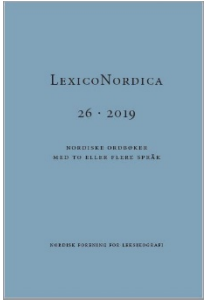


LexicoNordica

Titel:	Kva bruker vi minoritetsspråksordbøker til? Ein studie av brukarloggane for tolv tospråklege ordbøker	
Forfatter:	Trond Trosterud	
Kilde:	LexicoNordica 26, 2019, s. 177-198	
URL:	https://tidsskrift.dk/index.php/lexn/issue/archive	

© 2019 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.

Kva bruker vi minoritetsspråksordbøker til? Ein studie av brukarloggane for tolv tospråklege ordbøker

Trond Trosterud

The article analyses the user logs of twelve different bilingual minority language dictionaries. The logs show that the dictionaries are in frequent use, with lookup in the direction towards the minority language (production use) dominating by approximately 60% of the total queries. Query recall is high, around 90%, despite the dictionaries being medium-size to small, and most failed queries are due to orthographical errors. The dictionaries translating into the minority language are used for text production to a varying degree. This is clearly shown in the profile of the most commonly used query words.

1. Introduksjon

Artikkelen analyserer brukarloggen til tolv ulike e-ordbøker under ordbokportalen Neahttadigisánit for til saman fem ulike minoritetsspråk: nordsamisk, sørsamisk, enaresamisk, skoltesamisk og kvensk.

Kapittel 2 presenterer kort dei fem minoritetsspråka og andre tilgjengelege e-ordbøker for desse språka. Kapittel 3 presenterer ordbokportalen Neahttadigisánit og dei ordbøkene som blir hand-sama i denne artikkelen. Kapittel 4 viser korleis bruk av ordbøkene blir loggført, og gjev ein analyse av loggdata for dei tolv e-ordbøkene frå 2018, med vekt dels på dei mest frekvente oppslagsorda og dels på karakteristiske drag ved dei mislukka søka. Til slutt kjem ein konklusjon.

2. Bakgrunn

2.1. Dei fem minoritetsspråka som blir handsama

Temaet for artikkelen er ordbøker for fem ulike minoritetsspråk i to ulike land, Noreg og Finland. Tabell 1 viser kor mange talarar dei ulike språka har ifølgje Ethnologue, og kor livskraftige dei er målt etter EGIDS-skalaen. EGIDS-verdiane i tabellen er som følgjer: 2 = Provincial, 6b = Threatened, 7 = Shifting. For nærare presentasjon av skalaen, sjå EGIDS; for ei drøfting, sjå Lewis & Simons (2010).

Språk	Noreg			Finland		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Talarar	20 000	500	5000	1700	300	300
EGIDS	2	6b	7	6b	2	7

Tabell 1: Talarar og poeng på EGIDS-skalaen (Kjelde: Ethnologue).

Sør- og nordsamisk er vestsamiske språk, mens enare- og skolte-samisk er austsamiske. Som grannespråk er likevel nord- og enaresamisk meir eller mindre innbyrdes forståelege. Sør- og skolte-samisk ligg lenger unna. Alle dei fem språka har ulike ortografiske prinsipp. Sørsamisk har digrafar etter mønster av norsk, mens dei andre samiske språka bruker diakritiske teikn. Kvensk har same ortografiske prinsipp som finsk, men skil seg frå finsk både morfologisk og leksikalsk. Dei ortografiske prinsippa for enaresamisk er i hovuddrag dei same som for finsk, men også her er det skilnader. Det er med andre ord ikkje mogleg å skrive dei samiske språka med ortografiske reglar frå dei ulike majoritetsspråka.

2.2. Andre e-ordbøker

I tillegg til e-ordbøkene som blir analysert her, har brukarane for eitt av språkpara tilgang til fleire e-ordbøker.

For nordsamisk er den største e-ordboka *Nordsamisk-norsk-nordsamisk digital ordbok* på nett (NNNDO). Den er gjeven ut av forlaget Davvi girji og har 51 000 lemma i kvar retning. Denne e-ordboka bygger på papirordbøkene SNO (1995) og SNSO (2000), men inneheld berre lemma og omsetjing, ikkje forklaringsane eller eksempelsetningane frå papirordbøkene.

Det finst Wiktionary-ordbøker for fleire samiske språkpar, men berre nordsamisk har fleire enn 1000 lemma (til fransk 220 129, til engelsk 3867, til finsk 3060¹). To andre e-ordbokportalar tilbyr også ordbøker mellom nordsamisk og norsk bokmål, Glosbe og DinOrdbok. Desse ordbøkene inneheld primært leksikografisk materiale frå Wiktionary eller frå Neahttadigisánit. DinOrdbok viser i tillegg eksempelsetningar frå eit parallellkorpus.²

For språkparet nordsamisk–norsk finst det med andre ord alternativ til e-ordbøkene til Universitetet i Tromsø – Noregs arktiske universitet (UiT). For dei andre språkpara som blir handsama her, finst det ikkje andre e-ordbøker enn Neahttadigisánit.

