

# Les éditions circumpolaires de Wikipédia

Trond Trosterud, Université Arctique de Norvège (Tromsø)

## Résumé

L'encyclopédie multilingue Wikipédia semble posséder une place particulièrement forte dans la région circumpolaire. Sur les quelque 70 langues circumpolaires – c'est-à-dire les langues autochtones de cette région –, presque toutes possèdent une langue écrite, et 45 d'entre elles ont leur propre Wikipédia ou bien sont présentes dans l'incubateur. Par rapport à la situation mondiale, tant la position de Wikipédia que le degré d'alphabétisation dans le Grand Nord sont remarquables : selon *ethnologue.com*, il existe 7117 langues vivantes, dont plus de la moitié (3982) possèderaient une version écrite, mais seulement 620 ont un Wikipédia ou bien sont présentes dans l'incubateur. Si nous examinons d'un peu plus près ces éditions circumpolaires de Wikipédia, nous nous apercevons que ces chiffres sont trompeurs. Cet article examinera de quelle manière ils le sont afin de mieux comprendre la présence écrite des langues autochtones, la situation des langues circumpolaires dans leurs pays respectifs et le mode de création des éditions mineures de Wikipédia.

## Introduction

Cet article analysera le contenu des 45 éditions de Wikipédia ou des éditions de l'incubateur rédigées dans une langue circumpolaire minoritaire. Par *langue circumpolaire*, nous entendons une langue minoritaire parlée par une population autochtone dans une région située au nord ou aux alentours du 60<sup>e</sup> parallèle nord, hors langues officielles.

L'article est structuré comme suit : la section 2 présente les communautés linguistiques circumpolaires, leur présence écrite et leurs éditions de Wikipédia. La section 3 analyse les différences entre les différentes éditions d'une même langue circumpolaire, les raisons de cette situation et le rôle que jouent ces éditions dans la revitalisation linguistique. La section 4 résume ce que nous savons maintenant sur ces éditions et les conditions qui ont conduit au succès de certaines d'entre elles, puis nous achèverons par quelques recommandations sur ce qui pourrait être amélioré à l'avenir.

## Contexte

### Les langues circumpolaires

Si nous divisons la région circumpolaire en trois régions, la région nord-américaine compte 30 langues circumpolaires, la région russe, 31, et la région nordique, 10. Les principales familles linguistiques sont les langues ouraliennes (23 langues circumpolaires dans les pays nordiques et la Russie), les langues na-dénées (19 langues circumpolaires en Amérique du Nord) et les langues eskaléoutes (10 langues à travers les 3 régions). Les 20 langues restantes sont soit des isolats, soit des langues turciques ou des langues tounghouses, à l'exception des langues féroïennes germaniques.

La plupart des langues circumpolaires sont mineures ; le nombre moyen de locuteurs pour les langues qui nous intéressent est de 650. Dans l'ordre décroissant, les 7 langues les plus importantes de l'échantillon (de 50 000 à 500 000 locuteurs) sont le yakoute, le touvain, le bouriate, le komi, le

féroïen, le komi-permiak et le groenlandais. Tous possèdent leur propre entité administrative dans laquelle la langue a un statut officiel, que ce soit une république autonome ou une région de Russie, ou encore ses propres îles dotées d'un gouvernement local, comme c'est le cas pour le groenlandais et le féroïen.

Trois de ces langues sont majoritaires dans leur zone autonome<sup>1</sup>: le féroïen à 94 %, le touvain à 91 %, le groenlandais à 85 % et l'inuktitut à 63 %. Les autres langues sont minoritaires (le yakoute à 38 %, le bouriate à 27 % et le komi à 18 %). Le reste des langues ont un statut plus faible ou bien ne possèdent aucun statut officiel.

Sur le plan typologique, toutes les langues circumpolaires ont une morphologie complexe. Toutes les langues na-dénées et eskaléoutes sont polysynthétiques et utilisent l'inflexion verbale pour exprimer l'information grammaticale et pragmatique exprimée dans les langues germaniques par les pronoms et les adverbes. Les langues na-dénées sont surtout préfixales, alors que les langues eskaléoutes sont suffixales. Les langues circumpolaires parlées en Russie et dans les pays nordiques sont en majorité suffixales. Pour la plupart de ces langues, les systèmes casuels sont importants, avec généralement entre 6 et 10 cas, mais le komi et l'oudmourte en présentent entre 20 et 30. Les langues minoritaires des pays nordiques ont moins de cas (moins de 10), mais présentent des systèmes plus complexes d'alternance phonétique dans les radicaux pendant l'inflexion. Dans l'ensemble, les langues de la minorité circumpolaire diffèrent nettement de leurs langues majoritaires respectives. Les distances linguistiques les plus importantes par rapport à la langue de la majorité sont observées auprès des langues circumpolaires d'Amérique du Nord, du groenlandais et des langues sames en Norvège et en Suède. Les locuteurs du russe et du finnois, en particulier, auront une distance linguistique légèrement plus réduite par rapport aux langues minoritaires de leurs pays respectifs.

## **Alphabétisation des langues circumpolaires**

L'alphabétisation dans le Nord peut être imputée à trois initiatives différentes. La première est due au travail de missionnaires dans le cadre de la consolidation des États, sur une période allant du 17<sup>au</sup> au 19<sup>ème</sup> siècle inclus. Cela a entraîné l'utilisation des caractères latins afin de créer les orthographes dans les pays nordiques, du cyrillique en Russie et en Alaska, et du syllabique et du latin en Amérique du Nord. La seconde initiative n'a eu lieu qu'en Russie, où la révolution d'Octobre a lancé une création planifiée à grande échelle de langues littéraires basées sur le script latin de toutes ses langues minoritaires, y compris celles du Nord, dans l'idée, à terme, de remplacer les précédentes orthographes basées sur le cyrillique. Enfin, s'intégrant dans un processus qui a débuté dans les années 1960 et 1970, des initiatives populaires et un changement ultérieur (ou du moins un relâchement du processus d'assimilation) ont entraîné la création de nouvelles orthographes pour certaines des langues circumpolaires ainsi que, pour d'autres, la révision des anciennes orthographes.

