

The circumpolar Wikipedia editions

Trond Trosterud, UiT The Arctic University of Norway

Abstract

The multilingual encyclopedia Wikipedia seems to have a particular stronghold in the circumpolar region. Of the approximately 70 circumpolar languages, i.e. languages indigenous to that region, almost all have a literary language, and 45 of these have their own Wikipedia or are in the Wikipedia incubator. Compared to the global situation, both Wikipedia's position and the degree of literacy in the far North are remarkable: According to *ethnologue.com*, there are 7117 living languages, of which more than half (3982) have a literary language and only 620 have a Wikipedia or are in the incubator. If we look more closely at these circumpolar Wikipedia editions, it will become clear that these numbers are misleading. This article investigates in what way they are misleading in order for us to better understand indigenous language literacy, the situation for the circumpolar languages in their respective countries, and how small Wikipedia editions are created.

Introduction

This article will analyse the content of the 45 Wikipedia or Wikipedia incubator editions written in a circumpolar minority language. By *circumpolar language*, we refer to a minority language spoken by an autochthonous population in an area north of or close to the 60th parallel north, excluding the state languages.

The article is structured as follows: section 2 presents the circumpolar language communities, their literacy, and their Wikipedia editions. Section 3 analyzes the differences between the various editions in a circumpolar language, the reasons for this, and the role these editions play in language revitalization. Section 4 summarises what we now know about these editions and the conditions that have led to some of them being successful and concludes with a few recommendations on what we can do better in the future.

Background

The circumpolar languages

If we divide the circumpolar area into three regions, the North American region has 30 circumpolar languages, the Russian region 31, and the Nordic region 10. The dominating language families are Uralic (23 circumpolar languages in the Nordic countries and Russia), Dene (19 circumpolar languages in North America), and Eskaleut (10 languages across all 3 regions). The remaining 20 languages are either isolates or Turkic or Tungusic languages, with the exception of the Germanic Faroese.

Most circumpolar languages are small; the median number of speakers for the languages discussed here is 650. In decreasing order, the 7 largest languages (50,000 to 500,000 speakers) in the sample are Sakha, Tuvan, Buryaad, Komi, Faroese, Komi Permyak, and Greenlandic. All of them have their

own administrative entity where the language has official status, either an autonomic republic or area in Russia or its own islands with home rule, as with Greenlandic and Faroese.

Three of the languages are in the majority within their autonomous area¹: Faroese 94 %, Tuvan 91%, Greenland 85%, and Inuktitut 63 %. The other languages are in the minority (Sakha 38%, Buryaad 27%, and Komi 18%). The rest of the languages have either a weaker status or no official status.

Typologically, all the circumpolar languages have a complex morphology. Both the Na-Dene and Eskaleut languages are polysynthetic, and they use verbal inflection to express the grammatical and pragmatic information expressed by pronouns and adverbs in the Germanic languages. The Na-Dene languages are predominantly prefixing languages, whereas the Eskaleut ones are suffixing. The circumpolar languages spoken in Russia and the Nordic countries are predominantly suffixing. For most of these languages, the case systems are large, typically with between 6 and 10 cases, but the Komi languages and Udmurt have between 20 and 30. The minority languages of the Nordic countries have fewer cases (less than 10), but they have more complex systems of morphophonological stem alternation during inflection. All in all, the circumpolar minority languages differ markedly from their respective majority languages. The largest linguistic distances to the majority language are found for the circumpolar languages in North America, for Greenlandic, and for the Saami languages in Norway and Sweden. Speakers of Russian and of Finnish, in particular, will have a somewhat shorter linguistic distance to the minority languages in their respective countries.

Literacy among the circumpolar languages

Literacy in the North can be traced back to three different initiatives for alphabetisation. The first one was missionary work in connection with state consolidation from the 17th up to and including the 19th centuries. This resulted in the Latin script being used to create the orthographies in the Nordic countries, Cyrillic in Russia and Alaska, and both syllabic and Latin in North America. The second initiative took place only in Russia, where the October Revolution launched a planned, large-scale creation of literary languages based on the Latin script for all its minority languages, including the northern ones with the idea of eventually replacing the previous Cyrillic-based orthographies. Finally, in a process starting during the 1960s and 1970s, grassroot initiatives and a subsequent change or at least a relaxation in the assimilation has resulted new orthographies being created for some of the circumpolar languages and for others, the older orthographies being revised.

Looking closer at the different parts of the circumpolar areas, we see that in North America, early missionary initiatives using the native language from the 18th century onwards were replaced with an assimilatory policy that increased with the incorporation of northern people into modern society during the 20th century. As a part of reconciliation, in 1999 Canada separated Nunavut from the Northwest Territory into its own region. In 2008, it gave Inuktitut the status of official language and started publishing the proceedings of the Nunavut Legislative Assembly in both English and Inuktitut. Most circumpolar languages in Canada are taught on their First Nation reserves, but mostly as foreign languages with English as the language of instruction, and with poor results (see Meek 2010 for a thorough analysis of this practice). In Alaska, the assimilatory policy changed even less than it did in Canada. In the 1960s, the University of Fairbanks started creating practical

¹ Canada: [Statistics Canada 2016](#). Faroese Islands: [Statbank hagstova 2011](#) , Greenland: Frederiksen and Olsen 2017. Russia: [Всероссийская перепись населения 2010](#).

orthographies for several Native American languages in Alaska, but this did not result in systematic changes to the school curricula. The declaration of 20 Alaskan native languages as "co-official languages with English" in 2014 also did not result in any radical shift in policy².