3. UiT sine samiske og kvenske ordbøker

3.1. E-ordboka Neahttadigisánit

Forskningsgruppa for samisk språkteknologi ved UiT, Giellatekno, har det seinaste tiåret laga ein infrastruktur for e-ordbøker for morfologirike minoritetsspråk. Denne infrastrukturen kombinerer ei maskinlesbar ordbok med ein morfologisk analysator og tilgang til eit tekstkorpus. Med utgangspunkt i arbeid gjort av Sjur Moshagen integrerte vi først ei nordsamisk-norsk ordbok i ordbokapplikasjonen for Macintosh (jf. Antonsen et al. 2009), og

-
- 1 Data er henta frå dei respektive sidene på Wiktionary 3.1.2019. Det høge talet for fransk kjem av at kvar bøyingsform har sin eigen lemmaartikkel. Nordsamisk-fransk er likevel den klart største Wiktionary-ordboka for nordsamisk.
 - 2 Parallelltekstane er primært henta frå korpuset OPUS.

seinare for andre operativsystem via ordbokformatet StarDict. Desse e-ordbøkene var basert på ein kompileringsrutine der det for kvar bøyingsform for kvart lemma vart generert ein eigen ordbokartikkel, som inneheldt bøyingsforma sjølv, grunnforma og ein peikar til hovudartikkelen for lemmaet. Denne ordboka, Vuosttaš digisánit («Dei første digi-orda»), kunne lastast ned på alle datamaskinar og mobiltelefonar og brukast utan nettilgang. Deretter laga vi ein ny ordbokinfrastruktur, Neahttadigisánit, eit oppsett der ordboka er tilgjengeleg på ein nettserver. Når brukaren slår opp ei ordform, enten frå ordbokgrensesnittet eller i ein separat klikk-i-tekst-funksjon, blir ordforma sendt til ein morfologisk analysator, som finn grunnforma. Denne blir så sendt til ordboka, som returnerer lemmaartikkelen, med lenkjer til setningseksempel i eit korpus og til ein morfologisk generator som genererer bøyingsparadigmet for kvart lemma.

Dei ulike ordbøkene er tilgjengelege online (sjå under Neahttadigisánit i litteraturlista). I tillegg til den opprinnelege ordboka frå nordsamisk er det no 80 andre språkpar fordelt på 11 ulike ordbokportalar. Neahttadigisánit er presentert i Johnson, Antonsen & Trosterud (2013).

3.2. Ordbøkene

Ordbøkene som blir analysert i denne artikkelen, er nordsamisk↔norsk, nordsamisk↔finsk, sørsamisk↔norsk, enaresamisk↔finsk, skoltesamisk↔finsk og kvensk↔norsk. Referanse til dei ulike ordbøkene står under Neahttadigisánit i litteraturlista.

Nordsamisk↔norsk bygger opphavleg på Nils Jernslettens *Álgosátnegirji* (Jernsletten 1991), men er seinare nesten tidobla i omfang. Ordboka frå norsk til nordsamisk er utarbeidd av Gielatekno. *Álgosátnegirji* var, som namnet ('Nybyrjarordbok') viser, ei ordbok for dei med nordsamisk som andrespråk. Ho inneheldt

det samiske grunnordforrådet og forklarte korleis det vart brukt, for studentar på nybyrjarnivå. Av den grunn var også avvikande bøyingsformer sett opp i lemmalista, med referanse til grunnforma (til dømes står *boadán* ‘eg kjem’ i lemmalista, med referanse til infinitiven *boahtit*). I dei elektroniske ordbøkene er det mogleg å søke på alle ordformer, så tilvisingsformene frå papiordboka er fjerna, men ordboka har framleis eit fokus på L2-brukarar.

Ordboka nordsamisk↔finsk bygger på eit grunnordforråd frå den etymologiske databasen Ålgu ved Forskningscentralen för de inhemska språken i Finland, men er seinare utvida. Spesielt gjeld dette frekvente finske ord i den finsk-nordsamiske delen. Sørsamisk↔norsk bygger på leksikografisk arbeid gjort av Albert Jåma og Tove Brustad, publisert i Jåma (2001). Enaresamisk↔finsk bygger på ein tidlegare versjon av Olthuis & Valtonen (2019). Både dei sørsamiske og dei enaresamiske ordbøkene er seinare utvida ein god del. Skoltesamisk↔finsk bygger på Pekka Sammallahti og Jouni Mosnikoff si ordbok *Suomi-koltansaame sanakirja* (1991). Kvensk↔norsk bygde opprinneleg på ordbøker laga av Terje Aronsen (jf. Aronsen 2010) og Eira Söderholm (jf. Söderholm 2012), der ordforrådet er tilpassa ortografien vedtatt av Kvensk språkting. E-ordboka er meir enn dobbelt så stor som dei opprinnelege kjeldene, og det er primært Kvensk institutt som har arbeidd med å utvide ordboka. Jf. Haavisto et al. (2014) for ein presentasjon.

Storleiken på dei ulike ordbøkene går fram av tabell 2 nedanfor. Lmaj viser til dei respektive majoritetsspråka norsk og finsk, og Lmin til eitt av dei samiske språka eller kvensk. Ordbøker i dei ulike retningane mellom to språk er sett opp på kvar sin rad i same kolonne slik at den første raden viser tal for ordbøkene med minoritetsspråk som lemma og den andre raden viser tal for ordbøkene med majoritetsspråk som lemma. Tabellen viser med andre ord at nordsamisk-norsk ordbok har 39 862 lemma, norsk-nordsamisk har 26 021 lemma, og så bortetter.

Lmin	Lmaj = norsk			Lmaj = finsk		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Lemma = Lmin	39 862	11 174	9104	13 175	21 740	33 594
Lemma = Lmaj	26 021	14 754	8599	11 301	17 752	30 217

Tabell 2: Lemma i tolv av Giellateknos nettordbøker (Neahttagisánit), januar 2019.