En observant de plus près les différentes parties des zones circumpolaires, nous voyons qu'en Amérique du Nord, les premières initiatives missionnaires à partir du 18<sup>e</sup> siècle basées sur la langue maternelle ont été remplacées par une politique d'assimilation qui a augmenté avec l'intégration des populations du nord dans la société moderne au 20<sup>e</sup> siècle. En 1999, dans le cadre de la réconciliation, le Canada a fait du Nunavut – un territoire du Nord-Ouest – sa propre région. En 2008, il a octroyé à l'inuktitut le statut de langue officielle et a commencé à publier les actes de

<sup>1</sup> Canada : [Statistiques canadiennes, 2016](#). Îles Féroé : [Statbank hagstova, 2011](#). Groenland : FREDERIKSEN et OLSEN, 2017. Russie : [Всероссийская перепись населения, 2010](#).

L'Assemblée législative du Nunavut en anglais et en inuktitut. La plupart des langues circumpolaires au Canada sont enseignées dans leurs réserves respectives des Premières Nations, mais principalement en tant que langues étrangères, l'anglais étant la langue d'enseignement, et avec des résultats médiocres (voir MEEK, 2010, pour une analyse approfondie de cette pratique). En Alaska, la politique d'assimilation a encore moins changé qu'au Canada. Dans les années 1960, l'Université de Fairbanks a commencé à créer des orthographes pratiques pour plusieurs langues amérindiennes d'Alaska, mais cela n'a pas donné lieu à des changements systématiques dans les programmes scolaires. La reconnaissance de 20 langues autochtones de l'Alaska en tant que « langues co-officielles avec l'anglais » en 2014 n'a pas non plus donné lieu à un changement radical de politique<sup>2</sup>.

En Russie, la création d'orthographes dans les années 1930 s'est accompagnée de changements administratifs où les frontières administratives sans distinctions ethniques ont été remplacées par un système de républiques autonomes et, pour les minorités plus petites, de zones créées autour de toutes les minorités suffisamment grandes pour constituer un organe administratif. Durant les premières décennies qui ont suivi la Seconde Guerre mondiale, les langues avec leur propre république autonome (le komi, le yakoute, le touvain, le bouriate, le khakasse) ont été utilisées comme langue d'enseignement dans les écoles primaires. Les républiques ont publié des revues et des journaux dans leur propre langue, et des romans ont également été écrits et publiés dans la langue des républiques. Les locuteurs des autres langues circumpolaires de la Russie ont au moins appris à lire et à écrire leur langue maternelle, et il existait aussi des publications utilisant ces langues. Avec les nouvelles lois scolaires introduites en 1958-59, l'alphabétisation des autochtones a été affaiblie, mais elle est restée beaucoup plus forte que celle des langues minoritaires dans les autres régions circumpolaires<sup>3</sup>.

Après avoir été gouvernés en danois pendant des siècles, les Féroïens ont commencé à travailler sur leur propre langue écrite à la fin du 19<sup>e</sup> siècle. L'occupation britannique pendant la Seconde Guerre mondiale a affaibli l'influence danoise, et après l'introduction de l'autonomie gouvernementale en 1948, le féroïen est progressivement aussi devenu la langue écrite dominante (cf. KNUDSEN, 2010). Au Groenland, le groenlandais a été utilisé dans tout le système scolaire, y compris la formation des enseignants, et même dans un journal. Mais en 1953, cette tradition d'alphabétisation autochtone vieille de 100 ans a été mise au panier quand le Groenland est passé d'une colonie danoise à un comté (« *amt* », en danois) intégré au Danemark. La transition a été combinée à un changement à grande échelle de la langue écrite, passant du groenlandais au danois. Même si l'introduction de l'autonomie gouvernementale en 1979 et de l'autogouvernance en 2008 a fait du groenlandais la seule langue officielle du Groenland, le danois a encore une position forte, surtout en tant que langue dominante de l'ensemble de l'enseignement théorique après le cursus obligatoire. La langue écrite de l'administration, du système juridique et des médias de masse est bilingue, avec des textes typiquement rédigés en danois puis traduits en groenlandais (FREDERIKSEN et OLSEN, *op.cit.*).

2 Cf. <http://www.akleg.gov/basis/Bill/Text/28?Hsid=HB0216Z>, le nouveau paragraphe intitulé « Ajout de l'inupiaq, du youpik sibérien, du youpik de l'Alaska central, de l'alutiiq, de l'unangax, du dena'ina, de l'ingalik, du holikachuk, du koyukon, du kolchan, du kutchin, du tanana, du haut tanana, du tanacross, du hän, de l'ahtna, de l'eyak, du tlingit, du haida et du tsimshian comme langues officielles de l'État ». La loi stipule toujours explicitement que cela n'exige pas qu'un gouvernement d'État ou de municipalité imprime un document dans une autre langue que l'anglais.

3 Une présentation de la politique soviétique par un de ses architectes se retrouve dans ISAEV, 1977. Pour différentes analyses, voir par exemple KREINDLER, 1985, et TROSTERUD, 1995. BILINSKY, 1962, fournit une analyse de l'effet des lois scolaires de 1958-59.

Au cours des quatre dernières décennies, les pays nordiques ont lentement mis en œuvre une nouvelle politique linguistique pour leurs langues minoritaires, ce qui a donné lieu à un statut officiel au niveau régional ou à un statut de langue minoritaire nationale pour les langues sames, le kvèné et le meänkieli. Une grande partie du programme éducatif dans les écoles sames du Nord est enseignée en same du Nord et, dans une moindre mesure, en skolt et same d’Inari. Toutes les langues minoritaires circumpolaires des pays nordiques sont enseignées en tant que matières scolaires<sup>4</sup>. Il convient de noter le cas exceptionnel du same d’Inari, où un changement linguistique continu au cours de la dernière génération a donné lieu à une situation où la majorité des locuteurs appartiennent à la nouvelle génération, et où la communauté linguistique publie à la fois des revues, des manuels scolaires, de la fiction et des bulletins d’informations en same d’Inari<sup>5</sup>.

Ainsi, pour l’ensemble des langues circumpolaires, cela implique que seulement un quart des orthographes actuelles remontent à la Seconde Guerre mondiale. Ces orthographes sont soit basées sur le cyrillique (depuis 1937) ou bien sont des syllabaires. La moitié des orthographes circumpolaires ont moins de 40 ans. La plupart des langues circumpolaires tirent la plupart de leurs locuteurs des groupes d’âge supérieur, la situation prototypique étant que les langues circumpolaires ont des orthographes plus jeunes que leurs propres locuteurs.