In Russia, the creation of orthographies in the 1930s was accompanied by administrative changes where ethnically blind administrative borders were replaced with a system of autonomous republics and for the smaller minorities, areas created around all minorities large enough to constitute an administrative body. During the first decades after the Second World War, languages with their own autonomous republic (Komi, Sakha, Tuvan, Buryaad, Khakas) were used as the language of instruction in primary schools. The republics published journals and newspapers in their own language, and novels were written and published in the language of the republics as well. The speakers of the other circumpolar languages of Russia were at least taught to read and write their mother tongues, and there were publications using these languages as well. With the new school laws introduced in 1958-59, indigenous literacy was weakened, but it still remained on a level far stronger than the one for the minority languages in the other circumpolar regions³.

After having been governed in Danish for centuries, the Faroese started working on their own written language at the end of the 19th century. The British occupation during World War II weakened Danish influence, and after the introduction of home rule in 1948, Faroese gradually became the dominant language in writing too (cf. Knudsen 2010). In Greenland, Greenlandic was used across the whole school system, including teacher education, and even in a newspaper, but in 1953, this 100-year-old tradition of native literacy was shattered when Greenland went from being a Danish colony to a county (*amt*) integrated within Denmark. The transition was combined with a large-scale shift in literacy, from Greenlandic to Danish. Even though the introduction of home rule in 1979 and of self-governance in 2008 made Greenlandic the only official language of Greenland, Danish still has a strong position, especially as the dominating language of all theoretical education above the compulsory level. The literacy of both administration, the legal system, and mass media is bilingual, with text typically written in Danish and translated into Greenlandic (Frederiksen and Olsen, op.cit.).

Over the last four decades, the Nordic countries have slowly implemented a new language policy for their minority languages, resulting in official status on a regional level or status as a national minority language for the Saami languages, Kven and Meänkieli. Large parts of the curriculum for North Saami children is taught in North Saami and to a lesser extent this holds for Skolt and Inari Saami as well. All of the circumpolar minority languages in the Nordic countries are being taught as school subjects⁴. One should note the exceptional case of Inari Saami, where an on-going language shift during the last generation has been turned into a situation where the majority of speakers belong to the younger generation, and the language community publishes both journals, textbooks, fiction, and news broadcasts in Inari Saami⁵.

2 Cf. <http://www.akleg.gov/basis/Bill/Text/28?Hsid=HB0216Z>, the new paragraph entitled "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, and Tsimshian languages as official languages of the state.". The law still explicitly states that this does not require any state or municipality government to print a document in any language other than English.

3 A presentation of the Soviet policy by one of its own architects is Isaev 1977. For analyses, see e.g. Kreindler 1985 and Trosterud 1995. Bilinsky 1962 gives an analysis of the effect of the school laws of 1958-59.

4 As an illustration of the policy in Norway, see the Action Plan for Sami Languages (Norwegian Ministry of Labour and Social Inclusion 2009). A survey of the situation for Saami languages in Finland is given in Pasanen 2016.

5 Pasanen 2015 is the standard reference for the revitalisation process. An in-depth analysis of the restoring of the "middle generation" (the one between the older speakers and the children of the language nests) is Olthuis et al.

So for the circumpolar languages as a whole, this implies that only a quarter of the current orthographies date back to before the Second World War. These orthographies are either Cyrillic-based orthographies from 1937 or they are syllabic scripts. Half of the circumpolar orthographies are less than 40 years old. Most of the circumpolar languages have the lion's share of their speakers in the upper age cohorts, the prototypical situation thus being that circumpolar languages have orthographies that are younger than their own speakers.

As for language use in social media, both Faroese, Greenlandic, Meänkieli, North Saami, and Inari Saami and to a certain extent also Skolt Saami, South Saami, Lule Saami, Karelian, Livvi and Vepsian can be found in use on Facebook, both in groups devoted to the respective languages, and in personal postings. In North America, the situation seems to be different, as Facebook groups devoted to the largest languages (Inuktitut, Central Alaskan Yupik, Dene Suline, Iñupiaq) all used English only. Native language words were typically quoted (as in "what does X mean"), but the discussion was always in English. Buryaad, Komi, Tuvan and Khakas are in use on vkontakte⁶ groups devoted to use the language in question, Sakha less so. The other circumpolar languages of Russia with their own Wikipedia edition seem to either have their vkontakte language groups run in Russian, or do not have a group at all.

The circumpolar Wikipedia editions

We now move on to the concrete Wikipedia editions. After a short overview, we look at the size of the editions and their language community, then at the writers, and finally at the state of the articles.