3.3. Dei morfologiske analysatorane attom ordbøkene

Som referert til ovanfor er det for alle ordbøkene mogleg å slå opp på kva slags ordform som helst. Oppslaget blir sendt til ein morfologisk analysator som gjev både lemma og morfologisk analyse. Lemmaet blir brukt til å slå opp i ordbokartikkelen, og den morfologiske analysen blir presentert for brukaren på sida av sjølve ordbokartikkelen. Alle lemma i alle ordbøkene er inkludert i dei ulike analyseprogramma. Med unntak av språkparet nordsamisk–finsk, der det er mogleg å tenkje seg både samisk- og finskspråklege brukarar, vil brukarane ikkje vere interesserte i morfologisk informasjon om majoritetsspråket (norsk eller finsk). I og med at ordbøkene har ein klikk-i-tekst-funksjon, er ordbøkene likevel utstyrt med ein morfologisk analysator også for finsk og norsk. Brukarar i Noreg er nok i stand til å identifisere delane i eit ord som til dømes *partimedlemmene*, og tilsvarende for brukarar i Finland med det tilsvarende ordet *puoluejäsenet*, men med hjelp av den morfologiske analysatoren kan dei berre klikke på orda, eller lime dei inn i ordboka, og for delane *parti* og *medlem* (finsk: *puolue* og *jäsen*) få ut dei respektive nordsamiske omsetjingane *bellodat* og *miellahttu*. Dermed inneheld loggdata informasjon om både søkeord og lemma, som vist i neste avsnitt.

4. Analyse av loggdata

4.1. Logganalyse som forskingsområde

Herbert E. Wiegand set opp forskning på ordbokbruk som ein av dei fire deldisiplinane innafor metaleksikografien (Wiegand 1998), og det er innafor leksikografien ein rik forskingstradisjon på ordbokbruk. *LexicoNordica* 15 (2008) hadde til dømes ordbokbruk i Norden som tema, og fleire av artiklane der analyserte loggdata. Ordbokbruk kan bli undersøkt via observasjon og via spørjeundersøkingar. Med innføringa av e-ordbøker har ein ny innfallspport til kunnskap om ordbokbruk opna seg: loggoppføring av kvart oppslag gjort i ordboka.

Den mest opplagte måten å bruke loggdata på er til å forbedre lemmatilfanget i sjølve ordboka. Ifølgje de Schryver & Joffe (2004:187) er det få rapportar om slik bruk av loggdata. For utviklinga av Neahttadigisánit har loggdata tvert imot vorte brukt på nett denne måten. Haavisto et al. (2014) er eit godt døme på analyse av loggdata for å forbedre lemmatilfanget i ei ordbok. Då artikkelen vart skriven, inneheldt den kvensk-norske ordboka 4702 lemma, og loggføringa hadde vore operativ i kort tid. Loggen inneheldt berre 691 oppføringar for begge retningane. Likevel kunne loggen vise at sentrale ord som *olla* ‘å vere’ og *vasen* ‘venstre’ mangla i ordboka, og både dei og andre manglande ord kunne bli lagt til. Tilsvarande arbeid har blitt gjort for alle dei andre Neahttadigisánit-ordbøkene.

Med informasjon om den einskilde brukaren er det mogleg å undersøke søkemønster for individuelle brukarar. Almind (2008) inneheld ei drøfting av ulike aspekt ved ordbokbruken som det er mogleg å få ut av ei loggfil. Sentralt for han er i kor stor grad det er mogleg å seie ikkje berre kva det blir søkt etter, men også i kva grad ordboka oppfyller ein spesifikk funksjon (Almind 2008:36). I loggane som blir brukt i denne artikkelen, har vi ikkje tilgang til

den einskilde brukaren, så målsetjinga her er å analysere kva brukarane som kollektiv har søkt etter.

4.2. Loggføring

Bruken av Neahttagisánit blir loggført. Som illustrasjon viser tabell 3 nedanfor fire tilfeldige loggoppføringar, dvs. loggen for fire autentiske ordbokoppslag. Første kolonne gjev forma brukaren har søkt etter. Andre kolonne fortel om søket var vellukka (T, true) eller ikkje (F, false). Tredje kolonne fortel kva lemma (eitt eller fleire) søket får tilslag på. For dei første to oppslaga er det meir enn eitt lemma: Finsk *minä* er tvitydig mellom nominativ av *minä* 'eg' og essiv av *mikä* 'kva', og norsk *for* er fleirtydig mellom verba *fare* og *fore* og mellom substantivet, preposisjonen, adverbet og konjunksjonen *for*.

Fjerde kolonne gjev omsetjing. Der søket ikkje er vellukka (F), viser loggen naturleg nok verken lemma eller omsetjing. Andre opplysningar som kjem fram i lemmaartikkelen (eksempelsetningar, bruksrestriksjonar og så bortetter), er ikkje med i loggen. Dei neste to kolonnane identifiserer språkparet brukaren har søkt i. Siste kolonne viser dato og klokkeslett.

Søk	TF	Lemma	Omsetjing	L1	L2	Dato, tid
minä	T	mikä, minä	mii; mun	fin	sme	2018-12-29 T15:55:22
for	T	fare, fore, for	luoittihit, livkit; illativ...	nob	sme	2018-12-29 T15:55:23
fuotni	T	fuotni	därilig	sme	nob	2018-12-29 T15:55:37
forma- sjon	F			nob	sme	2018-12-29 T15:55:49

Tabell 3: Eksempel på loggoppføringar. Språka er identifisert med ISO 639-3-kodar (SIL International 2007): fin = finsk, nob = norsk bokmål, sme = nordsamisk.

