

De sirkumpolare Wikipedia-utgåvene

Trond Trosterud, UiT Noregs arktiske universitet

Samandrag

Det fleirspråklege oppslagsverket Wikipedia ser ut til å ha ei spesielt høgborg i den sirkumpolare regionen. Av dei rundt 70 sirkumpolare språka, dvs. innfødde språk i denne regionen, har nesten alle eit skriftspråk, og 45 av dei har sin eigen Wikipedia eller er i Wikipedia-inkubatoren. Samanlikna med den globale situasjonen, er både posisjon til Wikipedia og grada av lesekunne i nordområda oppsiktsvekkande: I følge ethnologue.com finst det 7117 levande språk, og meir enn halvparten (3982) har eit skriftspråk og berre 620 har ein Wikipedia eller er i inkubatoren. Viss vi ser nærare på desse sirkumpolare Wikipedia-utgåvene, vil det visa seg at desse tala er villeiande. Denne artikkelen undersøker på kva måte dei er villeiande for at vi betre skal kunna forstå skriftspråkskulturen til urfolk, situasjonen for dei sirkumpolare språka dei ulike landa dei er i, og korleis små Wikipedia-utgåver blir oppretta.

Introduksjon

Denne artikkelen vil analysere innhaldet i dei 45 inkubatorutgåvene frå Wikipedia eller Wikipedia-inkubatoren skrive på eit sirkumpolart minoritetsspråk. Med sirkumpolart språk viser vi til eit minoritetsspråk som talast av ei innfødd befolkning i eit område nord for eller nær 60. nordlege breiddegrad, unntatt statsspråka.

Artikkelen er strukturert som følger: Del 2 presenterer dei sirkumpolare språkfelleskapa lese- og skriveferdigheitene og Wikipedia-utgåvene deira. Del 3 analyserer forskjellane mellom dei forskjellige utgåvene i eit sirkumpolart språk, årsakene til dette, og kva rolle desse utgåvene spelar i den språklege revitaliseringa. Del 4 oppsummerer det vi no veit om desse utgåvene og forholda som har ført til at nokon av dei har lykkast, og konkluderer med nokre få oppmodingar om kva vi kan gjere betre i framtida..

Bakgrunn

Dei sirkumpolare språka

Viss vi deler det sirkumpolare området inn i tre regionar ser vi at den nordamerikanske regionen har 30 sirkumpolare språk, den russiske regionen 31 og den nordiske regionen 10. Dei dominerande språkfamiliane er *uralsk* (23 sirkumpolare språk i Norden og Russland), *dene* (19 sirkumpolare språk i Nord-Amerika) og *eskaleut* (10 språk i alle 3 regionane). Dei resterande 20 språka er enten isolat eller tyrkiske eller tungusiske språk, med unntak av det germanske språket færøysk.

Dei fleste sirkumpolare språka er små; mediantalet for dei som snakkar språka som er diskuterte her, er 650. I avtakande rekkefølge er dei 7 største språka (50 000 til 500 000 språkbrukarar) i utvalet sakha, tuvinsk, burjatisk, komi, færøysk, komipermjakisk og grønlandsk. Alle desse har si eiga administrative eining der språket har offisiell status, enten ein autonom republikk eller eit område i Russland eller sine egne øyar med sjølvstyre, som for grønlandsk og færøysk.

Tre av språka er i fleirtal innanfor sitt autonome område¹: færøysk 94 %, tuvinsk 91 %, grønlandsk 85 % og inuktitut 63 %. Dei andre språka er i mindretal (sakha 38 %, burjatisk 27 % og komi 18 %). Resten av språka har enten ein svakare status eller ingen offisiell status.

Typologisk har alle sirkumpolare språk ein kompleks morfologi. Både na-dene- og eskaleutspråka er polysyntetiske, og dei bruker verbal bøying for å uttrykke den grammatiske og pragmatiske informasjonen uttrykt med pronomen og adverb i dei germanske språka. Na-dene-språka er hovudsakleg prefikserande språk, medan eskaleutspråka er suffikserande. Dei sirkumpolare språka som blir snakka i Russland og dei nordiske landa er for det meste suffikserande. For dei fleste av desse språka er kasussystema store, vanlegvis med mellom 6 og 10 kasus, men komispråka har mellom 20 og 30. Minoritetsspråka i Norden har færre kasus (mindre enn 10), men dei har meir komplekse system for morfofonologisk stammeveksling under bøying. Alt i alt skil dei sirkumpolare minoritetsspråka seg markant frå dei respektive majoritetsspråka sine. Dei største språklege avstandane til majoritetsspråket finst for dei sirkumpolare språka i Nord-Amerika, for grønlandsk og for samiske språk i Noreg og Sverige. Spesielt dei som snakkar russisk og finsk vil ha ein noko kortare språkleg avstand til minoritetsspråka i dei respektive landa sine

Lese- og skrivedugleik blant dei sirkumpolare språka

Lese- og skrivedugleik i nord kan sporast tilbake til tre forskjellige initiativ for alfabetisering. Det første var misjonsarbeid i samband med statskonsolidering frå det 17. til og med det 19. hundreåret. Dette førte til at det latinske alfabetet vart brukt til å lage ortografiane i Norden, kyrillisk i Russland og Alaska, og både syllabisk og latin i Nord-Amerika. Det andre initiativet fann stad berre i Russland, der oktoberrevolusjonen lanserte ein planlagt, stortilt skiping av skriftspråk basert på det latinske alfabetet for alle minoritetsspråka, inkludert dei nordlege med ideen om til slutt å erstatte dei tidlegare kyrilliskbaserte ortografiane. Til slutt, i ein prosess som starta på 1960- og 1970-talet, har grasrotinitiativ og ei følgjande endring, eller i det minste ei oppmjuking av assimileringspolitikken, ført til at nye ortografiar vart oppretta for nokon av dei sirkumpolare språka og for andre vart dei eldre ortografiane revidert.

Ser vi nærare på dei forskjellige delane av dei sirkumpolare områda, ser vi at i Nord-Amerika vart tidlege misjonsinitiativ som bruker morsmålet frå det 18. hundreåret og framover erstatta med ein assimilerande politikk som auka med integreringa av dei nordlege folkegruppene i det moderne samfunnet i løpet av det 20. hundreåret. Som ein del av forsoningsprosessen skilde Canada i 1999 ut Nunavut frå Nordvestterritoriet som eigen region. I 2008 gav Canada inuktitut status som offisielt språk og byrja å publisere forhandlingane i Nunavuts lovgivande forsamling på både engelsk og inuktitut. Dei fleste sirkumpolare språka i Canada blir undervist i First Nation-reservata (reservata til urfolka), men mest som framandspråk med engelsk som undervisningsspråk, og med dårlege resultat (sjå Meek 2010 med ein grundig analyse av denne praksisen). I Alaska endra assimileringspolitikken endå mindre enn han gjorde i Canada. På 1960-talet byrja University of Fairbanks å lage praktiske ortografiar for fleire indianarspråk i Alaska, men dette førte ikkje til systematiske endringar i læreplanane i skulane. Fråsegna om 20 innfødde språk i Alaska som «med-offisielle språk med engelsk» i 2014 førte heller ikkje til noko radikalt skifte i politikken¹.