The Wikipedia editions

The map in Figure 1 shows the approximate location of the Wikipedia editions with more than 200 articles with orange ovals and with more than 1000 articles with blue ovals. As can be seen from Figure 1, the Wikipedia editions containing more than 1000 words are found mostly in the Nordic countries, but also in Russia with the republic languages. In addition to the languages shown on the map, there are 24 Wikipedia editions containing less than 200 words. From largest to smallest, 6 of these are in North America (Central Alaskan Yupik, Tlingit, Hän, Gwich'in, Aleut, and Dene Suline), 15 in Russia (Kildin Saami, Mansi, Ket, Tundra Nenets, Ludic, Tundra Yukaghir, Nanai, Evenki, Siberian Yupik, Khanty, Dolgan, Udege, Ulch, Oroch, and Negidal), and in the Nordic countries (South Saami, Skolt Saami, and Lule Saami). Characteristic of the languages in this group is that they have few speakers and weak or no literary traditions. The Wikipedias are typically started by non-speakers from outside the language community, some of the articles are in the majority language, and there is currently no activity on any of them.

2013.

6 vkontakte is the Russian version of Facebook, cf. <https://vk.com/>

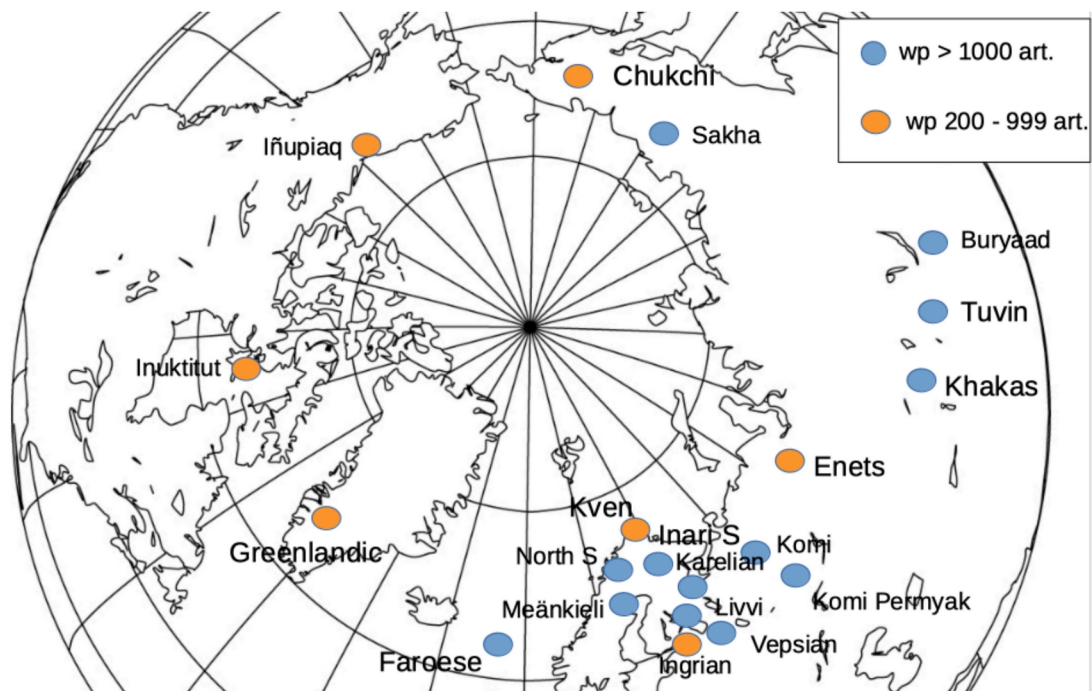


Figure 1: Circumpolar Wikipedia editions with more than 200 articles

Wikipedia size

Let us first take stock of the mere size of the Wikipedia editions containing more than 1000 words. They are shown in Figure 2, ordered first according to area (North America, Russia, the Nordic countries), and then by the number of words. As can be seen, the three areas have distinct profiles. For North America (left), the three largest Wikipedias barely have 100 words. The Wikipedia editions of Russia can (with a few exceptions, to which we will return) be divided into two groups: The ones with the support of its own republic, and the ones without. In the Nordic countries, Faroese is the outlier. North Saami is the largest one after Faroese, with Inari Saami following suit, and the rest being far smaller.