Tabell 4 viser talet på alle loggførte oppslag i dei tolv ordbøkene i 2018 (i perioden 1.1.–29.12.2018). Data for ordbøkene der lemma er på minoritetsspråket, står i raden «Ordbokoppslag på Lmin», og data for ordbøkene der lemma er på majoritetsspråket, står i raden «Ordbokoppslag på Lmaj». Siste rad viser fordelinga språkretningane imellom: 40,7 % av oppslaga i ordbøkene mellom nord-samisk og norsk er til dømes gjort for språkparet norsk–nordsamisk (Lmaj–Lmin). Oppslaga frå nordsamisk til norsk utgjer med andre ord 59,3 % av oppslaga for dette språkparet (i og med at $59,3\% + 40,7\% = 100\%$).

Lmin = ...	Lmaj = norsk			Lmaj = finsk		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Ordbokoppslag på Lmin	953 807	317 452	70 449	247 590	176 662	340 494
Ordbokoppslag på Lmaj	655 928	206 378	42 529	273 459	127 654	222 793
Lmaj / alle oppslag	40,7 %	39,4 %	37,6 %	52,5 %	41,9 %	39,6 %

Tabell 4: Loggførte oppslag i 2018 for dei tolv ordbøkene.

Som det går fram av tabellen, er ordbøkene mykje i bruk. Målt i oppslag per morsmålstalar er bruksfrekvensen langt høgare enn for e-ordbøker på majoritetsspråka. Med unntak av paret finsk–nordsamisk, der oppslag frå finsk til nordsamisk er i knapt fleirtal, går rundt rekna fire av ti ordbokoppslag frå majoritetsspråket til minoritetsspråket.

Loggdøma i tabell 3 viser at det blir loggført kor mange av ordbokoppslaga som ordbøkene kan svare på. Tabell 5 viser kor stor prosentdel av oppslaga som ikkje blir funne i dei ulike ordbøkene (dvs. av tala i tabell 4). Ordbøkene til og frå majoritets- og minoritetsspråka står på kvar si linje, på same måte som i tabell 4.

	Lmaj = norsk			Lmaj = finsk		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Oppslag utan svar						
Ordbokoppslag på Lmin	10,7 %	77 %	10,2 %	12,8 %	11,3 %	2,7 %
Ordbokoppslag på Lmaj	11,8 %	6,8 %	10,8 %	22,4 %	9,1 %	4,5 %

Tabell 5: Prosent av oppslaga som ikkje resulterte i treff.

Êi ordbok peiker seg ut med langt høgare prosent av oppslag utan svar enn dei andre, nemleg finsk-nordsamisk, med 22,4 % oppslag utan treff (tabell 5). Dette korresponderer med storleiken på ordboka: Den finsk-nordsamiske ordboka er berre halvparten så stor som den norsk-nordsamiske (jf. tabell 2) og har dermed også dårlegare treffprosent for omsetjing til same målspråk.

4.3. Analyse av søk som ikkje gjev treff

No kan det vere fleire grunnar til at eit oppslag ikkje finst i ordboka. Êin er at ordet manglar, ein annan er at det er skrive feil. Eg har sett på dei orda som ikkje gav søketreff i ordbøkene, og kontrollert dei opp mot ein stavekontroll for kvart av dei relevante språka. Resultatet går fram av tabell 6, der «Ordbokoppslag med skrivefeil» viser til oppslag som ikkje blir kjent att av stavekontrollar for dei respektive språka.

	Lmaj = norsk			Lmaj = finsk		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Ordbokoppslag med skrivefeil						
Ordbokoppslag på Lmin	73,7 %	63,1 %	89,0 %	56,3 %	83,9 %	74,4 %
Ordbokoppslag på Lmaj	39,5 %	32,1 %	34,8 %	11,6 %	34,0 %	30,8 %

Tabell 6: Feilskrivingsprosent i søkeorda som ikkje fekk treff.

Tabell 6 viser svært stor variasjon. Det mest slåande er at ordbokbrukarane er langt støare (mellom to og fire gonger så stø) i ortografien til majoritetsspråket enn i ortografien til minoritetsspråket. Delar av skilnaden i feilskrivning må også bli sett i lys av treffprosent. Når ein stor del av oppslaga ikkje finst i ordboka, blir feilskrivingsprosenten lågare, som for finsk-nordsamisk, mens der ein låg prosent av oppslaga manglar, er prosenten av skrivefeil høgare, som for finsk-skoltesamisk.

Skilnadene i feilskrivning frå språk til språk er likevel store. Den største feilskrivingsprosenten har kvensk, som også har det yngste skriftspråket, og der avgrensinga mot finsk er vanskeleg å dra. Vi skal òg merke oss at resultatata for andre ordbøker også viser at feilskrivning er den viktigaste kjelda til mislukka søk: For ei undersøking av 500 vilkårleg valde feilsøk i *Bokmålsordboka* viser Nygaard & Fjeld (2008) at så mykje som 46 % kjem av skrivefeil.