En ce qui concerne l’utilisation de la langue sur les réseaux sociaux, tant le féroïen, le groenlandais, le meänkieli, le same du Nord, le same d’Inari et dans une moindre mesure le skolt, le same du Sud, le same de Lule, le carélien, le livvi et le vepse sont en usage sur Facebook, à la fois dans les groupes consacrés à ces langues respectives, et dans des posts personnels. En Amérique du Nord, la situation semble différente, car les groupes Facebook consacrés aux langues les plus importantes (l’inuktitut, le yupik d’Alaska central, le déné suline, l’inupiaq) n’utilisaient que l’anglais. Les mots de langue maternelle étaient généralement cités (pour expliquer ce qu’ils signifiaient), mais les discussions se faisaient toujours en anglais. Le bouriate, le komi, le touvain et le khakasse sont en usage sur vkontakte<sup>6</sup> dans des groupes consacrés à l’utilisation de la langue en question. Le yakoute un peu moins. Les autres langues circumpolaires de Russie avec leur propre édition Wikipédia possèdent apparemment des groupes linguistiques sur vkontakte en russe, ou bien n’ont pas de groupe du tout.

## **Les éditions circumpolaires de Wikipédia**

Passons à présent aux éditions concrètes de Wikipédia. Après une brève vue d’ensemble, nous examinerons la taille des éditions et de leur communauté linguistique, puis les auteurs, et enfin l’état des articles.

### **Les éditions de Wikipédia**

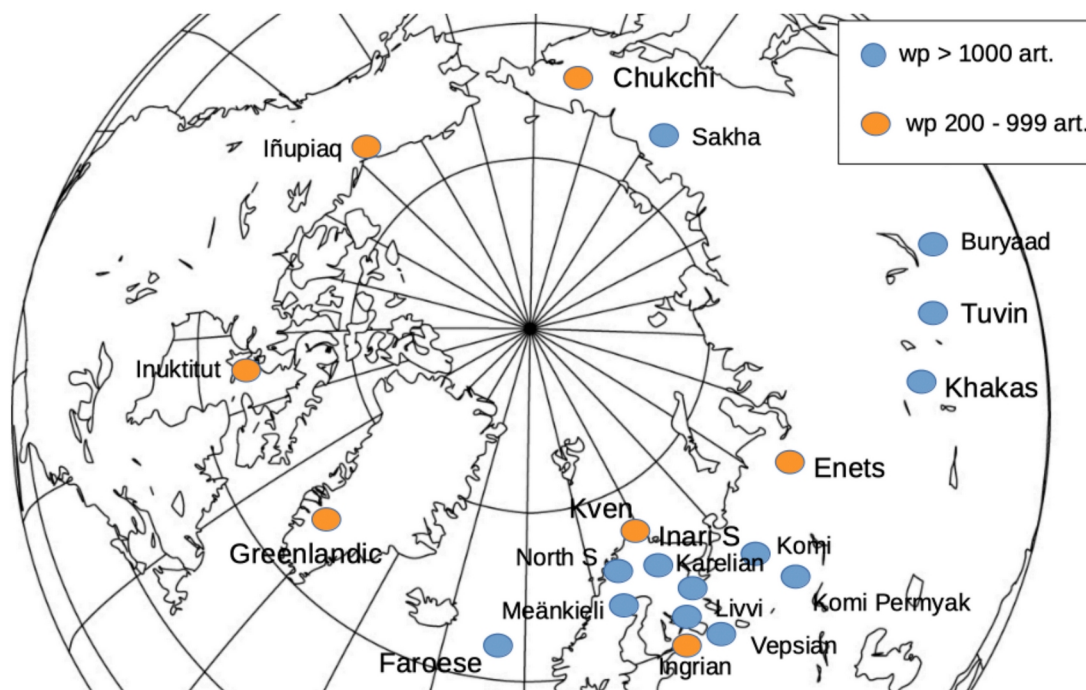
La carte du graphique 1 désigne l’emplacement approximatif des éditions de Wikipédia dotées de plus de 200 articles par des ovales orange, et celles avec plus de 1000 articles par des ovales bleus. Comme le montre le graphique 1, les éditions de Wikipédia contenant plus de 1000 mots se retrouvent surtout dans les pays nordiques, mais également en Russie avec les langues de la République. En plus des langues indiquées sur la carte, il existe 24 éditions de Wikipédia contenant

4 Pour observer les effets de la politique en Norvège, voir le « Plan d’action pour les langues sames » (Ministère norvégien du travail et de l’inclusion sociale, 2009). Une étude de la situation des langues sames en Finlande est retranscrite dans PASANEN, 2016.

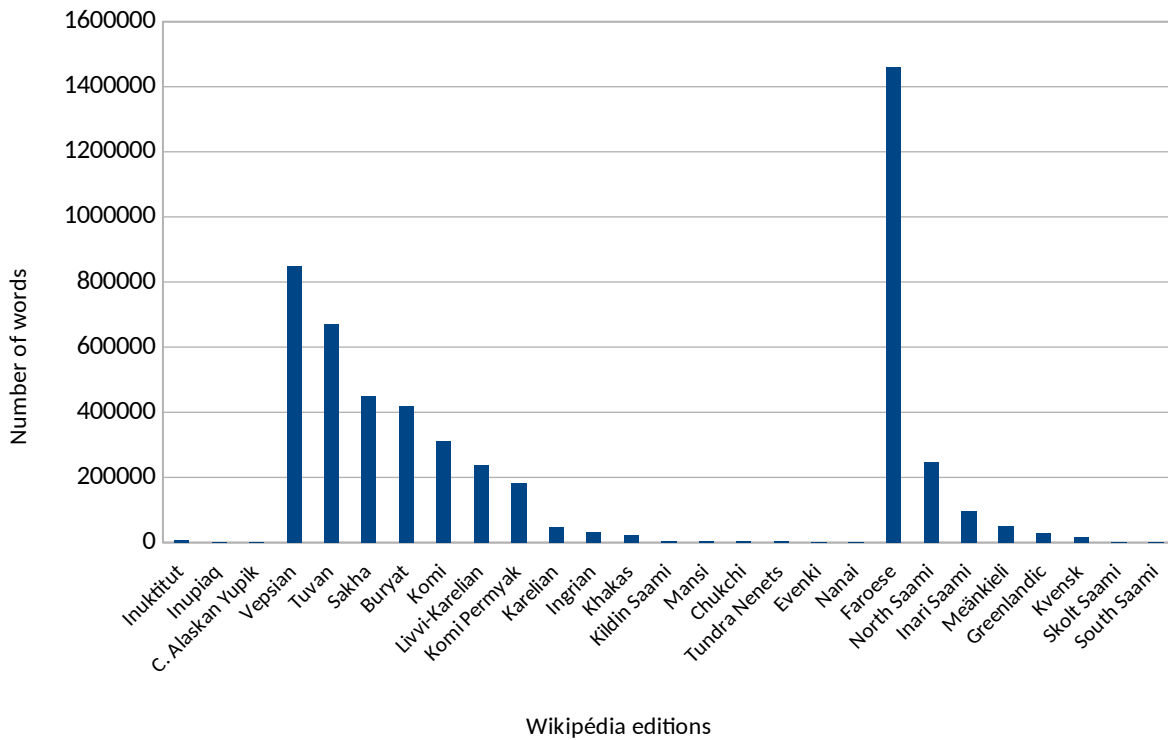
5 PASANEN, 2015, est la référence principale pour ce processus de revitalisation. Une analyse approfondie de la restauration de la « génération intermédiaire » (celle entre les personnes âgées et les enfants des niches linguistiques) est OLTHUIS et al. 2013.