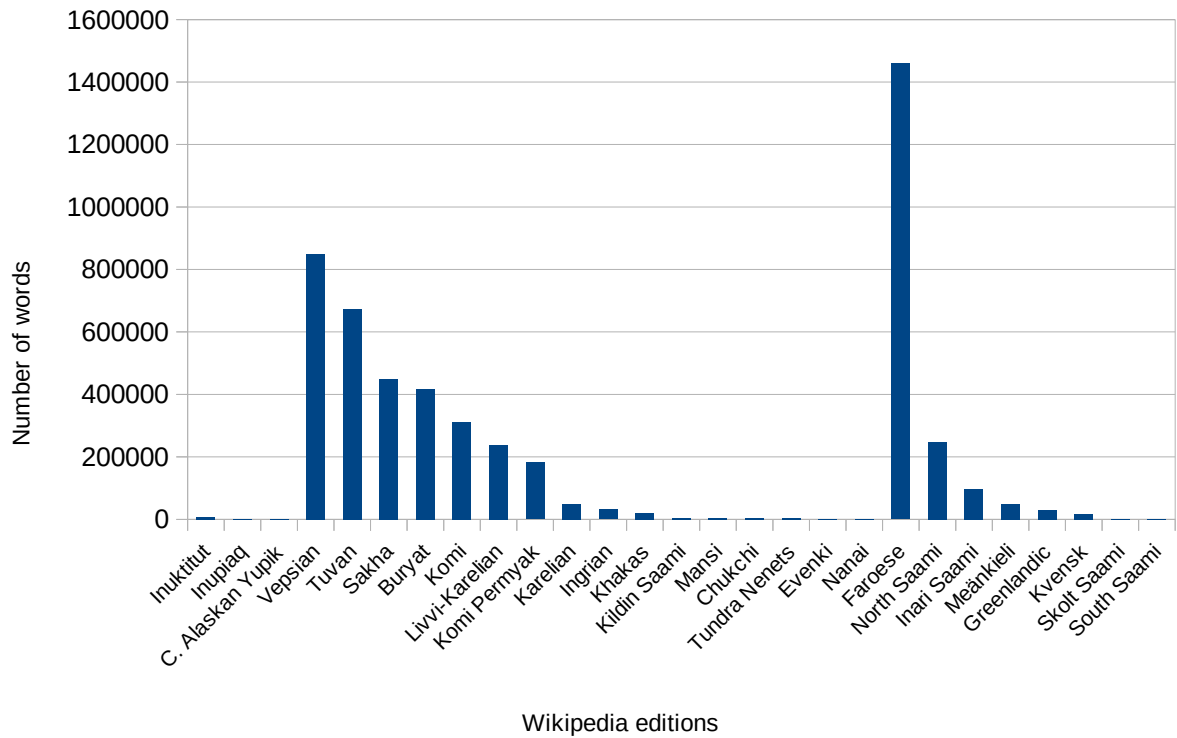


Figure 2: Wikipedia editions > 1000 words in North America, Russia, and the Nordic countries

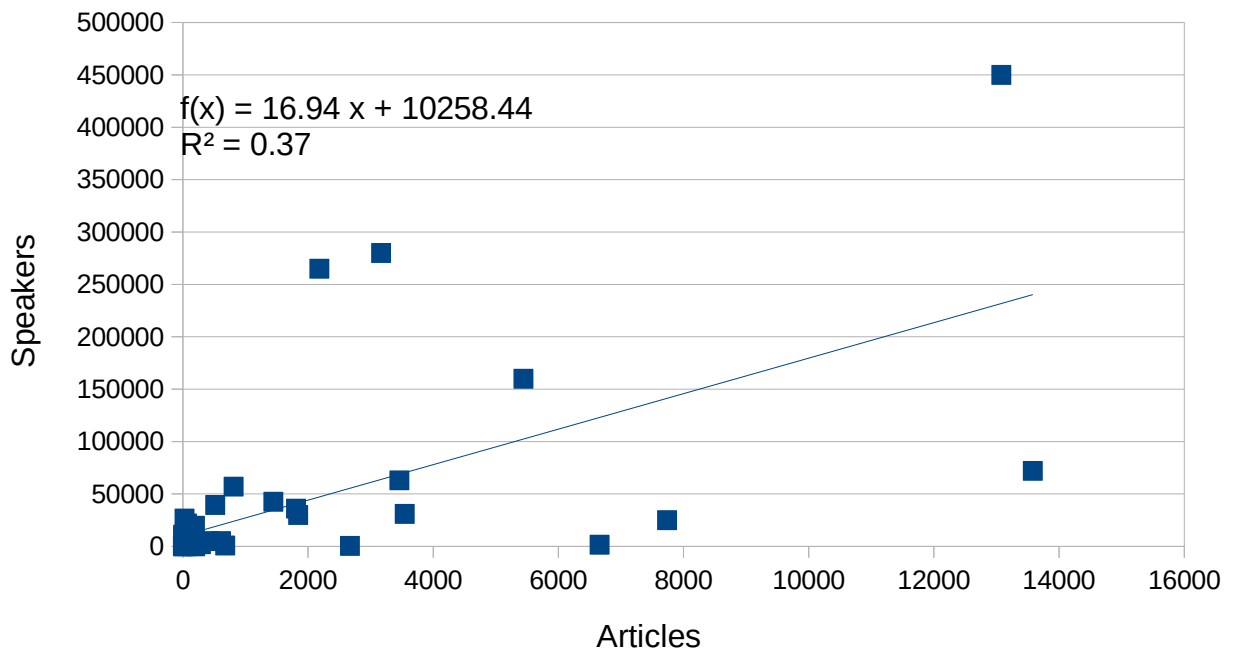


Figure 3: Correlation between the number of articles and the size of the language community for the circumpolar Wikipedia editions