¹ Jfr. <http://www.akleg.gov/basis/Bill/Text/28?Hsid=HB0216Z>, den nye paragrafen med tittelen «Adding the Inupiaq, Siberian Yupik, Central Alaskan Yup'ik, Alutiiq, Unangax, Dena'ina, Deg Xinag, Holikachuk, Koyukon, Upper Kuskokwim, Gwich'in, Tanana, Upper Tanana, Tanacross, Hän, Ahtna, Eyak, Tlingit, Haida og Tsimshian languages as official languages of the state.» Lova seier framleis eksplisitt at dette ikkje krev at statlege eller kommunale styresmakter skriv ut dokument på noko anna språk enn engelsk.

I Russland vart skipinga av nye ortografiar på 1930-talet kombinert med administrative endringar der etnisk blinde administrative grenser vart erstatta med eit system med autonome republikkar danna i samsvar med språkgrensene, og for dei mindre minoritetane vart område oppretta rundt alle minoritetar store nok til å utgjere ei administrativ eining. I løpet av dei første tiåra etter andre verdskrigen vart språk med sin eigen autonome republikk (Komi, Sakha, Tuva, Burjatia, Khakassia) brukt som undervisningsspråk i grunnskular. Republikane gav ut tidsskrift og aviser på sitt eige språk, og romanar vart òg skrivne og publisert på språket til republikken. Dei som snakka dei andre sirkumpolare språka i Russland vart i det minste lært opp til å lese og skrive på dei respektive morsmåla sine, og enkelte publikasjonar brukte òg desse språka. Med dei nye skulelovene som vart introduserte i 1958–59, vart urfolka sine lese- og skriveferdigheiter svekt, men dei heldt seg framleis på eit nivå som var langt sterkare enn for minoritetsspråka i dei andre sirkumpolare regionane².

Etter å ha vore styrt på dansk i fleire hundreår, byrja færingane å arbeide med sitt eige skriftspråk på slutten av det 19. hundreåret. Den britiske okkupasjonen under andre verdskrigen svekte dansk innverknad, og etter innføringa av lokalstyret i 1948 vart færøysk gradvis òg det dominerande skriftspråket (jfr. Knudsen 2010). På Grønland vart grønlandsk brukt i heile skolesystemet, inkludert lærarutdanninga, og til og med i den dominerande avisa på Grønland, men i 1953 vart denne 100 år gamle tradisjonen med innfødt lese- og skrivedugleik knust då Grønland gjekk frå å vere ein dansk koloni til eit fylke (amt) integrert i Danmark. Overgangen vart kombinert med eit stort skifte i lese- og skrivedugleik, frå grønlandsk til dansk. Sjølv om innføringa av lokalstyret i 1979 og av sjølvstyret i 2008 gjorde grønlandsk til det einaste offisielle språket på Grønland, har dansk framleis ein sterk posisjon, spesielt som det dominerande språket i all teoretisk utdanning over obligatorisk nivå. Både administrasjonen, rettssystemet og massemedia er tospråklege, med tekst som vanlegvis er skrive på dansk og omsett til grønlandsk (Frederiksen og Olsen, op.cit.).

I løpet av dei siste fire tiåra har dei nordiske land langsamt implementert ein ny språkpolitikk for minoritetsspråka sine. Dette har ført til offisiell status på regionalt nivå eller status som eit nasjonalt minoritetsspråk for samiske språk, kvensk og meänkieli. Store deler av læreplanen i nordsamiske skolar har nordsamisk og i mindre grad skolte- og enaresamisk som undervisningsspråk. Alle dei sirkumpolare minoritetsspråka i Norden er også skolefag i grunnskolen³. Eitt språk som skil seg frå dei andre er enaresamisk, der eit pågåande språkskifte i løpet av den siste generasjonen har vorte omgjort til ein situasjon der fleirtalet av språkbrukarane er frå den yngre generasjonen, og språkmiljøet publiserer både tidsskrift, lærebøker, skjønnlitteratur, og nyhendendingar på enaresamisk⁴.

Så for dei sirkumpolare språka som heilskap inneber dette at berre ein firedel av dagens ortografiar er frå tida før andre verdskrigen. Desse ortografiane er enten kyrilliskbaserte ortografiar frå 1937, eller dei er skrive med syllabiske alfabet (stavingsalfabet). Halvparten av dei sirkumpolare ortografiane er under 40 år. Dei fleste sirkumpolare språka har størstedelen av språkbrukarane i dei øvre aldersgruppene, den prototypiske situasjonen er dermed at sirkumpolare språk har ortografiar som er yngre enn sine eigne språkbrukarar.

2 Ein presentasjon av den sovjetiske politikken skrive av ein av arkitektene attom politikken er Isaev 1977. For analysar sjå f.eks. Kreindler 1985 og Trosterud 1995. Bilinsky 1962 gir ein analyse av effekten av skulelovene frå 1958–59.

3 Ein illustrasjon av politikken i Noreg er *Handlingsplan for samiske språk* (Arbeids- og inkluderingsdepartementet 2009). Eit oversys over situasjonen for samiske språk i Finland er Pasanen 2016.

4 Pasanen 2015 er standardreferansen for revitaliseringsprosessen. Ein grundig analyse av gjenopprettinga av «mellomgenerasjonen» (den mellom dei eldre språkbrukarane og språka barna bruker) er Olthuis m.fl. 2013.