Tala for minoritetsspråk er likevel høgare enn for majoritetsspråk. At over 56 %, og i dei fleste tilfella over 70 %, av dei mislukka oppslaga på minoritetsspråket inneheld skrivefeil, bør få konsekvensar for korleis vi presenterer lemma i minoritetsspråklege ordbøker. Eitt mogleg svar på dette problemet finn vi i den norsk-nordsamiske ordboka SNSO. Som vist i Trosterud (2001) har denne ordboka lemmaartiklar strukturert som i ei nordsamisk-norsk ordbok. Det vil seie at ordbokartiklane ikkje er inndelt i tydingar og tydingsgrupper med norske bruksdøme og nordsamiske omsetjingar av dei. I staden listar ordboka opp alle dei samiske omsetjingane alfabetisk. Denne alfabetiseringa kan i lys av brukarloggen presentert her bli sett på som ein strategi for å gje ordbokbrukarane tilgang til ordartiklar for lemma dei ikkje veit korleis dei skal skrive.

I Trosterud (2001) illustrerte eg problemet med redigeringsprinsippet for SNSO ved å sitere artikkelen for verbet *gå*. *Gå* hadde 28 omsetjingar, og desse var altså lista opp alfabetisk. Viss vi heller ser SNSO som ei nordsamisk rettskrivingsordbok, kan sa-

miskspråklege brukarar i Noreg som lurar på korleis *vázzit* blir skrive, slå opp på ordet *gå* og lese seg nedover til V, der dei finn den sentrale omsetjinga, *vázzit*. At det er *gå* dei skal slå opp på for å finne *vázzit*, veit brukarane fordi dei er tospråklege. Fordelen med å slå opp i den norskspråklege ordboka framfor direkte i ei samiskspråkleg ordbok er at ein kjem unna rettskrivingsordbok-paradokset: For å finne ordet du ikkje veit korleis blir stava, må du kunne stave det.

Problemfri er denne metoden likevel ikkje. For enkle ordpar går det fint, men i mange tilfelle er det ikkje opplagt kva norske oppslagsord det samiske ordet ein er interessert i, står under. For å hjelpe e-ordbokbrukarane til å bruke ordbøkene med nord-, enare- og skoltesamisk som L1 har vi bygd inn ein viss toleranse i systemet: I tillegg til den vanlege e-ordboka som blir evaluert her, har Neahttagisániit også ein tolerant inputmodus som aksepterer bruk av grunnbokstavar utan diakritiske teikn, slik at ordboka for kvar av bokstavane *acdnstz* i input undersøker om det å bruke *áčđņšțž* i staden vil gje treff på eit oppslag (til dømes **odas* for korrekt *odas* 'ny'). Oppslag der brukaren vel denne modusen, er elles haldne utanfor analysen i denne artikkelen, men dei utgjer omtrent 9 % av den totale ordbokbruken. 12,6 % av dei vellukka oppslaga som er loggført for nordsamisk ved bruk av den tolerante modusen, ligg utanfor den normative grammatikken, og utan bruk av tolerant modus ville dei ikkje resultert i treff i ordboka. Bruken av ein slik tolerant modus vil med andre ord auke treffprosenten tilsvarande.

Ein spesiell form for toleranse er inkludert i den kvensk-norske ordboka. Kvensk er standardisert i tre ulike dialektar: Porsanger-, Varanger- og Nordreisa-dialekt. Skilnadene er primært morfologiske og morfofonologiske, men dei er svært djuptgripande, og store delar av ordforrådet har systematiske skilnader i bøyingsmorfologien. I ordboka er dette løyst ved at den morfologiske analysatoren inneheld morfofonologien for alle tre dialektane.

Ordboka er i utgangspunktet skriven på Porsanger-dialekt, og den morfologiske generatoren, som genererer ordformene til paradigma som blir presentert, genererer også berre Porsanger-former. Det er dermed mogleg å slå opp på Varanger-forma *pimeä* og få vite at ordet betyr *mørk*, og at lemmaet (Porsanger-forma) er *pim-mee*. På denne måten blir ei ordbok som opprinneleg er skriven for éin dialekt, i stand til å gje omsetjingar av ord frå tre ulike dialektar.

4.4. Dekningsgrad

Det er stor variasjon i dekningsgraden til dei ulike ordbøkene. Tabell 7 viser dekningsgraden for dei 1000 og 10 000 mest vanlege lemmaa for kvart språk (skoltesamisk manglar fordi eit relevant korpus ikkje var tilgjengeleg). Frekvensdata for dei samiske språka og kvensk er henta frå SIKOR. Sjangrane tekstane er henta frå, er primært administrative tekstar. For nordsamisk er det også avis-språk og for enaresamisk også tidsskriftsartiklar av blanda innhald. For norsk brukte eg NOWAC og for finsk frekvenslister frå Kielipankki (sjå litteraturlista for referansar).

Dekningsgrad for topp-1000	Lmaj = norsk			Lmaj = finsk	
	nord-samisk	sør-samisk	kvensk	nord-samisk	enaresamisk
Ordbokoppslag på Lmin	84,0 %	85,5 %	79,9 %	78,5 %	87,6 %
Ordbokoppslag på Lmaj	87,8 %	74,3 %	66,0 %	88,3 %	86,8 %
Dekningsgrad for topp-10 000	nord-samisk	sør-samisk	kvensk	nord-samisk	enaresamisk
Ordbokoppslag på Lmin	68,0 %	40,5 %	28,8 %	40,8 %	61,6 %
Ordbokoppslag på Lmaj	57,5 %	32,9 %	32,2 %	46,7 %	48,9 %

Tabell 7: Dekningsgrad for dei 1000 og 10 000 vanlegaste orda i 10 ordbøker.