Wikipedia size relative to the size of the language community

At the outset, one may think more speakers write more articles. As shown in Figure 3, there is indeed such a correlation, but it is weak ($R^2=0.37$, where $R^2=1.0$ would have been a perfect

correlation). The outliers with "too few articles" are Sakha, Komi, Tuvan, and Buryaad, whereas the ones with "too many articles" are Faroese, North Saami, Vepsian, and Inari Saami. The outliers with too few articles are also the four languages in Russia with the largest language communities. Compared to larger Wikipedia editions outside the circumpolar area they are still not too far off though. With 88 speakers/article, Buryaad is on a par with Albanian (72), and the 29 and 34 speakers/article for Komi and Sakha are compatible with Danish (21), Hungarian (30) and Bulgarian (33). Faroese is on a par with Icelandic (both have approximately 5.5 speakers/article).

Of the circumpolar languages with more than 10,000 speakers, 5 stand out with "too many articles", all of them within the Nordic countries or in Karelia: Faroese, North Saami, Vepsian, Livvi (Olonets), and Inari Saami. For the smallest language communities of Vepsian and Inari Saami, it is easy to have a low number of speakers/article. Still, these language communities stand out with higher activity than the rest, especially compared to their size.

The correlation between language community size and Wikipedia size is thus weak, and there must be other factors than language community size determining the size of the Wikipedias.

The writers

When the language communities are as small as they are for the circumpolar languages, the behaviour of individual writers plays a larger role than it does for larger language communities.

For each Wikipedia edition, there is a page under *Special pages* > *Active users list* that shows who has been active in the last 30 days. The Wikipedia term *Users* refers to the people or computer programs (bots) editing the Wikipedia editions.. Some of the listed users do know the language of the Wikipedia in question, but a large part or even most users do not. These non-speakers may be divided into three categories: users who do not touch the linguistic content, users with enough knowledge of both the language and their own limitations as a writer to produce acceptable text, and users who have little or no knowledge of the language yet still produce text, which typically varies between very poor quality and utter gibberish.

In order to classify the users into speakers and non-speakers, I looked at both self-declared proficiency and writing behaviour. Many users declare language proficiency on their user page. I trusted the ones declaring no proficiency and looked closer at the ones that either declared language proficiency or did not reveal their proficiency. For these users, I went through their contributions and classified the ones with contributions that did not require language skills as non-speakers⁷.

Looking first at the group of users who know the language in question, the following picture emerged: Only 3 Wikipedia editions had more than one active user with knowledge of the language in the last month: *Sakha* (15 users), *Faroese* (9), and *Inari Saami* (8). The following Wikipedia editions had one (1) user with knowledge of the language in the same period: *North Saami*, *Vepsian*, *Komi*, and *Buryaad*. The other editions with more than 1000 words (*Livvi*, *Komi Permyak*, *Tuvan*, *Greenlandic*, *Inuktitut*, and *Inupiaq*) were not edited by users with knowledge of the language in the last 30 days.

If we then look at the set of active users as a whole, we see that 1/3 of the contributors (44 of 146) are active on more than one of the circumpolar Wikipedia editions. These are the contributors that do not know the language in question. As long as these users know their limitations, their contributions are certainly welcome. One example is the user *Glorious 93*, a French user who in the

⁷ The evaluation was done June 6th, 2021.

last month, has been active on 8 of the 13 circumpolar Wikipedia editions, as well as on several Incubator editions. Their contributions typically involve adding or updating references to pictures in infoboxes, very useful work with no negative consequences. A representative of the second type of a (more) knowledgeable Wikipedian is the user *Frhdkazan*, who is active on many of the Wikipedia editions in the languages of Russia. According to his user page, his mother tongue is Tatar, which is a Turkic language. The language editions he writes on have a grammatical structure similar to the one of Tatar. He is thus able to understand the structure of the sentences of articles he edits or creates, and with a careful editing practice, the chances of getting article text of acceptable quality is quite good.

The third type of non-speakers writing on Wikipedia I call the *Wikipedia hijackers*. These Wikipedians typically pick a Wikipedia with little or no activity, and their contributions quickly dominate the Wikipedia edition. One such example is the South Saami Wikipedia edition⁸. The two most central headlines ("Front page" and "Welcome to Wikipedia, the free encyclopedia") both contain catastrophic errors. For the word for "page", the user used the word *bielie* ("side") and for "free", he used *såångedidh* ("to propose, to ask someone to become one's spouse"). Behind these mistakes are homonymies in Norwegian and thus in the Norwegian - South Saami dictionary. The Norwegian word *side* means both "side" and "page", whereas the word *fri* means both "free" and "propose". There are also grammatical errors: in the phrase *såångedidh daajroegærjase*, which attempts to say "[Welcome] to the free encyclopedia", the inflected form *gærjase* (< *gærja*, "book") is wrong; the correct version is *gærjese*. These phrases are not even taken from content pages, they are part of the localisation of the South Saami Wikipedia. "Work" of this type may be done in good faith, by people with a total lack of insight into their own linguistic capabilities, but it has grave consequences for the language community in question. Even if the language community is totally unaware of having a Wikipedia edition in their own language, these erroneous forms will show up in Google searches, thereby making it impossible to use the Internet to check for language usage.