6 vkontakte est la version russe de Facebook, cf. <https://vk.com/>

moins de 200 mots. De la plus grande à la plus petite, 6 d'entre elles se trouvent en Amérique du Nord (yupik de l'Alaska du centre, tlingit, hän, gwich'in, aléoute et déné suline), 15 en Russie (same de Kildin, mansi, ket, nénètse de la Toundra, ludic, youkaguir de la Toundra, nanaï, evenki, yupik sibérien, khanty, dolgane, oudihé, oultche, orotche et négidal), et dans les pays nordiques (same du Sud, skolt et same de Lule). La caractéristique des langues de ce groupe est qu'elles ont peu de locuteurs et peu ou pas de traditions littéraires. Leurs Wikipédias sont généralement lancés par des non-locuteurs externes à la communauté linguistique. Certains articles sont dans la langue majoritaire, et il n'y a actuellement aucune activité sur aucun d'entre eux.

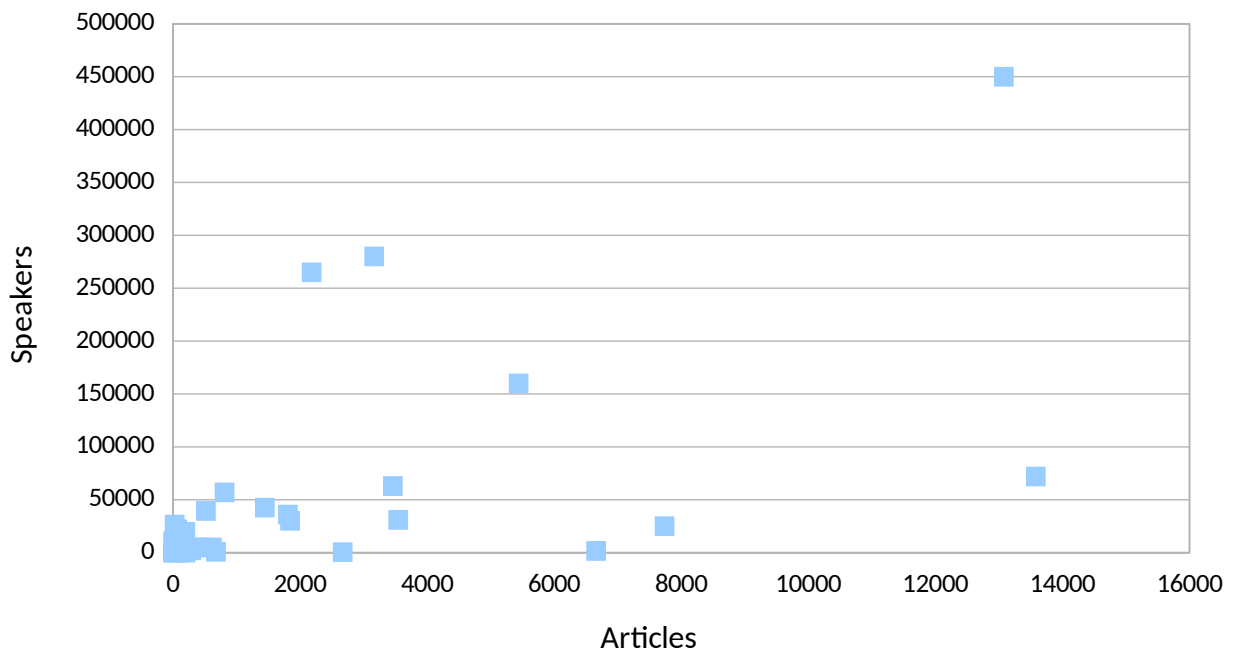


Faisons d'abord le point sur la taille des éditions de Wikipédia contenant plus de 1000 mots. Elles sont présentées dans le graphique 2, classées d'abord selon la région (Amérique du Nord, Russie, pays nordiques), puis selon le nombre de mots. Force est de constater que les trois zones affichent des profils distincts. Pour l'Amérique du Nord (à gauche), les trois plus grands Wikipédias contiennent à peine 100 mots. Les éditions Wikipédia de la Russie peuvent (à quelques exceptions près, sur lesquelles nous reviendrons) être divisées en deux groupes : ceux qui ont le soutien de leur propre république, et les autres. Dans les pays nordiques, le féroïen est le cas particulier. Le same du Nord est le plus important après le féroïen, suivi de près par le same d'Inari, le reste étant loin derrière.



Graphi

que 2 : Les éditions de Wikipédia de plus de 1000 mots en Amérique du Nord, en Russie et dans les pays nordiques



### Taille des Wikipédias par rapport à la taille de la communauté linguistique

D'entrée de jeu, on pourrait penser que plus de locuteurs équivaldraient à plus d'articles écrits. Comme le montre le graphique 3, il existe effectivement une telle corrélation, mais elle est faible ( $R^2=0.37$ , où  $R^2=1.0$  aurait été une corrélation parfaite). Les cas particuliers avec « trop peu d'articles » concernent le yakoute, le komi, le touvain et le bouriate, alors que celles avec « trop

d'articles » sont le féroïen, le same du Nord, le vepse, et le same d'Inari. Les aberrations avec trop peu d'articles sont également les quatre langues de Russie qui possèdent les plus grandes communautés linguistiques. Cela étant, comparées aux éditions plus importantes de Wikipédia hors de la zone circumpolaire, elles ne sont pas si hors-normes que cela. Avec 88 locuteurs par article, le bouriate est sur un pied d'égalité avec l'albanais (72), et les 29 et 34 locuteurs par article pour le komi et yakoute sont comparables avec le danois (21), le hongrois (30) et le bulgare (33). Le féroïen est sur un pied d'égalité avec l'islandais (les deux ont environ 5,5 locuteurs par article).