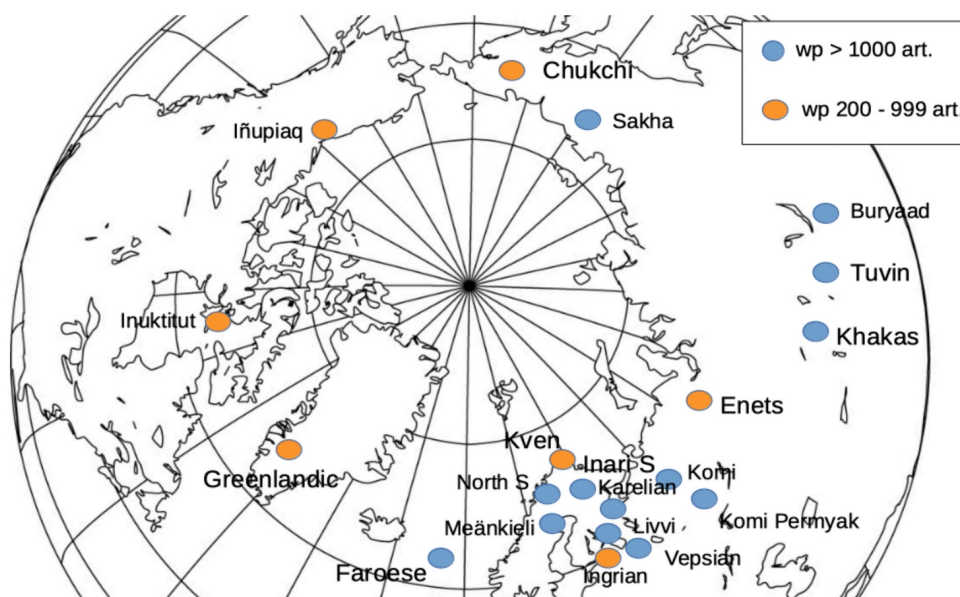
Når det gjeld språkbruk i sosiale media, er både færøysk, grønlandsk, meänkieli, nordsamisk og enaresamisk og til ei viss grad òg skoltesamisk, sørsamisk, lulesamisk, karelsk, livvisk og vepsisk i bruk på Facebook, både i grupper vigde til dei respektive språka, og i personlege innlegg. I Nord-Amerika ser det ut til å vere annleis. Her er alle Facebookgruppene vigde til dei største språka (inuktitut, sentralalaskajupik, dene suline, iñupiaq) berre brukte engelsk. Ord på urfolksspråka vart vanlegvis sett i hermeteikn (som i «kva betyr X»), men diskusjonen var alltid på engelsk. Burjatisk, komi, tuvinskog khakassisk er i bruk i *vkontakte*-grupper⁵ vigde til bruk av det aktuelle språket, Sakha i mindre grad. Dei andre sirkumpolare språka i Russland med si eiga Wikipedia-utgåve ser ut til enten å ha *vkontaktespråk*gruppene sine på russisk, eller å ikkje ha ei gruppe i det heile.

Dei sirkumpolare Wikipedia-utgåvene

Vi går nå over til de konkrete Wikipedia-utgåvene. Etter en kort oversikt ser vi på størrelsen på utgåvene og språkfellessamfunna, deretter på forfattarane, og til slutt på tilstanden på artiklane.

Wikipedia-utgåvene

Kartet i Figur 1 viser den omtrentlege plasseringa av Wikipedia-utgåvene med meir enn 200 artiklar med oransje ovalar og dei med meir enn 1000 artiklar med blå ovalar. Som det kjem fram av Figur 1, finst Wikipedia-utgåvene som inneheld meir enn 1000 ord, hovudsakleg i dei nordiske landa, men også i Russland med republikkspråka. I tillegg til språka som blir viste på kartet, er det 24 Wikipedia-utgåver som inneheld mindre enn 200 ord. Frå størst til minst er 6 av desse i Nord-Amerika (sentralalaskayupik, tlingit, hän, wich'in, aleutisk og dene suline), 15 i Russland (kildinsamisk, mansisk, ketisk, tundranenetsisk, lydisk, tundrajukaghirsk, nanaisk, evenkisk, sibirsk yupik, khantisk, dolgansk, udege, ulch, oroch og negidalsk), og i dei nordiske landa (sørsamisk, skoltesamisk og lulesamisk). Karakteristisk for språka i denne gruppa er at dei har få språkbrukarar og svake eller ingen litterære tradisjonar. Wikipedia-utgåvene blir vanlegvis starta av ikkje-språkbrukarar utanfor språkfellesskapet, nokre av artiklane er på majoritetsspråket, og det er for tida ingen aktivitet på nokon av desse Wikipedia-utgåvene.



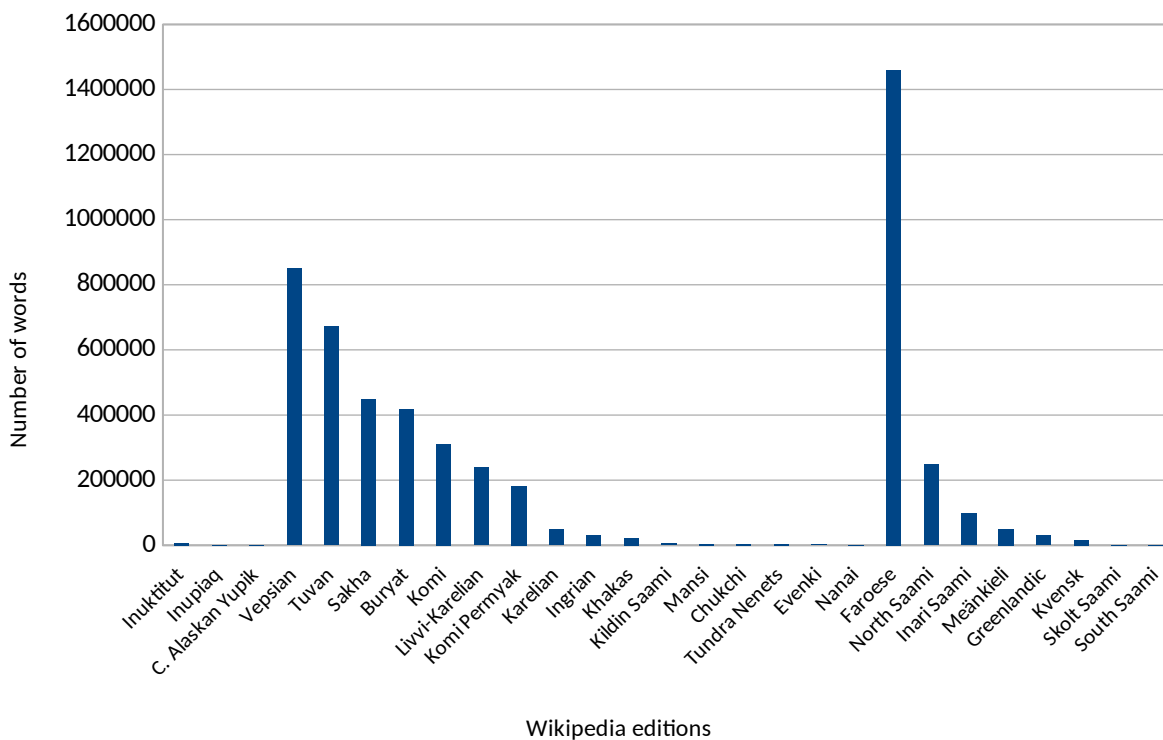
Figur 1:

Sirkumpolare Wikipedia-utgåver med meir enn 200 artiklar

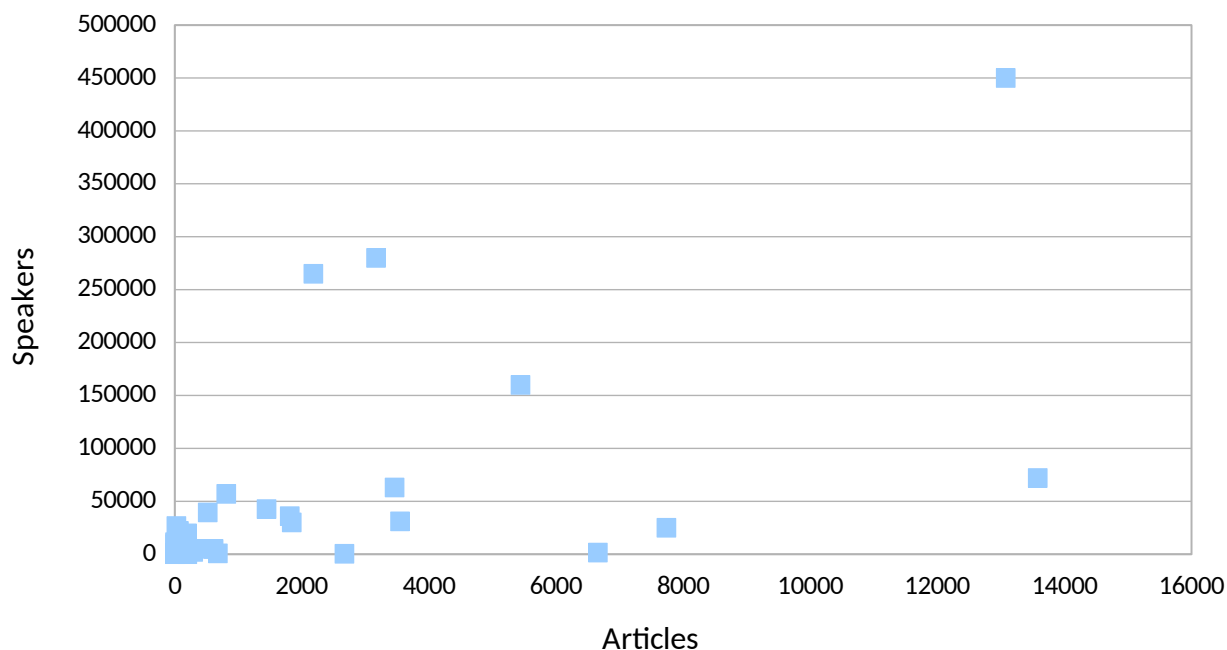
⁵ *vkontakte* er den russiske parallellen til Facebook, jf. <https://vk.com/>

Wikipediastorleik

La oss først gjere opp status for storleiken på Wikipedia-utgåvene som inneheld meir enn 1000 ord. Dei er viste i Figur 2, ordna først etter område (Nord-Amerika, Russland, Norden), og deretter etter kor mange ord dei inneheld. Som det kjem fram har dei tre områda forskjellige profilar. For Nord-Amerika (til venstre) har dei tre største Wikipedia-utgåvene knapt 100 ord. Wikipedia-utgåvene frå Russland kan (med nokre få unntak, som vi kjem attende til) blir delte inn i to grupper: Dei som har støtte frå sin eigen republikk, og dei utan. I dei nordiske landa er færøysk avvikande. Den nordsamiske utgåva er den største etter den færøyske, med enaresamisk etterpå. Resten av utgåvene er langt mindre.



Figur 2: Wikipedia-utgåver med over 1000 ord i Nord-Amerika, Russland og dei nordiske landa



Wikipedia-storleik samanlikna med storleiken på språksamfunnet

I utgangspunktet skulle ein tru at fleire språkbrukarar skriv fleire artiklar. Som vist i Figur 3, er det verkeleg ein slik samanheng, men han er svak (korrelasjonen, eller rettare kvadratet av korrelasjonen, er $R^2=0,37$, der ein R^2 -verdi på 1.0 ville ha vore ein perfekt samanheng mellom språkbrukarar og Wikipediastorleik). Dei avvikande utgåvene med «for få artiklar» er sakha, komi, tuvinsk og burjatisk, medan dei med «for mange artiklar» er færøysk, nordsamisk, vepsisk og enaresamisk. Dei avvikande utgåvene med for få artiklar er òg dei fire språka i Russland med flest talarar og størst språkmiljø. Samanlikna med større Wikipedia-utgåver utanfor det sirkumpolare området er dei framleis ikkje så langt unna eit samsvar mellom talarar og artiklar. Med 88 språkbrukarar per artikkel er burjatisk på nivå med albansk (72), og høvesvis 29 og 34 språkbrukarar per artikkel for komi og sakha er kompatibelt med tilsvarende data for dansk (21), ungarsk (30) og bulgarsk (33). Færøysk er på nivå med islandsk (begge har omtrent 5,5 språkbrukarar/artikkel).

Av dei sirkumpolare språka med meir enn 10 000 språkbrukarar skil 5 seg ut med «for mange artiklar», alle i dei nordiske landa eller i Karelen: Færøysk, nordsamisk, vepsisk, livvisk (olonetsisk) og enaresamisk. For dei minste språkmiljøa, vepsisk og enaresamisk er det enkelt å ha eit lågt tal på språkbrukarar per artikkel. Likevel skil desse språkmiljøa seg ut med høgare aktivitet enn resten, spesielt i forhold til storleiken.

Korrelasjonen mellom storleik på språksamfunnet og Wikipediastorleik er dermed svak, og det må vere andre faktorar enn kor stort språket er som bestemmer storleiken på Wikipedia-utgåvene.

Forfattarane

Når språkmiljøa er så små som dei er for sirkumpolare språk, spelar åtferda til individuelle forfattarar ei større rolle enn den gjer for større språkmiljø. For kvar Wikipedia-utgåve er det ei side under *Spesialsider > Liste over aktive brukarar* som viser kven som har vore aktive dei siste 30 dagane. Wikipedia-omgrepet *brukarar* refererer til personar eller dataprogram (bots) som redigerer

Wikipedia-utgåvene. Nokon av dei oppførte brukarane kjenner språket til den aktuelle Wikipedia-utgåva, men ein stor del, og i mange tilfelle dei fleste brukarane, gjer ikkje det. Desse skribentane utan språkkunnskap («ikkje-språkbrukarar») kan delast inn i tre kategoriar: brukarar som ikkje rører det språklege innhaldet, brukarar med tilstrekkeleg kunnskap om både språket og sine eigne avgrensingar som forfattar til å produsere akseptabel tekst, og brukarar som har liten eller ingen kunnskap om språk, men framleis produserer tekst, som vanlegvis varierer mellom veldig dårleg kvalitet og fullstendig sludder.

For å klassifisere brukarane i språkbrukarar og ikkje-språkbrukarar, har eg sett på både kva dei sjølv seier om språkkunnskapen sin og på teksten dei produserer. Mange brukarar erklærer språkkunnskap på brukarsidene sine, etter ein 4-punktskala (0-3 og morsmålskunnskap). Eg gjekk ut i frå at dei som erklærte at dei ikkje kunne språket («0») snakka sant, og så nærare på dei som enten ikkje oppgav språkkunnskap eller erklærte at dei kunne språket på eitt eller anna nivå. For desse brukarane gjennomgjekk eg bidraga deira og klassifiserte dei som berre hadde redigeringar som ikkje krev språkkunnskap (t.d. dei som berre la til fotografi) som ikkje-språkbrukarar⁶.