Organised help for the users: The Wikimedia chapters

Both in Russia, Norway, and Finland, the national Wikimedia chapters support the minority language editions of the Wikipedias of the language communities in their respective countries. Wiki seminars have been held for several of the languages in Russia. A presentation of the work may be found in Fatkullin 2017. Similarly, the Wikimedia chapters of Norway and Finland have arranged courses for potential writers and helped with templates, infoboxes, and formatting.

Article length

If we look at the 12 editions with more than 1000 articles (in this context referred to as "large Wikipedia editions"), only one has shorter articles, the one for Khakas, with 15 words/article, as shown in Figure 4. Tuvan and Buryaad have the longest articles, with an average article length of 212 and 192 words, respectively.

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

32 of the circumpolar Wikipedias have less than 1000 articles⁹; these will be referred to as the smaller Wikipedia editions. In addition to having a small number of articles, the articles themselves were smaller. Looking at each Wikipedia edition separately, the median words/article count for the larger ones was 52.7 words/article, whereas for the smaller ones it was 11.5.

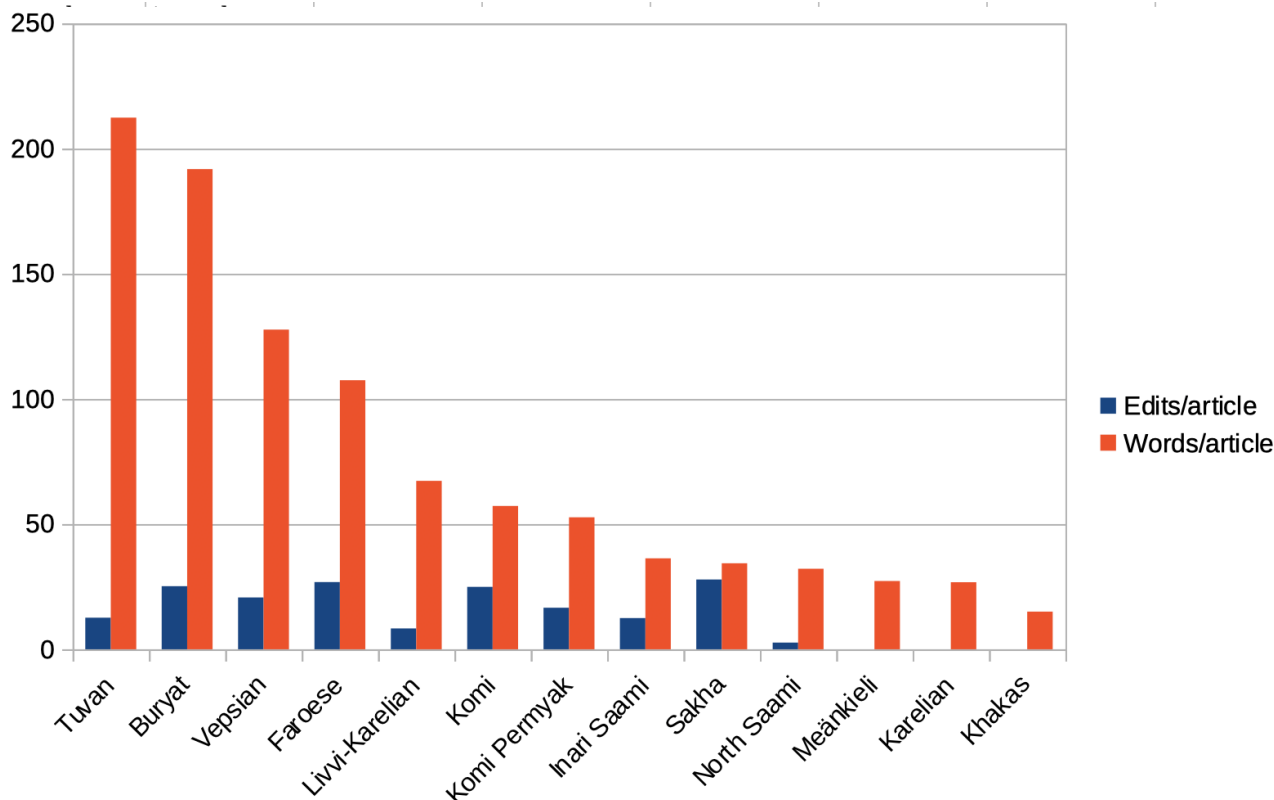


Figure 4: Edits and words per article for Wikipedia editions with more than 1000 articles

Number of edits

We then turn to the number of edits for each article. Taking the world's 321 Wikipedia editions¹⁰ as a reference point, their 56 million articles have been edited an average of 53 times each.

Data for the edits per article was available only for the Wikipedia editions not in the incubator. Figure 5 shows the circumpolar Wikipedia editions that are no longer in the incubator, ordered by the number of edits per article, but at the same time giving the number of articles.