Parmi les langues circumpolaires comptant plus de 10 000 locuteurs, 5 se distinguent par « trop d'articles », toutes dans les pays nordiques ou en Carélie : le féroïen, le same du Nord, le vepse, le livvi (olonetsien) et le same d'Inari. Pour les plus petites communautés linguistiques du vepse et du same d'Inari, il n'est pas rare d'avoir un petit nombre de locuteurs par article. Cela étant, ces communautés linguistiques se distinguent par une activité plus élevée que les autres, surtout par rapport à leur taille.

La corrélation entre la taille de la communauté linguistique et celle de Wikipédia est donc faible, et il doit exister d'autres facteurs que la taille de la communauté linguistique pour déterminer la taille des Wikipédias.

### **Les auteurs**

Lorsque les communautés linguistiques sont aussi petites qu'elles le sont pour les langues circumpolaires, le comportement des rédacteurs individuels joue un rôle plus important que celui des communautés linguistiques plus grandes.

Pour chaque édition de Wikipédia, il existe une page située à l'emplacement *Pages spéciales > Liste des utilisateurs actifs*, qui indique qui a été actif au cours des 30 derniers jours. Le terme « utilisateurs » sur Wikipédia fait référence aux personnes ou aux programmes informatiques (bots) qui éditent les éditions de Wikipédia. Certains des utilisateurs énumérés connaissent la langue du Wikipédia en question, mais une grande partie ou même la plupart des utilisateurs ne la maîtrisent pas. Ces non-locuteurs peuvent être divisés en trois catégories : les utilisateurs qui ne touchent pas au contenu linguistique, les utilisateurs ayant suffisamment de connaissance de la langue et de leurs propres limites en tant que rédacteur pour produire un texte acceptable, et les utilisateurs qui ont peu ou pas de connaissance de la langue, mais qui produisent tout de même du texte, ce qui varie généralement entre une qualité très médiocre et un charabia complet.

Afin de répartir les utilisateurs en locuteurs et non-locuteurs, j'ai examiné à la fois la maîtrise de la langue qu'ils affirment posséder et leur comportement en tant que rédacteur. De nombreux utilisateurs annoncent leur maîtrise de la langue sur leur profil d'utilisateur. J'ai fait confiance à ceux qui ne déclaraient aucune compétence et j'ai examiné de plus près ceux qui déclaraient une certaine compétence linguistique ou ne révélaient pas leur maîtrise exacte. Pour ces utilisateurs, j'ai parcouru leurs contributions et classé celles qui ne requéraient pas de compétences linguistiques en tant que non-locuteurs<sup>7</sup>.

En observant en premier lieu le groupe d'utilisateurs qui connaissent la langue en question, l'image suivante est apparue : seulement 3 éditions de Wikipédia avaient plus d'un utilisateur actif ayant une connaissance de la langue au cours du mois dernier : le *yakoute* (15 utilisateurs), le *féroïen* (9) et le *same d'Inari* (8). Les éditions suivantes de Wikipédia ont eu un (1) utilisateur ayant une connaissance de la langue au cours de la même période : *Le same du Nord*, *le vepse*, *le komi* et le

<sup>7</sup> L'évaluation a été effectuée le 6 juin 2021.

*bouriate*. Les autres éditions avec plus de 1000 mots (le *livvi*, le *komi-permiak*, le *touvain*, le *groenlandais*, l'*inuktitut* et l'*inupiaq*) n'ont pas été éditées par des utilisateurs ayant une connaissance de la langue au cours des 30 derniers jours.

Si l'on regarde ensuite l'ensemble des utilisateurs actifs, on se rend compte que 1/3 des contributeurs (44 sur 146) sont actifs sur plus d'une des éditions circumpolaires de Wikipédia. Ce sont ces mêmes contributeurs qui ne connaissent pas la langue en question. Tant que ces utilisateurs connaissent leurs limites, leurs contributions sont certainement bienvenues. Un exemple est l'utilisateur *Glorious 93*, un utilisateur français qui, au cours du dernier mois, a été actif sur 8 des 13 éditions circumpolaires de Wikipédia, ainsi que sur plusieurs éditions de l'incubateur. Ses contributions impliquent généralement l'ajout ou la mise à jour de références d'images situées dans des boîtes d'information, un travail très utile sans conséquences négatives. Un représentant du deuxième type d'utilisateur (plus) expert est le compte *Frhdkazan*, qui est actif sur de nombreuses éditions de Wikipédia dans les langues de la Russie. Selon son profil utilisateur, sa langue maternelle est le tatar, qui est une langue turque. Les éditions linguistiques auxquelles il participe ont une structure grammaticale semblable à celle du tatar. Il est ainsi en mesure de comprendre la structure des phrases des articles qu'il édite ou crée, et s'il reste un éditeur prudent, les chances d'obtenir un texte d'article de qualité acceptable sont tout à fait probables.

Le troisième type de non-locuteurs écrivant sur Wikipédia est ce que j'appelle les *pirates de Wikipédia*. Ces Wikipédiens choisissent généralement un Wikipédia avec peu ou pas d'activité, et leurs contributions dominent rapidement l'édition Wikipédia. Un exemple est l'édition du same du Sud<sup>8</sup>. Les deux titres les plus capitaux (« Page d'accueil » et « Bienvenue sur Wikipédia, l'encyclopédie libre ») contiennent tous deux des erreurs catastrophiques. Pour le mot « page », l'utilisateur a utilisé le mot *bielie* (« côté ») et pour « libre », il a utilisé *sångedidh* (qui est l'acte de demander en mariage). Derrière ces erreurs se cachent des homonymies en norvégien reflétées dans le dictionnaire norvégien — same du Sud. Le mot norvégien *side* signifie à la fois « côté » et « page », alors que le mot *fri* signifie à la fois « libre » et « demander en mariage ». Il y a également des erreurs grammaticales : dans l'expression *sångedidh daajroegærjase*, qui tente de dire « [Bienvenue] dans l'encyclopédie libre », la forme infléchie *gærjase* (< *gærja*, « livre ») est erronée ; la version correcte est *gærjese*. Ces phrases ne proviennent même pas des pages de contenu ; elles font partie de la localisation du Wikipédia en same du Sud. Un « travail » de ce type est peut-être effectué en toute bonne foi, par des personnes qui ont un manque total de lucidité sur leurs propres capacités linguistiques, mais il a de graves conséquences pour la communauté linguistique en question. Même si la communauté linguistique n'est absolument pas consciente de l'existence d'une édition de Wikipédia dans sa propre langue, ces formes erronées apparaissent dans les recherches Google, rendant ainsi impossible l'utilisation d'Internet pour vérifier le bon usage de la langue.

