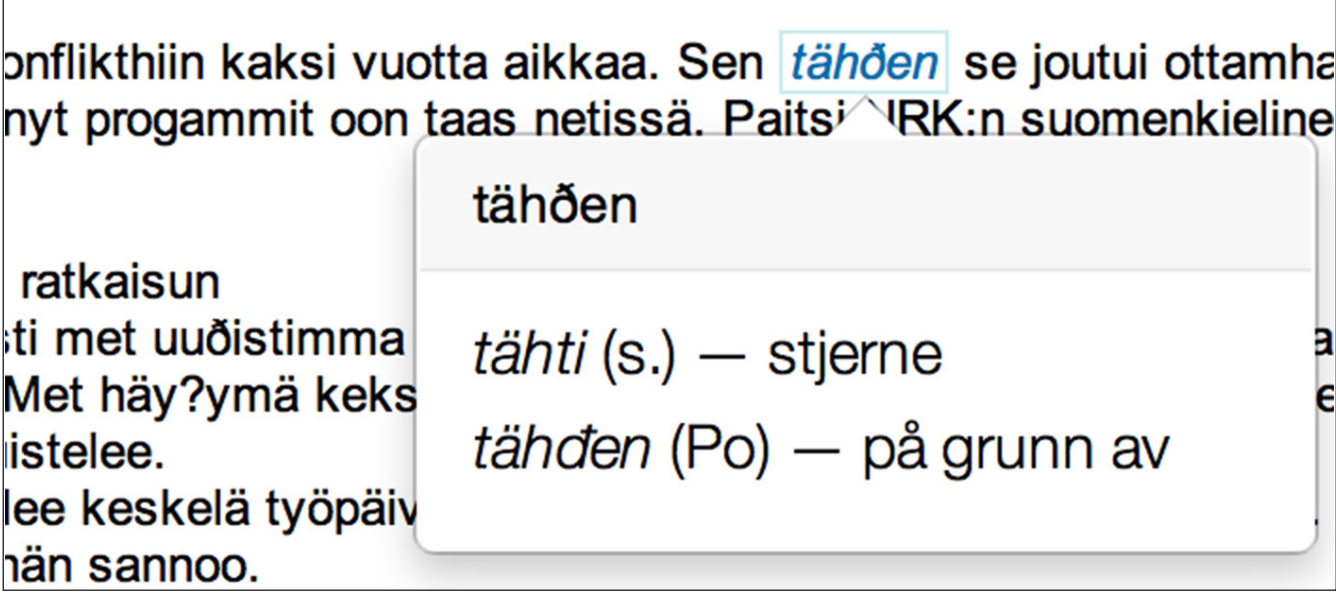
Det fins en kvensk ordbok (Aronsen 2010), men som de fleste ordbøker inneholder den bare hovedformen av ordene. Når kvenske ord blir bøyd i 90% av tilfellene blir det nødvendig å finne grunnforma ved hjelp av et analyseprogram. Vi setter ordboka og analyseprogrammet sammen i grensesnittet Neahttagisánit (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 + ^A (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.



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

Framtidige muligheter for analysatoren

- interaktiv språklæring
- retteprogram
- terminologiutvikling
- språkforskning

Hvordan kan en digital ordbok gjøres mer effektiv?

Selv om analysatoren ikke er ferdig, kan den allerede hjelpe leseren til å forstå ord i løpende tekst. Analysatoren håndterer også dynamiske sammensetninger, dvs. sammensetninger ordboka ikke har sett før, og den gjør det også mulig å behandle ortografisk variasjon: Den kvenske analysatoren aksepterer *d* for *đ* (saada/saada ‘å få’) og *æ/ø* for *ä/ö* (mæki/mäki ‘bakke’).

Ortografisk variasjon	Dynamisk sammensetning	%
Nei	Nei	54,2 %
Ja	Nei	54,6 %
Ja	Ja	55,9 %

Tabell 2. Dekningsgrad med og uten ortografisk variasjon og dynamisk sammensetning. Forskjellene er små fordi teksten inneholder få dynamiske sammensetninger og lite ortografisk variasjon.

Ruijan Kaiku numero 5-2013 tuli ulos jo juunikuussa! Siinä muistelhaan mm. - Paaskiviikosta - kainun kielestä, musikista ja teatterista - kahesta sanakirjasta - Kveenimaayhistyksestä - Kilpisjärven uutisii Tillaa sieki <i>paperiaviisi</i>	<i>paperiaviisi</i> <i>aviisi</i> (s.) — avis <i>paperi</i> (s.) — papir
LEXICON n_21 +Sg+Nom: R ; +Sg+Gen:n R ; LEXICON R +Cmp#: NounRoot ;	

Bilde 3. Dynamisk sammensetning i praksis.

Bilde 4. Dynamisk sammensetning i analysatoren.

Statusplanlegging - utvidet bruk av kvensk

Ordboka er nyttig for språkbrukere som kan snakke kvensk, men ikke skrive det. Størst utbytte av en digital kvensk ordbok har de som vil lære seg kvensk og ta språket i bruk, og som mangler både ordforråd og kunnskap om ordbøyninger. Ordboka vil derfor gi dem verdifull hjelp til å kunne forstå kvensk og ikke minst til å kunne produsere egne tekster. Fordi ordboka er med på å utvide bruken av kvensk i samfunnet, bidrar den til statusplanlegginga for kvensk.

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

Tabell 4. De 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 for kvensk. En grammatisk analysator, selv med dårlig dekningsgrad, bidrar til ei fungerende elektronisk resepsjonsordbok. For morfologirike språk spiller intelligente ordbøker som denne ei sentral rolle i språkrevitaliseringa.

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, Eira 2012: En liten kvensk grammatikk. Vesta – Porsanger – Nordreisa -varianten. Ms, UiT.