Den beste dekningsgraden har finsk-nordsamisk, noko som kjem av at den same frekvenslista for finsk som ligg attom denne tabellen, vart brukt også i utarbeidinga av sjølve ordboka. Elles illustrerer det brå fallet i dekningsgrad frå topp-1000 til topp-10 000 at alle ordbøkene (bortsett frå nordsamisk-norsk) er relativt små (jf. tabell 3).

4.5. Grunnform versus bøyingsform som søkeord

I og med at alle ordbøkene er knytt til ein morfologisk analysator, er det mogleg å slå opp også på alle ordformene i bøyingsparadigmet til kva som helst lemma i ordboka. I loggen er søk på oblike former kjenneteikna ved at søkeord (kolonne 1 i tabell 3) og lemma (kolonne 3 i tabell 3) er ulike. Tabell 8 viser at dei fleste ordbokoppslaga i ordbøkene frå minoritetsspråk til majoritetsspråk blir gjort på lemma. Språket med flest oppslag på oblike former er nordsamisk, og dette er også språket med mest tekst tilgjengeleg elektronisk.

Ordbok Lemma = Lmin	Lmaj = norsk			Lmaj = finsk		
	nord-samisk	sør-samisk	kvensk	nord-samisk	enare-samisk	skolte-samisk
Oppslag ≠ lemma	42,0 %	35,8 %	26,5 %	37,5 %	31,4 %	22,5 %

Tabell 8: Ordbokoppslag på oblike former i ordbøker med lemma på minoritetsspråket, i prosent av alle oppslaga.

Som vist i Antonsen et al. (2009:278) er delen av ord i grunnform i løpande nordsamisk tekst 7,9 %. Viss den dominerande bruken hadde vore lesing av e-tekst med klikk-i-tekst-funksjonen, hadde prosenten av oblike former med andre ord vore langt høgare enn dei rundt 40 % vi har. Ein rimeleg konklusjon er at ordbøkene ikkje primært blir brukt til resepsjon av minoritetsspråkstekst. I og med at dei fleste oppslaga går frå minoritets- til majoritetsspråket

(jf. tabell 4), er det heller ikkje mogleg å seie at dei primært blir brukt til produksjon. Loggen viser ikkje i kor mange av oppslaga brukarane går vidare til å generere bøyingsparadigme for ordet eller å slå opp i korpus, men ei mogleg forklaring på dataa i desse to tabellane kan vere at det er L1-brukarar som slår opp i ordboka for å få opplysningar om rettskriving og bøyning.

4.6. Innhaldet i loggane

Ein annan innfallsvinkel til loggdata er å sjå nærare på orda brukarane slår opp. Tabell 9 viser dei 15 vanlegaste bokmålsorda som brukarane slår opp i ordbøkene til kvensk, sørsamisk og nordsamisk.

Rang	Norsk-nordsamisk	Norsk-sørsamisk	Norsk-kvensk
1	skulle	måtte	vel
2	ha	få	jeg
3	kunne	forstå	hus
4	gå	ha	være
5	se	samisk språk	hei
6	bli	gjøre	gå
7	komme	se	kaste
8	få	snø	bli
9	dra	møte	kunne
10	være	komme	du
11	fare, fore, for	språk	se
12	det	nå	eie
13	snakke	løpe	følge etter
14	som	høre	ha
15	på	være ordstyrer	skulle

Tabell 9: Dei 15 vanlegaste norske oppslagsorda.

For den nordsamiske ordboka er dei 10 vanlegaste oppslagsorda hjelpeverb eller andre viktige verb, og dei andre orda er sentrale funksjonsord. Legg særleg merke til det nest vanlegaste ordet, verbet *ha*, som ikkje har noko direkte motsvar på nordsamisk. Brukaren treng med andre ord hjelp til å produsere habitive setningar, og slår opp på *ha*. Samla sett består lista av verb som opptre i ein stor del av setningane i språket, ord med ulike tydingar i ulike samanhengar. Inntrykket loggen gjev, er at han er dominert av brukarar som har norsk og ikkje nordsamisk som morsmål, og at dei bruker ordboka for å produsere nordsamisk tekst.

For sørsamisk ser vi same tendens, sjølv om den ikkje er like klår. Også her dominerer hjelpeverb og andre sentrale verb. Verbet *ha* er ikkje like høgt oppe, men sørsamisk skil seg frå nordsamisk ved å ha eit verb som svarer til det norske *ha*. Eit par substantiv har også snike seg med. Oppslagsorda *språk* og *samisk språk* bør nok bli sett i samheng med verbet *forstå* (som i *forstå samisk*), og den høge plasseringa av *snø* kan tyde på at ein stor del av brukarane er meir interesserte i eksotiske drag ved språket enn i å verkeleg produsere sørsamisk tekst. Det 15. oppslaget er *støy*. I det heile er inntrykket også her at ordboka er ei produksjonsordbok, men delen av brukarar som produserer normal prosa, ser likevel ut til å vere lågare enn for nordsamisk.

For kvensk er det vanskelegare å dra nokon konklusjonar. Hovudintrykket er mangelen på sentrale verb. Verba *ha*, *skulle* og *vere* er med, men ikkje like høgt oppe som for nordsamisk. Derimot er *eie* med, noko som sannsynlegvis er uttrykk for at brukarar er på utkik etter ein alternativ strategi for å uttrykke eigarkonstruksjonen. Kvensk har eit verb *omistaat* 'å eige', men den vanlege måten å uttrykke at nokon har noko, er, som på nordsamisk, med på-kasus og verbet for å vere («Eg har ein båt» blir uttrykt som *Minula oon venet* 'På meg er båt'). I det heile er det vanskelegare å dra konklusjonar basert på materialet for norsk-kvensk. At det empiriske materialet berre er ein brøkdel av det det er for nord-

samisk (42 000 mot 655 000, jf. tabell 4), gjer også at rangeringa av oppslagsord på frekvenslista blir meir tilfeldig. Det er likevel klårt at den kvenske ordboka i mindre grad enn dei to samiske blir brukt til produksjon av L2-tekst.