Ser vi først på gruppa brukarar som kan det aktuelle språket, dukkar følgjande bilde opp: Berre tre Wikipedia-utgåver hadde i juni 2021 meir enn éin aktiv brukar som kunne språket den siste månaden: Sakha (15 brukarar), færøysk (9), og enaresamisk (8). Følgjande Wikipedia-utgåver hadde éin (1) brukar med kunnskap om språket i same periode: Nordsamisk, vepsisk, komi og burjatisk. Resten av Wikipedia-utgåvene med meir enn 1000 ord (livvisk, komipermjakisk, tuvinsk, grønlandsk, inuktitut, og iñupiaq) vart ikkje redigert av brukarar med kunnskap om språket dei siste 30 dagane før målinga i det heile.

Viss vi så ser på settet med aktive brukarar som heilskap, ser vi at 1/3 av bidragsytarane (44 av 146) var aktive på meir enn éin av dei sirkumpolare Wikipedia-utgåvene. Dette er bidragsytarane som ikkje kan det aktuelle språket. Så lenge desse brukarane kjenner avgrensingane sine, er bidraget deira absolutt velkomne. Eit døme er brukaren *Glorious 93*, ein fransk brukar som i evalueringssperioden var aktiv på 8 av dei 13 sirkumpolare Wikipedia-utgåvene, og dessutan i fleire inkubatorutgåver. Bidraget til desse brukarane inneber vanlegvis å legge til eller oppdatere referansar til bilete i infoboksar, eit veldig nyttig arbeid utan negative konsekvensar. Ein representant for den andre typen av ein (meir) kunnskapsrik Wikipedia-brukar er *Frhdkazan*, ein brukar som er aktiv på mange av Wikipedia-utgåvene i Russland. I følgje brukarsida hans er morsmålet tatarisk, som er eit tyrkisk språk. Språkutgåvene han skriv på har ein grammatisk struktur som liknar på tatarisk. Han er såleis i stand til å forstå strukturen i setningane i artiklane han redigerer eller opprettar, og med ein streng redigeringspraksis er sjansane for at teksten han skriv skal vere av akseptabel kvalitet ganske gode.

Den tredje typen ikkje-språkbrukarar som skriv på Wikipedia, kallar eg *Wikipediakaprarar*. Disse brukarane vel vanlegvis ei Wikipedia-utgåve med liten eller ingen aktivitet, og bidraget deira dominerer raskt Wikipedia-utgåva. Eit slikt eksempel er den sørsamiske Wikipedia-utgåva⁷. Dei to mest sentrale overskriftene («Forside» og «Velkommen til Wikipedia, gratis oppslagsverk») inneheld begge katastrofale feil. For å omsetje «side» i ordet «forside» brukte brukaren det sørsamiske ordet *bielie* («halvpart, den eine av to sider, t.d. oversida eller høgresida»), der det korrekte ville ha vore *sæjroe* («side i ei bok»). For å omsetje ordet «gratis» brukte skribenten det sørsamiske ordet *såangedidh* («å fri, å be nokon om å bli ektefelle»). Begge desse feila kjem av at

⁶ Evalueringen ble gjort 6. juni 2021.

⁷ <https://incubator.wikimedia.org/wiki/Wp/sma/Aalkoebielie>

orda «side» og «fri» er tvetydige på norsk og dermed har to omsetjingar i *norsk - sørsamisk ordbok*. I staden for prøve å finne rett omsetjing tok skribenten berre den første og beste omsetjinga, med ei meiningslaus forside som resultat «Forhalvpart. Velkomen til å fri til kunnskapsbok». Teksten er òg grammatiske feil: i uttrykket *såångedidh daajroegærjase*, som prøver å seie «[Velkommen] til det gratis oppslagsverket», er den bøygde forma *gærjase* (< gærja, «bok») feil; den riktige versjonen er gærjese. Desse setningane er ikkje eingong henta frå innhaldssidene, dei er ein del av lokaliseringa av den sørsamiske Wikipedia-utgåva. «Arbeid» av denne typen er i beste fall gjort i god tru, av menneske med total mangel på innsikt i sine egne språkkunnskapar, men det har alvorlege konsekvensar for det aktuelle språkmiljøet. Sjølv om språkfelleskapet er heilt uvitande om at dei har ei Wikipedia-utgåve på sitt eige språk, vil disse feilaktige formene dukke opp i Google-søk, og dermed gjere det umogleg å bruke Internett for å verifisere korrekt språkbruk. Wikipedia-utgåvene blir kopiert og lagt ut på mange andre serverar, og eit søk vil dermed gje inntrykk av at desse feilaktige formene har eit solid brukargrunnlag.

Organisert hjelp for brukarane: Wikimedia-avdelingane

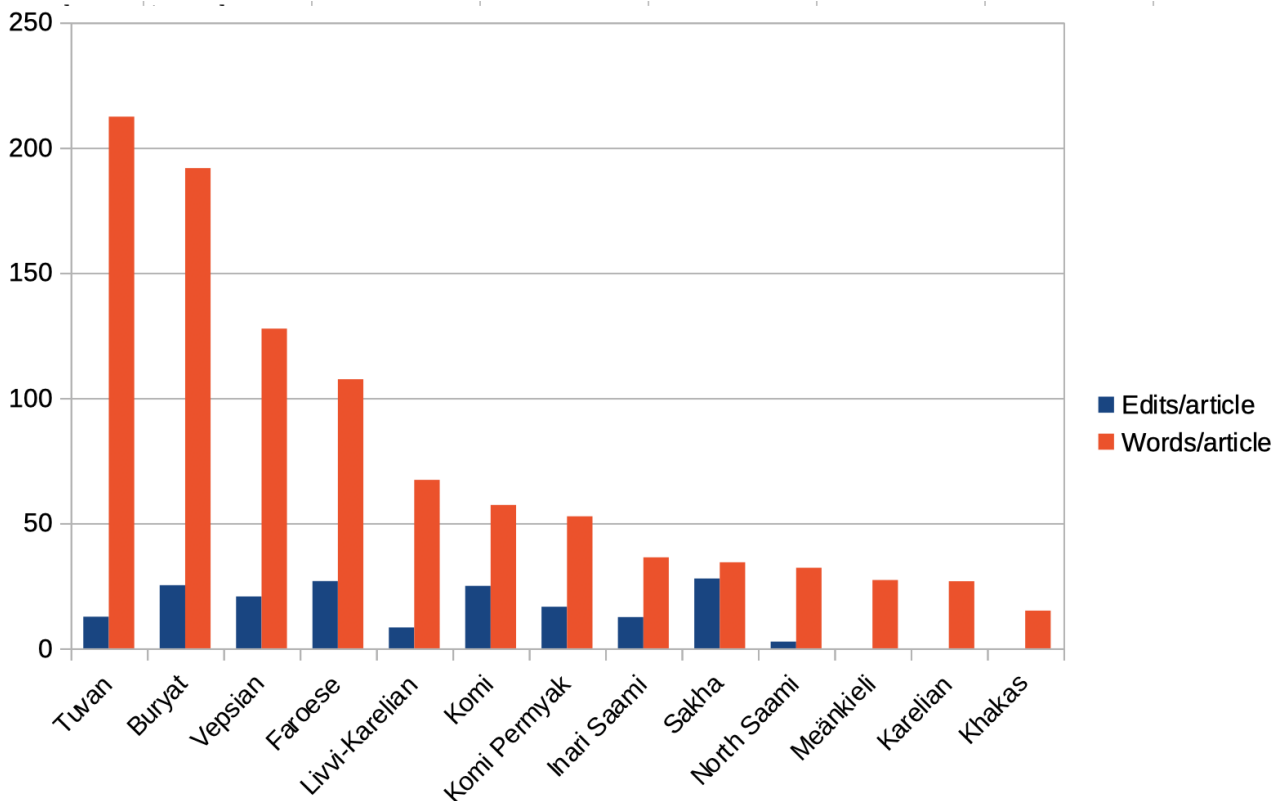
Både i Russland, Noreg og Finland støttar dei nasjonale Wikimedia-avdelingane minoritetsspråklege utgåver av Wikipedia-utgåvene på språka i dei respektive landa sine. Det har t.d. vorte halde wikiseminar for fleire av språka i Russland. Ein presentasjon av arbeidet er Fatkullin 2017. Tilsvarande har Wikimedia-avdelingane i Noreg og Finland arrangert kurs for potensielle forfattarar og hjelpt med malar, infoboksar og formatering.