As can be seen from the figure, there is no correlation between the number of articles and the number of edits per article. Instead, the figure reveals different types of Wikipedias. The Greenlandic, Inuktitut, and Inupiaq Wikipedias stand out with very few articles but a lot of edits per article. Faroese, Sakha, and to a certain extent also Komi, are larger Wikipedias with more than 20 edits per article. At the other extreme stands Livvi and North Saami, in particular, with very few edits per article.

⁹ The small Wikipedia editions are, in descending order: 100-999 articles: Greenlandic, Ingrian, Kven, Inuktitut, Chukchi, Inupiaq, Enets, Central Alaskan Yupik, Southern Saami, Kildin Saami, and Mansi. 10-99 articles: Skolt Saami, Ket, Tundra Nenets, Tlingit, Ludic, Lule Saami, Tundra Yukaghir, Nanai, Evenki, Siberian Yupik, and Gwich'in. 1-9 articles: Aleut, Khanty, Dolgan, Udege, Dene Suline, Oroch, Ulch, Negidal, Tlingit, and Hän.

¹⁰ For an overview of the Wikipedia editions, see <http://wikipedia.org>.

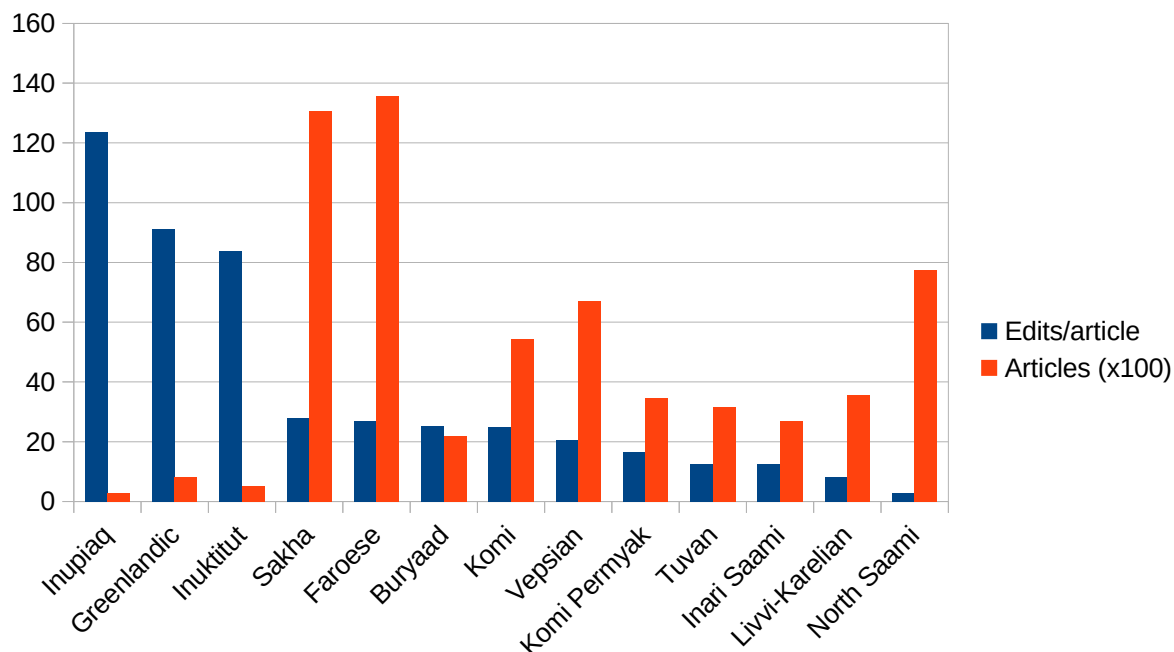


Figure 5: Edits/ article and number of articles for the non-incubator Wikipedia editions

Analysis

The analysis will look closer at the differences between the circumpolar Wikipedia editions and investigate why some editions are larger than others. It will also discuss what role these editions may have in language revitalisation work, not only as writing exercises, but also as encyclopedias.

Differences and their reasons

The strongest Wikipedia editions seem to fall into two categories. The first category consists of editions that have both a high articles per speaker ratio and to some extent, a high edits per article ratio. There are two Wikipedias in this category: *Inari Saami* and *Vepsian*. These are thus editions where the writers both write many articles and devote some time to each article. For Inari Saami, there are reason is clear: The Inari Saami revitalisation is conducted in a planned fashion by the Inari Saami Language Association, targeting one generation of speakers at a time. Their current priority is on strengthening literacy; one strategy is to use Wikipedia (cf. Olthuis et al 2021, Brecciaroli 2021). The writers behind the Vepsian Wikipedia may have been inspired by work on the Finnish and Estonian Wikipedia editions.

In the second category, are the Wikipedia editions with a relatively large population backed up by a strong native language administration. Within this category fall the *Faroese*, *Sakha* and *Komi* Wikipedias. The writers may be part of a successful revitalisation movement and have a strong motivation, as the Inari Saami writers do, or they may belong to a language community with a

strong literacy in school and society, spanning several generations, as in the case of Faroese, Sakha, and Komi.