Mønsteret for finsk-nordsamisk (tabell 10) minner om mønsteret for dei to ordbøkene frå norsk til høvesvis nord- og sør-samisk. Bortsett frå *voida* manglar dei modale hjelpeverba, men ut over det er listene dominert av dei mest sentrale finske verba. For finsk-enaresamisk er det fleire verb enn for finsk-nordsamisk, noko som kan spegle det aktive revitaliseringsarbeidet for enaresamisk. På begge listene er det også substantiv, men det er sentrale substantiv til bruk i språkproduksjonen. Også her ber lemmautvalet preg av at brukarane har dei samiske språka som L2.

Rang	Finsk-nordsamisk	Finsk-enaresamisk
1	olla 'vere'	olla
2	saada 'få'	saada
3	fiinnis 'fin'	tulla
4	se 'den, det'	voida
5	tulla 'kome'	pitää
6	voida 'kunne'	lapsi 'barn'
7	mennä 'gå, dra'	tehdä
8	tehdä 'gjere'	ruoka 'mat'
9	aika 'tid'	mennä
10	työ 'arbeid'	haluta
11	katsoa 'sjå'	tarvita 'trengje'
12	haluta 'ville'	käydä 'vitje'
13	pitää 'halde, like'	kirjoittaa 'skrive'
14	tämä 'denne, dette'	syödä 'ete'
15	ihminen 'menneske'	se

Tabell 10: Dei 15 vanlegaste finske oppslagsorda for to ordbøker.

5. Konklusjon

Ein analyse av loggane for tolv ulike ordbøker mellom fem ulike minoritetsspråk og dei respektive majoritetsspråka deira viser at ordbøkene er mykje brukt. Dei fleste oppslaga går frå minoritetsspråket til majoritetsspråket, men skilnaden er ikkje overveldande (40 % vs. 60 %). Dei ulike språksamfunna bruker til ein viss grad ordbøker på ulike måtar. Ordbokbrukarane for dei samiske ordbøkene slår opp sentrale verb, hjelpe- og modalverb. For kvensk er mønsteret meir tilfeldig og dei sentrale verba ikkje like dominerande.

Oppslaga på minoritetsspråkslemma er dominert av søk på grunnform; over 60 % av søka er av denne typen. Dette står i kontrast til løpande tekst, der grunnformene utgjør mindre enn 10 % av ordformene. Det ser med andre ord ikkje ut til at bruk av ordbøkene som resepsjonsordbøker er den dominerande bruken.

Graden av feilskrivning av oppslagsord er ulik for minoritets- og majoritetsspråksoppslag. Blant mislukka søk på minoritetsspråka er prosenten av feilskrivne ord så høg som mellom 60 og 90. For majoritetsspråka er prosenten også høg, mellom 30 og 40, men altså langt lågare. Dette resultatet viser ikkje berre at oppslagsord svært ofte blir skrivne feil i e-ordbøker, men også at dette er eit stort problem for minoritetsspråk. Måtar å hjelpe ordbokbrukarane på slik at dei finn oppslagsorda dei er på jakt etter, bør med andre ord få høg prioritet ved vidare arbeid med ordbokgrensesnitt.

Litteratur

Ordbøker

Álgu = Etymologische Datenbank für die saamischen Sprachen.

Forskningscentralen för de inhemska språken i Finland.

<<http://kaino.kotus.fi/algu/>> (mars 2019).

- Aronsen, Terje (2010): *Kvensk-norsk-kvensk elektronisk ordbok*. Redigert av Verena Schall & Trond Trosterud. Universitetet i Tromsø.
- Bokmålsordboka*. Oslo: Kunnskapsforlaget. <ordbok.uib.no> (mars 2019).
- DinOrdbok. <<http://dinordbok.no>> (mars 2019).
- Glosbe = Glosbe – the multilingual online dictionary. <<http://glosbe.com>> (mars 2019).
- Jernsletten, Nils (1991): *Álgosátnegirji sámii-dáru sátnegirji = samisk-norsk ordbok*. Kárášjohka: Davvi girji.
- Jåma, Albert (2001): *Norsk-sydsamisk ordliste. Gærjiste vaalteme*. Hemnes sameforening. <http://www.ruovatsijte.no/gaerjiste-vaalteme_2001.pdf> (mars 2019).
- Neahttadigisánit. E-ordbøker frå Giellatekno ved Universitetet i Tromsø – Noregs arktiske universitet: nordsamisk: <<http://sanit.oahpa.no>> (mars 2019), sørsamisk: <<http://baakoah.oahpa.no>> (mars 2019), enaresamisk: <<http://saanih.oahpa.no>> (mars 2019), skoltesamisk: <<http://saan.oahpa.no>> (mars 2019) og kvensk: <<http://sanat.oahpa.no>> (mars 2019). Felles ordbokportal: <<http://dicts.uit.no/>> (mars 2019). Kjeldekode for grensesnittet: <<https://gtsvn.uit.no/langtech/trunk/apps/dicts/nds/>> (mars 2019).
- NNNDO = *Nordsamisk-norsk-nordsamisk digital ordbok på nett*. Kárášjohka: Davvi girji. <<https://533.davvi.no/>> (mars 2019).
- Olthuis, Marja-Liisa & Taarna Valtonen (2019): *Suomi-inarinsaame sanakirja*. Saamelaiskäräjät.
- Sammallahti, Pekka & Jouni Mosnikoff (1991): *Suomi-koltansaame sanakirja*. Ohcejohka: Girjegiisa.
- SNO = John Henrik Eira, Johan Jernsletten, Brita Kåven, Ingrid Nordal & Aage Solbakk (1995): *Sámii-dáru sátnegirji = Samisk-norsk ordbok*. Kárášjohka: Davvi girji.
- SNSO = Brita E. Kåven (red.) (2000): *Stor norsk-samisk ordbok / Dáru-sámii sátnegirji*. Kárášjohka: Davvi girji.