Lengda på artiklane

Viss vi ser på dei 12 utgåvene med meir enn 1000 artiklar (i denne samanhengen kallar vi dei «store Wikipedia-utgåver»), er det berre ei som har kortare artiklar, den for khakassisk, med 15 ord/artikkel, som vist i Figur 4. Tuvinsk og burjatisk har dei lengste artiklane, med ei gjennomsnittleg artikkellengde på høvesvis 212 og 192 ord.

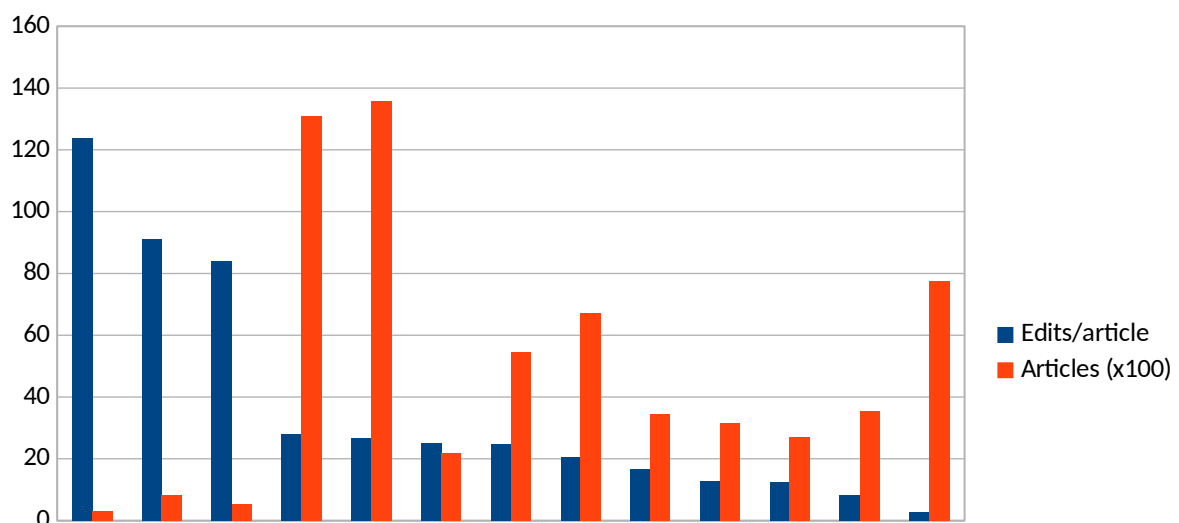
32 av dei sirkumpolare Wikipedia-utgåvene har færre enn 1000 artiklar⁸; disse vil eg referere til som «dei små Wikipedia-utgåvene». I tillegg til å ha få artiklar, har kvar artikkel også få ord. Ser vi på kvar Wikipedia-utgåve separat, var mediantalet for de større utgåvene 52,7 ord per artikkel, mens det for dei små var 11,5.

⁸ De små Wikipedia-utgåvene er i fallende rekkefølge: Grønlandsk, ingrisk, kvensk, inuktitut, tsjuksji, inupiaq, enetsisk, sentralalaskayupik, sørsamisk, kildinsamisk og mansisk. 10–99 artiklar: Skoltesamisk, ketisk, tundranenetsisk, tlingit, ludisk, lulesamisk, tundrajukagirsk, nanai, evenkisk, sibirsyupik og gwich'in. 1–9 artiklar: Aleutsik, khantisk, dolgansk, udegejsk, dene suline, orotsjisk, ultsjisk, negidal, tlingit og hän.



Vi går deretter over til talet på redigeringar for kvar artikkel. Viss vi tar verdas 321 Wikipedia-utgåver⁹ som referansepunkt, har dei 56 millionar artiklane der vorte redigert i gjennomsnitt 53 gonger kvar.

Data for redigeringane per artikkel var berre tilgjengeleg for Wikipedia-utgåvene, ikkje i inkubatoren. Figur 5 viser dei sirkumpolare Wikipedia-utgåvene som ikkje lenger er i inkubatoren, ordna etter tal redigeringar per artikkel (blå søyler), men samtidig med tal på artiklar (raude søyler, verdiane viser *artiklar x 100* for å kunne få dei i same tabell som talet på redigeringar).



9 Du finn eit oversyn over Wikipedia-utgåvene her: <http://wikipedia.org>.

Som det kjem fram av figuren, er det ingen samanheng mellom tal artiklar og tal redigeringar per artikkel. I staden avslører figuren forskjellige typar Wikipedia-utgåver. Utgåvene på grønlandsk, inuktitut og iñupiaq skil seg ut med svært få artiklar, men mange redigeringar per artikkel. Færøysk, Sakha, og til ei viss grad òg Komi, er større Wikipedia-utgåver med meir enn 20 redigeringar per artikkel. På det andre ytterpunktet står spesielt livvisk og nordsamisk, med svært få redigeringar per artikkel.

Analyse

Analysen vil sjå nærare på forskjellane mellom dei sirkumpolare Wikipedia-utgåvene og undersøke kvifor nokre utgåver er større enn andre. Han vil òg diskutere kva rolle desse utgåvene kan ha i språkvitaliseringsarbeidet, ikkje berre som skriveøving, men òg som oppslagsverk.

Dei sterkaste Wikipedia-utgåvene ser ut til å falle i to kategoriar. Den første kategorien består av utgåver som har både mange artiklar per språkbrukar og til ei viss også mange redigeringar per artikkel. Det er to Wikipedia-utgåver i denne kategorien: enaresamisk og vepsisk. Dette er altså utgåver der forfatarane både skriv mange artiklar og bruker tid på kvar artikkel. For enaresamisk er grunnen til dette klar: Revitaliseringa av enaresamisk blir gjennomført på ein planlagt måte av Anarâškielâ servi (Foreininga for enaresamisk), og det blir retta mot éin generasjon språkbrukarar av gangen. Den noverande prioritetten deira er å styrke lese- og skrivekunnskap og ein del av strategien er å bruke Wikipedia (jf. Olthuis et al 2021, Brecciaroli 2021). Forfatarane bak vepsisk Wikipedia kan ha vorte inspirert av arbeidet med dei finske og estiske Wikipedia-utgåvene.