The *North Saami* and *Livvi* Wikipedia editions fall outside this picture. Compared to the size of the language community, they contain many articles, but there is little activity on each article. Buryaad and Tuvan have a lot of speakers compared to the relatively few articles. The contrast to Sakha and Komi could also be coincidental.

Conclusion

The main conclusion is that very few of the circumpolar Wikipedia editions are active projects run by users able to write texts in the language in question. Only 3 Wikipedia editions saw edits from more than one editor actually knowing the language during a random month in 2021, and another 4 editions saw edits from only one editor knowing the language. Most Wikipedia editions are small, the median number of articles is 23, and only 13 out of 45 circumpolar Wikipedia editions have more than 1000 articles. The majority of Wikipedias are thus created and written by people who do not know the language and who do not belong to the language community.

The successful circumpolar Wikipedia editions are either rooted in active and successful revitalisation movements, as for Inari Saami, or they constitute a large speaker community with an official language policy favouring the circumpolar language as well as an active literacy. This is the case for Sakha, Faroese, and Komi.

The conclusions to be drawn for the Wikimedia chapters are that minority language communities are not like other language communities, and expecting them to behave like the majority language communities will lead us astray. A better understanding of the circumpolar language communities will improve our ability to offer Wikipedia to these communities, and in some rare cases, even contribute to strengthening their literacies.

References

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?* Poster at the Wikimania 2017 conference. <https://meta.wikimedia.org/wiki/File:WLR-Wikimania2017-poster-frhdkazan.pdf>

Frederiksen, Katti and 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.

Norwegian Ministry of Labour and Social Inclusion 2009: [Action Plan for Sami Languages](#). Oslo: Norwegian Government Administration Services.

Olthuis, Marja-Liisa, Suvi Kivelä and 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 and Eljas Niskanen 2021: [Methodological Implications of the Project 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 saamenkielisistä 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.

Appendix: Tables showing the circumpolar Wikipedia editions

Legend:

Fam = language family (dene = Na Dene, ea = Eskaleut, ie = Indo-European, mo = Mongolic, ps = Paleo-Siberian, tun = Tungusic, tur = Turkic, ur = Uralic).

Reg = region (a = North America, n = the Nordic countries, r = Russia)

Arts = Number of articles

Ed/art = Numbers of edits per articles

last30 = Active users (all users), last 30 days

eds l30 = Number of edits last 30 days

Spkrs = Language speakers (Source as given on their respective pages on <http://en.wikipedia.org>, mostly Ethnologue.

Art/Spkr = Wikipedia articles per speaker

Wrds = Total number of words in the Wikipedia edition, counted as content pages from the respective wikipedia dumps in April or May 2021, counted as text between blanks.

Wrds/art = Words per article

Orth = Year for current orthography, source mostly taken from <http://ru.wikipedia.org>

The circumpolar Wikipedia editions not in the incubator

Language	Fam	Reg	Address	Arts	Edits	Ed/art	last30	eds l.30	Spkrs	Art/spkr	Wrds	Wrds/art	Orth
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
Faroese	ie	n	fo.wikipedia.org	13581	364420	26,83	43	124	72000	0,19	1459615	107,47	1854
North Saami	ur	n	se.wikipedia.org	7737	20470	2,65	29	75	25000	0,31	248741	32,15	1979
Inari Saami	ur	n	smn.wikipedia.org	2670	33252	12,45	8	959	400	6,68	96942	36,31	1996
Greenlandic	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
Buryaad	mo	r	bxr.wikipedia.org	2181	54907	25,18	15	89	265000	0,01	418338	191,81	1939

The circumpolar Wikipedia editions in the incubator

Language	Fam	Cntr	Address	Arts	Spkrs	Art/spkr	Words	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
Hän	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
Southern Saami	ur	n	incubator.wikimedia.org/wiki/Wp/sma	160	600	0,27	1031	6,44	1974
Skolt Saami	ur	n	incubator.wikimedia.org/wiki/Wp/sms	90	320	0,28	1149	12,77	1973
Lule Saami	ur	n	incubator.wikimedia.org/wiki/Wp/smj	43	650	0,07	379	8,81	1983
Karelian	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
Kildin Saami	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
Ket	ps	r	incubator.wikimedia.org/wiki/Wp/ket	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
Negidal	ps	r	incubator.wikimedia.org/wiki/Wp/neg	2	75	0,03	117	58,50	1992

The circumpolar languages without any Wikipedia edition are:

Ahtna, Aluutiq, Coast Tsimshian, Deg Xinaq, Dena'ina, Forest Nenets, Gitxsan, Haida, Inuttitut, Inuvialuktun, Kaska, Kokyma Yukaghir, Koyukon, Kuskokwim, Lower Tanana, Naukan Yupik, Nganasan, Nisga'a, North Slavey, Northern Tutchone, Pite Saami, South Slavey, Southern Tutchone, Tł̨chq̨ (Dogrib), Tanacross, Ume Saami, and Upper Tanana.