### **Aide organisée pour les utilisateurs : Les chapitres de Wikimédia**

En Russie, en Norvège et en Finlande, les divisions nationales de Wikimédia apportent leur soutien aux éditions en langue minoritaire des Wikipédias des communautés linguistiques de leurs pays respectifs. Des séminaires wiki ont été organisés pour plusieurs des langues de Russie. Une présentation de ces initiatives se trouve dans FATKULLIN, 2017. De même, les divisions de

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

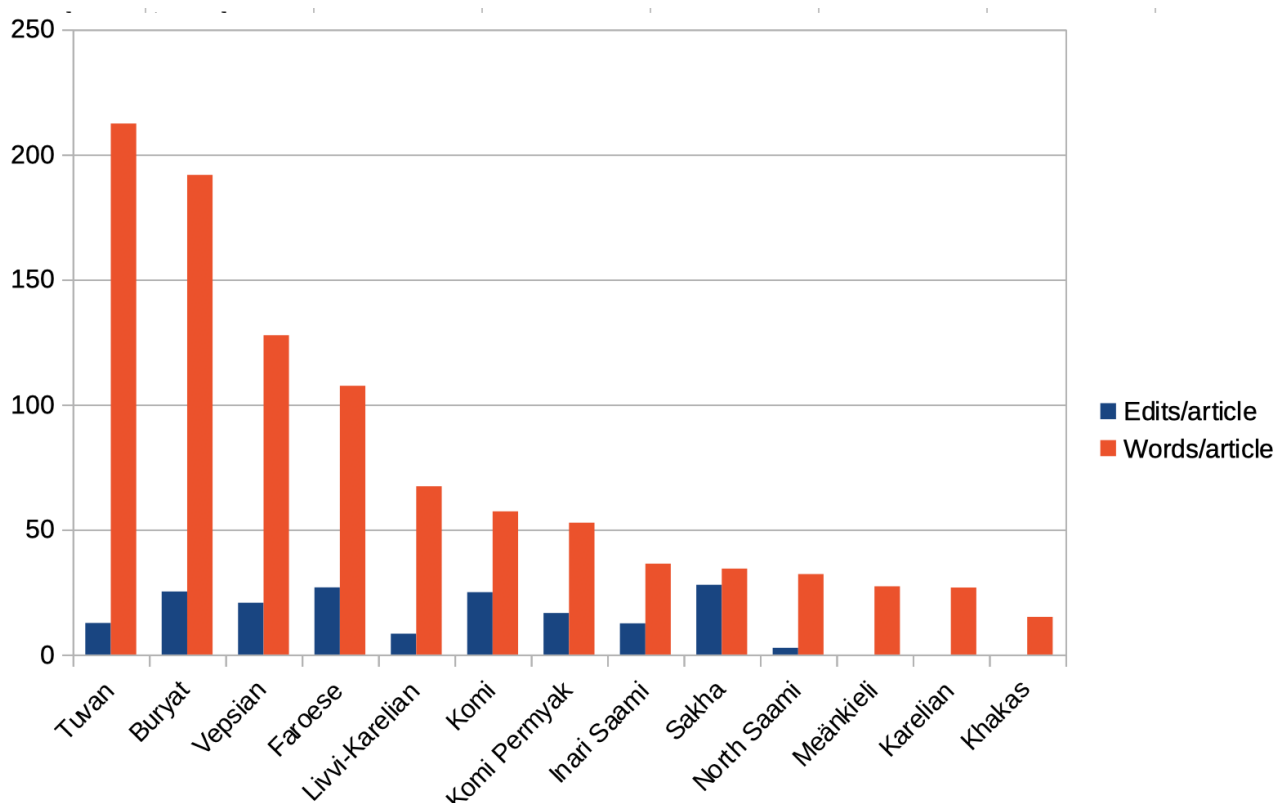


Wikimédia de Norvège et de Finlande ont organisé des cours pour les rédacteurs potentiels et ont apporté leur aide pour des templates, des boîtes d'infos et du formatage.

### Longueur des articles

Si nous regardons les 12 éditions avec plus de 1000 articles (dans ce contexte appelé « éditions majeures de Wikipédia »), seulement une a des articles plus courts : celle pour le khakasse, avec 15 mots par article, voir graphique 4. Le touvain et le bouriate présentent les articles les plus longs, avec une longueur moyenne de 212 et 192 mots, respectivement.

32 des Wikipédias circumpolaires ont moins de 1000 articles<sup>9</sup> ; ceux-ci seront désignés sous l'appellation des « éditions mineures de Wikipédia ». En plus d'avoir un nombre d'articles restreint, les articles eux-mêmes étaient plus courts. Si l'on regarde chaque édition de Wikipédia séparément, le nombre moyen de mots par articles pour les plus grandes était de 52,7 mots par article, alors que pour les plus petites, il était de 11,5.