- StarDict. <<http://www.huzheng.org/stardict/>> (mars 2019).
- Söderholm, Eira (2012): *Sannoi ja haamui. Porsangin ja Raisin variantti. Kainun sana- ja sanahaamulista Aikamatkaa varten. Kvensk ord- og ordformliste til Aikamatka*. Manuskript. Universitetet i Tromsø – Noregs arktiske universitet. <<https://tinyurl.com/y6xmfh45>> (mars 2019).
- Wiktionary. <<http://wiktionary.org>> (mars 2019).

Annan litteratur

- Almind, Richard (2008): Søgemønstre i logfiler. I: *LexicoNordica* 15, 33–55.
- Antonsen, Lene, Ciprian-Virgil Gerstenberger, Sjur N. Moshagen & Trond Trosterud (2009): Ei intelligent elektronisk ordbok for samisk. I: *LexicoNordica* 16, 271–283.
- de Schryver, Gilles-Maurice & David Joffe (2004): On How Electronic Dictionaries are Really Used. I: G. Williams & S. Vessier (eds.): *Proceedings of the Eleventh EURALEX International Congress*. EURALEX 2004, Lorient, France, July 6–10, 2004. Lorient: Faculté des Lettres et des Sciences Humaines, Université de Bretagne Sud, 187–196.
- EGIDS = Language Status. I: David M. Eberhard, Gary F. Simons & Charles D. Fennig (eds.) (2019): *Ethnologue: Languages of the World. Twenty-second edition*. Dallas, Texas: SIL International. <<https://www.ethnologue.com/about/language-status>> (mars 2019).
- Ethnologue = David M. Eberhard, Gary F. Simons & Charles D. Fennig (eds.) (2019): *Ethnologue: Languages of the World. Twenty-second edition*. Dallas, Texas: SIL International. <<http://www.ethnologue.com>> (mars 2019).
- Haavisto, Mervi, Kaisa Maliniemi, Leena Niiranen, Pirja Paavaniemi, Tove Reibo & Trond Trosterud (2014): Kvensk ordbok på nett – hvem har nytte av den? I: Ruth Vatvedt Fjeld & Marit

- Hovdenak (red.): *Nordiske studier i leksikografi* 12. Oslo: Nordisk forening for leksikografi, 176–192. <<https://tidsskrift.dk/nsil/article/view/20997>> (mars 2019).
- Johnson, Ryan, Lene Antonsen & Trond Trosterud (2013): Using finite state transducers for making efficient reading comprehension dictionaries. I: *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013), May 22–24, 2013, Oslo University, Norway*. NEALT Proceedings Series 16, 59–71. <https://munin.uit.no/bitstream/handle/10037/12884/paper_5.pdf> (mars 2019).
- Kielipankki = Kielipankin konkordanssihakuohjelma. <<http://korp.csc.fi>> (mars 2019).
- Lewis, M. Paul & Gary F. Simons (2010): Assessing endangerment: Expanding Fishman's GIDS. I: *Revue Roumaine de Linguistique* 55(2), 103–120. <<https://www.lingv.ro/RRL%202%202010%20arto1Lewis.pdf>> (mars 2019).
- NOWAC = *Norwegian web as a corpus*. <<https://www.hf.uio.no/iln/tjenester/kunnskap/sprak/korpus/skriftsprakskorpus/nowac/index.html>> (mars 2019).
- Nygaard, Lars & Ruth Vatvedt Fjeld (2008): Analyse av søkelogger for bedre søkemuligheter i elektroniske ordbøker. I: *LexicoNordica* 15, 57–72.
- OPUS. <<http://opus.nlpl.eu>> (mars 2019).
- SIKOR. Universitetet i Tromsø – Noregs arktiske universitets og det norske Sametingets samiske tekstsamling, versjon 06.11.2018. <<http://gtweb.uit.no/korp/>> (samiske språk) og <http://gtweb.uit.no/f_korp> (austersjøfinske språk) (begge mars 2019).
- SIL International (2007): ISO 639-3, *Codes for the representation of names of languages – Part 3: Alpha-3 code for comprehensive coverage of languages*. <<https://iso639-3.sil.org/>> (september 2019).
- Trosterud, Trond (2001): *Stor norsk-samisk ordbok / Dáru-sámi sátnegirji* [bokmelding]. I: *LexicoNordica* 8, 283–306.

Wiegand, H.E. (1998): *Wörterbuchforschung: Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilband. Berlin: Walter de Gruyter.

Trond Trosterud
professor, dr.art.
Institutt for språk og kultur
UiT – Noregs arktiske universitet
Postboks 6050 Langnes
NO-9037 Tromsø
trond.trosterud@uit.no