I den andre kategorien er Wikipedia-utgåvene med ei relativt stor befolkning støtta av ein sterk administrasjon som fungerer på morsmålet deira. Innanfor denne kategorien fell Wikipedia-utgåvene på færøysk, sakha og komi. Forfatarane kan vere ein del av ei vellykka revitaliseringsrørsle og ha ein sterk motivasjon, slik enaresamiske forfatarar gjer, eller dei kan tilhøyre eit språkfelleskap med sterk skriftkultur i skule og samfunn, som strekker seg over fleire generasjonar, som i tilfellet med færøysk, sakha og komi.

Dei nordsamisk og livviske Wikipedia-utgåvene fell utanfor dette bildet. Samanlikna med storleiken på språkfelleskapet, inneheld dei mange artiklar, men det er lite aktivitet på kvar artikkel. burjatisk og tuvinsk har mange språkbrukarar samanlikna med dei relativt få artiklane. Kontrasten til Sakha og Komi kan òg vere tilfeldig.

Konklusjon

Hovudkonklusjonen er at svært få av dei sirkumpolare Wikipedia-utgåvene er aktive prosjekt som blir drivne av brukarar som kan skrive tekstar på det aktuelle språket. Berre 3 Wikipedia-utgåver så redigeringar frå meir enn ein redaktør som faktisk kjente språket i løpet av ein tilfeldig månad i 2021, og ytterlegare 4 utgåver hadde redigeringar frå berre éin redaktør som kunne språket. Dei fleste Wikipedia-utgåvene er små, median tal artiklar er 23, og berre 13 av 45 sirkumpolare Wikipedia-utgåver har meir enn 1000 artiklar. Fleirtalet av Wikipedia-utgåvene blir dermed oppretta og skriva av folk som ikkje kan språket og som ikkje tilhøyre språksamfunnet.

Dei vellykka sirkumpolare Wikipedia-utgåvene er enten forankra i aktive og vellykka revitaliseringsrørsler, til dømes enaresamisk, eller dei utgjer eit stort fellesskap av brukarar av språket med ein offisiell språkpolitikk som favoriserer det sirkumpolare språket og skriftkulturen. Dette er i alle fall til ein viss grad tilfelle for sakha, færøysk og komi.

Konklusjonane som kan trekkast for Wikimedia Foundation er at minoritetsspråklege samfunn ikkje er som andre språksamfunn, og det å forvente at dei skal oppføre seg som majoritetsspråklege samfunn vil føre oss på villspor. Ei betre forståing av dei sirkumpolare språkmiljøa vil forbetre evna vår til å tilby Wikipedia til desse samfunna, og i enkelte sjeldne tilfelle til og med bidra til å styrke skriftkulturen deira.

Litteratur

Bilinsky, Yaroslav, 1962: The Soviet Education Laws of 1958-9 and Soviet National Policy. *Soviet Studies* 14:138-157. Oxford

Brecciaroli, Fabrizio 2021: *The Inari Saami Wikipedia*. Wikimedia: Arctic Knot

Fatkullin, Farhad 2017: *Wikipedias in the languages of Russia: in line with the wider Movement strategy?* Plakat på Wikimania 2017-konferansen. <https://meta.wikimedia.org/wiki/File:WLR-Wikimania2017-poster-frhdkazan.pdf>

Frederiksen, Katti og Carl Christian Olsen 2017: *Det grønlandske sprog i dag*. Rapport over det grønlandske sprog, standpunkt og anbefalinger. Nuuk.

Isaev, M.I. 1977: *National Languages in the USSR: Problems and Solutions*. Russian original: Resjenie natsional'no-jazykovykh problem v SSSR. Progress Publishers, Moskva.

Knudsen, Karin Jóhanna L. 2010. Language use and linguistic nationalism in the Faroe Islands, *International Journal of Multilingualism*, 7:2, 128-146, DOI: 10.1080/14790710903518404

Kreindler, Isabelle (red.) 1985: *Sociolinguistic Perspectives on Soviet National Languages*. Contributions to the Sociology of Language 40. Mouton de Gruyter.

Meek., A. B. 2010: *We Are Our Language. An Ethnography of Language Revitalization in a Northern Athabaskan Community*. Tucson, Az.: The University of Arizona Press.

Arbeids- og inkluderingsdepartementet 2009: [Handlingsplan for samiske språk](#). Oslo: Statens forvaltningstjenester.

Olthuis, Marja-Liisa, Suvi Kivelä og Tove Skutnabb-Kangas 2013: *Revitalising Indigenous Languages. How to Recreate a Lost Generation*. Bristol: Multilingual Matters.

Olthuis, Marja-Liisa, Trond Trosterud, Erika Katjaana Sarivaara, Petter Morottaja og Eljas Niskanen 2021: [Metodiske implikasjoner av prosjektet Čyeti čälled anarâškielân, 'One Hundred Writers for Aanaar Saami'](#) In Indigenous Research Methodologies in Sámi and Global Contexts. New Research - New Voices, Volume: 11

Pasanen, Annika 2015: *Kuávsui já peeivičuová. 'Sarastus ja päivänvalo': Inarinsaamen kielen revitalisaatio*. PhD. dissertation, Helsinki: Helsingin yliopisto.

Pasanen, Annika 2016: *Saamebarometri 2016 : selvitys saamenkielistä palveluista saamelaisalueella = Sámi giellabaromehter 2016 : čielggadus sámegielat bálvalusain sámeguovllus*. Helsinki: Oikeusministeriö 2016.

Trosterud, Trond 1995: Sovjetisk språkpolitikk. Finsk-ugriske språk i Russland i eit historisk-politisk perspektiv. *Nordisk Østforum* 3:40-45.

Vedlegg: Tabeller som viser de sirkumpolare Wikipedia-utgavene

Tegnforklaring:

Fam = språkfamilie (dene = Na Dene, ea = Eskaleut, dvs. = indoeuropeisk, mo = mongolsk, ps = paleo-sibirisk, tun = tungusisk, tur = tyrkisk, ur = uralsk).

Reg = region (a = Nord-Amerika, n = Norden, r = Russland)

Arts = Antall artikler

Ed/art = Antall redigeringer per artikkel

last30 = Aktive brukere (alle brukere), siste 30 dager

eds l30 = Antall redigeringer siste 30 dager

Spkrs = Språkbrukere (Kilde angitt på deres respektive sider på <http://en.wikipedia.org>, hovedsakelig Ethnologue.