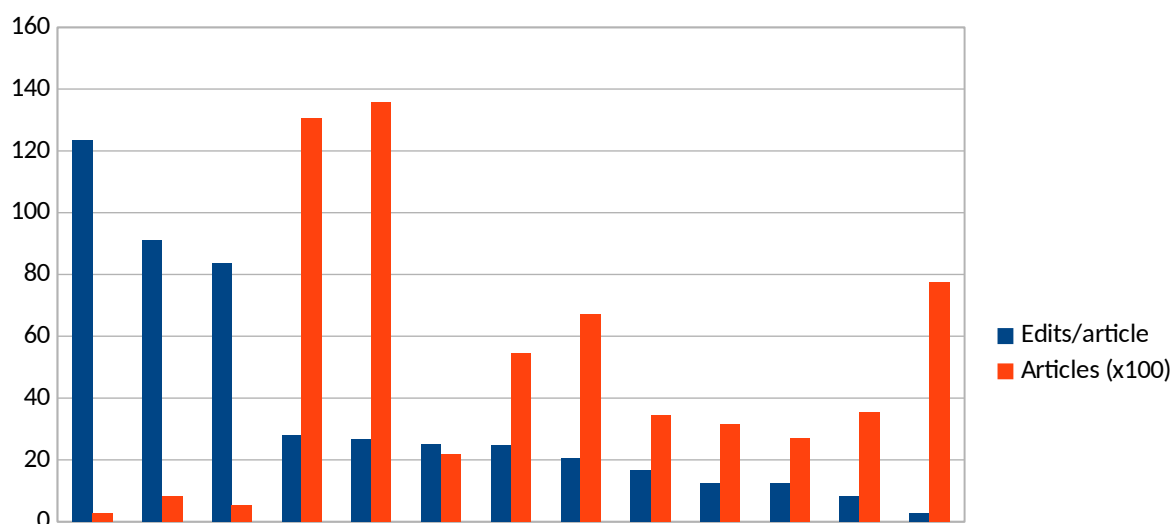
Passons maintenant au nombre de modifications pour chaque article. Si l'on prend les 321 éditions de Wikipédia dans le monde<sup>10</sup> comme point de référence, leurs 56 millions d'articles ont été édités en moyenne 53 fois chacun.

<sup>9</sup> Voici la liste des éditions mineures de Wikipédia dans l'ordre décroissant. De 100 à 999 articles : le groenlandais, l'ingrien, le kvène, l'inuktitut, le chukchi, l'inupiaq, l'énètse, le yupik de l'Alaska central, le same du Sud, le same de Kildin et le mansi. De 10 à 99 articles : le skolt, le ket, le nénétsse de la Toundra, le tlingit, le ludic, le same de Lule, le youkaguir de la Toundra, le nanaï, l'évenki, le yupik de Sibérie et le gwich'in. De 1 à 9 articles : l'aléoute, le khanty, le dolgane, l'oudihé, le déné suline, l'orotche, l'oultche, le négidal, le tlingit et le hän.

<sup>10</sup> Pour obtenir un aperçu des éditions de Wikipédia, voir <http://wikipedia.org>.

Les données pour les modifications par article n'étaient disponibles que pour les éditions de Wikipédia qui ne se trouvent pas dans l'incubateur. Le graphique 5 montre les éditions circumpolaires de Wikipédia qui ne sont plus dans l'incubateur, classées par le nombre de modifications par article, tout en précisant le nombre d'articles.

Comme on peut le voir sur le graphique, il n'existe aucune corrélation entre le nombre d'articles et le nombre de modifications par article. Au lieu de cela, le graphique révèle différents types de Wikipédias. Les Wikipédias pour le groenlandais, l'inuktitut et l'inupiaq se distinguent par très peu d'articles, mais beaucoup de modifications par article. Le féroïen, le yakoute et, dans une certaine mesure, le komi, sont des Wikipédias plus grands avec plus de 20 modifications par article. À l'autre extrême, nous avons le livvi et le same du Nord – en particulier – avec très peu de modifications par article.



## Analyse

L'analyse se penchera sur les différences entre les éditions circumpolaires de Wikipédia et étudiera pourquoi certaines éditions sont plus importantes que d'autres. Nous discuterons également du rôle que ces éditions peuvent jouer dans le travail de revitalisation des langues, non seulement comme exercices d'écriture, mais aussi en tant qu'encyclopédies.

## Les différences et leurs causes

Les éditions les plus fortes de Wikipédia semblent appartenir à l'une de deux catégories. La première regroupe des éditions qui ont à la fois un rapport élevé d'articles par locuteur et, dans une certaine mesure, un rapport élevé de modifications par article. Deux Wikipédias sont présents dans cette catégorie : le *same d'Inari* et le *vepse*. Ce sont donc des éditions où les auteurs écrivent beaucoup d'articles et consacrent un certain temps à chaque article. Pour le *same d'Inari*, il existe une raison précise : la revitalisation de cette langue est menée de manière planifiée par l'Association de la langue *same d'Inari*, ciblant une génération de locuteurs à la fois. Leur priorité actuelle est de renforcer l'alphabétisation et une de leurs stratégies est d'utiliser Wikipédia (cf. OLTHUIS et al., 2021, BRECCIAROLI, 2021). Les rédacteurs du Wikipédia en langue *vepse* ont peut-être été inspirés par les travaux sur les éditions en finnois et en estonien.

Dans la deuxième catégorie se trouvent les éditions de Wikipédia qui représentent une population relativement importante soutenue par une solide administration en langue maternelle. Dans cette catégorie se trouvent les Wikipédias pour le *féroïen*, le *yakoute* et le *komi*. Les rédacteurs font peut-être partie d'un mouvement de revitalisation réussi et ont une forte motivation, comme ceux du same d'Inari, ou alors, ils appartiennent à une communauté linguistique dotée d'une solide alphabétisation à l'école et dans la société, et ce sur plusieurs générations, comme dans le cas du féroïen, du yakoute et du komi.

Les éditions de Wikipédia pour le same du Nord et le livvi sortent de ce cadre. Par rapport à la taille de leur communauté linguistique, ils contiennent de nombreux articles, mais il y a peu d'activité sur chaque article. Le bouriate et le touvain ont de nombreux locuteurs par rapport aux articles relativement peu nombreux. Le contraste avec le yakoute et le komi pourrait également être une coïncidence.

## Conclusion

La conclusion principale est que très peu d'éditions circumpolaires de Wikipédia sont des projets actifs gérés par des utilisateurs capables d'écrire des textes dans la langue en question. Durant un mois de 2021 sélectionné au hasard, seulement 3 éditions de Wikipédia ont présenté des modifications de la part de plus d'un éditeur connaissant réellement la langue, et 4 autres éditions ont été modifiées par un rédacteur connaissant la langue. La plupart des éditions de Wikipédia sont petites. Le nombre moyen d'articles est de 23, et seulement 13 des 45 éditions circumpolaires de Wikipédia ont plus de 1000 articles. La majorité des Wikipédias sont ainsi créés et rédigés par des personnes qui ne connaissent pas la langue et qui n'appartiennent pas à la communauté linguistique.

Les éditions circumpolaires réussies de Wikipédia sont soit enracinées dans des mouvements de revitalisation actifs et réussis, comme pour le same d'Inari, soit elles constituent une grande communauté de locuteurs avec une politique de langue officielle favorisant la langue circumpolaire ainsi qu'une alphabétisation active. C'est le cas du yakoute, du féroïen et du komi.