Art/Spkr = Wikipedia-artikler per språkbruker

Wrds = Totalt antall ord i Wikipedia-utgaven, regnet som innholdssider fra de respektive wikipedia-dumpene i april eller mai 2021, regnet som tekst mellom tomme

Wrds/art = Ord per artikkel

Orth = År for gjeldende ortografi, kilde hovedsakelig tatt fra <http://ru.wikipedia.org>

De sirkumpolare Wikipedia-utgavene som ikke er i inkubatoren

Språk	Fam	Reg	Adresse	Redige		last30	eds l.30	Spkrs	Art/spkr	Wrds	Wrds/art	Orth	
				Arts	ringer								
Inuktitut	ea	a	iu.wikipedia.org	517	43322	83,79	13	36	39475	0,01	6898	13,34	1855
Inupiaq	ea	a	ik.wikipedia.org	289	35723	123,61	7	11	2144	0,13	1803	6,24	1946
Færøysk	ie	n	fo.wikipedia.org	13581	364420	26,83	43	124	72000	0,19	1459615	107,47	1854
Nordsamisk	ur	n	se.wikipedia.org	7737	20470	2,65	29	75	25000	0,31	248741	32,15	1979
Enaresamisk	ur	n	smn.wikipedia.org	2670	33252	12,45	8	959	400	6,68	96942	36,31	1996
Grønlandsk	ea	n	kl.wikipedia.org	810	73832	91,15	15	13	57000	0,01	29683	36,65	1973
Sakha	tur	r	sah.wikipedia.org	13077	364499	27,87	39	438	450000	0,03	449228	34,35	1939
Vepsian	ur	r	vep.wikipedia.org	6659	137729	20,68	21	487	1600	4,16	850353	127,70	2007
Komi	ur	r	kv.wikipedia.org	5441	135522	24,91	21	99	160000	0,03	311417	57,24	1938
Livvi-Karelian	ur	r	olo.wikipedia.org	3545	29478	8,32	14	151	31000	0,11	238428	67,26	2007
Komi Permyak	ur	r	koi.wikipedia.org	3460	57409	16,59	13	107	63000	0,05	182428	52,72	1938
Tuvan	tur	r	tyv.wikipedia.org	3167	39925	12,61	9	101	280000	0,01	672601	212,38	1943
burjatisk	mo	r	bxr.wikipedia.org	2181	54907	25,18	15	89	265000	0,01	418338	191,81	1939

De sirkumpolare Wikipedia-utgavene i inkubatoren

Språk	Fam	Cntr	Adresse	Arts	Spkrs	Art/spkr	Ord	wrds/art	Orth
Central Alaskan Yupik	ea	a	incubator.wikimedia.org/wiki/Wp/esu	193	19750	0,01	1681	8,71	1984
Tlingit	dene	a	incubator.wikimedia.org/wiki/Wp/tli	47	1360	0,03		0,00	1973
Gwich'in	dene	a	incubator.wikimedia.org/wiki/Wp/gwi/	20	560	0,04		0,00	1970
Aleut	ea	a	incubator.wikimedia.org/wiki/Wp/ale	8	150			0,00	1990
Dene Suline	dene	a	incubator.wikimedia.org/wiki/Wp/chp	4	11325	0,00		0,00	1904
Han	dene	a	incubator.wikimedia.org/wiki/Wp/haa	1	20	0,05	1	1,00	1970
Tlingit	dene	a	incubator.wikimedia.org/wiki/Wp/tli	1	1360	0,00	9	9,00	1973
Meänkieli	ur	n	incubator.wikimedia.org/wiki/Wp/fit	1841	30000	0,06	50167	27,25	1985
Kvensk	ur	n	incubator.wikimedia.org/wiki/Wp/fkv	610	5000	0,12	16191	26,54	2008
Sørsamisk	ur	n	incubator.wikimedia.org/wiki/Wp/sma	160	600	0,27	1031	6,44	1974
Skoltesamisk	ur	n	incubator.wikimedia.org/wiki/Wp/sms	90	320	0,28	1149	12,77	1973
Lulesamisk	ur	n	incubator.wikimedia.org/wiki/Wp/smj	43	650	0,07	379	8,81	1983
Karelsk	ur	r	incubator.wikimedia.org/wiki/Wp/krj	1808	36000	0,05	48383	26,76	2007
Khakas	tur	r	incubator.wikimedia.org/wiki/Wp/kjh	1445	42604	0,03	21745	15,05	1939
Ingrian	ur	r	incubator.wikimedia.org/wiki/Wp/izh	677	820	0,83	31391	46,37	2014

Chukchi	ps	r	incubator.wikimedia.org/wiki/Wp/ckt	415	5095	0,08	3923	9,45	1937
Enets	ur	r	incubator.wikimedia.org/wiki/Wp/enf	197	43	4,58	654	3,32	1980
Kildinsamisk	ur	r	incubator.wikimedia.org/wiki/Wp/sjd	143	340	0,42	5165	36,12	1985
Mansi	ur	r	incubator.wikimedia.org/wiki/Wp/mns	118	940	0,13	4053	34,35	1938
Falt av	ps	r	incubator.wikimedia.org/wiki/Wp/kef	71	213	0,33	734	10,34	1980
Tundra Nenets	ur	r	incubator.wikimedia.org/wiki/Wp/yrk	66	21926	0,00	3555	53,86	1937
Ludic	ur	r	incubator.wikimedia.org/wiki/Wp/lud	43	3000	0,01	393	9,14	1993
Tundra Yukaghir	ps	r	incubator.wikimedia.org/wiki/Wp/ykg	41	370			0,00	1983
Nanai	ps	r	incubator.wikimedia.org/wiki/Wp/gld	40	1400	0,03	1628	40,70	1937
Evenki	tun	r	incubator.wikimedia.org/wiki/Wp/evn	26	26580	0,00	2526	97,15	1937
Siberian Yupik	ea	r	incubator.wikimedia.org/wiki/Wp/ess	26	1300	0,02	183	7,04	1937
Dolgan	tur	r	incubator.wikimedia.org/wiki/Wp/dlg	4	1100	0,00	135	33,75	1961
Khanty	ur	r	incubator.wikimedia.org/wiki/Wp/kca	4	9600	0,00	159	39,75	1937
Udege	tun	r	incubator.wikimedia.org/wiki/Wp/ude	4	100	0,04	133	33,25	1989
Oroch	tun	r	incubator.wikimedia.org/wiki/Wp/oac	3	8	0,38	123	41,00	1994
Ulch	tun	r	incubator.wikimedia.org/wiki/Wp/ulc	3	150	0,02	116	38,67	1989
Negal	ps	r	incubator.wikimedia.org/wiki/Wp/neg	2	75	0,03	117	58,50	1992

De sirkumpolare språkene uten Wikipedia-utgave er:

Ahtna, Aluutiq, Kyst-tsimshian, Deg Xinaq, Dena'ina, Forest Nenets, Gitxsan, Haida, Inuttitut, Inuvialuktun, Kaska, Kokyma Yukaghir, Koyukon, Kuskokwim, Lower Tanana, Naukan Yupik, Nganasan, Nisga'a, Nord-Slaa Pitesamisk , Sørslavey, Sørtutchone, Th̄chq (Dogrib), Tanacross, Umesamisk og Upper Tanana.