Les conclusions à tirer pour les chapitres de Wikimédia sont que les communautés de langue minoritaire ne sont pas comme les autres communautés de langue, et s'attendre à ce qu'elles se comportent comme les communautés de langue majoritaire nous induira en erreur. Une meilleure compréhension des communautés linguistiques circumpolaires améliorera notre capacité à offrir Wikipédia à ces communautés, et en de rares occasions, contribuera même à renforcer leurs compétences linguistiques.

## Références

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 Wikipédia*. Wikimédia : Arctic Knot

FATKULLIN, Farhad, 2017. *Wikipédias in the languages of Russia: in line with the wider Movement strategy?* Affiche de la conférence Wikimania, 2017. <https://meta.wikimedia.org/wiki/File:WLR-Wikimania2017-poster-frhdkazan.pdf>

- FREDERIKSEN, Katti et 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*. Original en russe : *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 (éd.), 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.
- Ministère norvégien du Travail et de l'inclusion sociale (2009) : [Plan d'action pour les langues sames](#). Oslo. Services administratifs du gouvernement norvégien.
- OLTHUIS, Marja-Liisa, Suvi KIVELÄ et 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 et Eljas NISKANEN, 2021. [Methodological Implications of the Project Čyeti čälled anarâškielân, 'One Hundred Writers for Aanaar Saami'](#) dans *Indigenous Research Methodologies in Sámi and Global Contexts. New Research - New Voices*, Volume : 11
- PASANEN, Annika, 2015. *Kuávsui já peeivičuovâ. 'Sarastus ja päivänvalo' : Inarinsaamen kielen redynamalizaatio*. Thèse de doctorat, 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.

## Annexe : Tableaux montrant les éditions circumpolaires de Wikipédia

Légende :

Fam = famille linguistique (déné = na-déné, es = eskaléoute, ie = indo-européen, mo = mongol, ps = paléo-sibérien, tou = tOUNGouse, tur = turcique, ou = ouralien).

Reg = région (a = Amérique du Nord, n = pays nordiques, r = Russie)

Arts = nombre d'articles

Mo/ar = nombre de modifications par article

der30 = utilisateurs actifs (tous les utilisateurs), 30 derniers jours

mod30 = nombre de modifications des 30 derniers jours

Loc = locuteurs (source comme indiquée sur leurs pages respectives sur <http://en.Wikipédia.org>, principalement Ethnologue.

Art/Loc = articles Wikipédia par locuteur

Mots = nombre total de mots dans l'édition Wikipédia, comptés comme pages de contenu provenant de leurs dumps Wikipédia respectifs en avril ou mai 2021, comptés comme texte entre les espaces.

Mots/art = mots par article

Orth = année pour l'orthographe actuelle, source principalement tirée de <http://ru.Wikipédia.org>

### Les éditions circumpolaires de Wikipédia non présentes dans l'incubateur

Langue	Fam	Reg	URL	Arts	Mods	Mo/ar	der30	Mod30	Loc	Art/loc	Mots	Mots/art	Orth
Inuktitut	es	a	<a href="http://iu.Wikipédia.org">iu.Wikipédia.org</a>	517	43 322	83,79	13	36	39 475	0,01	6898	13,34	1855
Inupiaq	es	a	<a href="http://ik.Wikipédia.org">ik.Wikipédia.org</a>	289	35 723	123,61	7	11	2144	0,13	1803	6,24	1946
				13	364						1 459		
Féroïen	ie	n	<a href="http://fo.Wikipédia.org">fo.Wikipédia.org</a>	581	420	26,83	43	124	72 000	0,19	615	107,47	1854
Same du Nord	ou	n	<a href="http://se.Wikipédia.org">se.Wikipédia.org</a> <a href="http://smn.Wikipédia.org">smn.Wikipédia.org</a>	7737	20 470	2,65	29	75	25 000	0,31	248 741	32,15	1979
Same d'Inari	ou	n	<a href="http://g">g</a>	2670	33 252	12,45	8	959	400	6,68	96 942	36,31	1996
Groenlandais	es	n	<a href="http://kl.Wikipédia.org">kl.Wikipédia.org</a>	810	73 832	91,15	15	13	57 000	0,01	29 683	36,65	1973
				13	364				450				
Yakoute	tur	r	<a href="http://sah.Wikipédia.org">sah.Wikipédia.org</a>	077	499	27,87	39	438	000	0,03	449 228	34,35	1939
					137								
Vepse	ou	r	<a href="http://vep.Wikipédia.org">vep.Wikipédia.org</a>	6659	729	20,68	21	487	1600	4,16	850 353	127,70	2007
					135				160				
Komi	ou	r	<a href="http://kv.Wikipédia.org">kv.Wikipédia.org</a>	5441	522	24,91	21	99	000	0,03	311 417	57,24	1938
Livvi-carélien	ou	r	<a href="http://olo.Wikipédia.org">olo.Wikipédia.org</a>	3545	29 478	8,32	14	151	31 000	0,11	238 428	67,26	2007
Komi-permiak	ou	r	<a href="http://koi.Wikipédia.org">koi.Wikipédia.org</a>	3460	57 409	16,59	13	107	63 000	0,05	182 428	52,72	1938
									280				
Touvain	tur	r	<a href="http://tyv.Wikipédia.org">tyv.Wikipédia.org</a>	3167	39 925	12,61	9	101	000	0,01	672 601	212,38	1943
									265				
Bouriate	mo	r	<a href="http://bxr.Wikipédia.org">bxr.Wikipédia.org</a>	2181	54 907	25,18	15	89	000	0,01	418 338	191,81	1939

### Les éditions circumpolaires de Wikipédia dans l'incubateur

Langue	Fam	Pays	URL	Arts	Loc	Art/loc	Mots	Mots/art	Orth
Yupik de l'Alaska central	es	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/esu">incubator.Wikimédia.org/wiki/Wp/esu</a>	193	750	0,01	1681	8,71	1984
Tlingit	déné	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/tli">incubator.Wikimédia.org/wiki/Wp/tli</a> <a href="http://incubator.Wikimédia.org/wiki/Wp/">incubator.Wikimédia.org/wiki/Wp/</a>	47	1360	0,03		0,00	1973
Gwich'in	déné	a	<a href="http://gwi/">gwi/</a>	20	560	0,04		0,00	1970
Aléoute	es	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/ale">incubator.Wikimédia.org/wiki/Wp/ale</a>	8	150			0,00	1990
					11				
Déné-suline	déné	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/chp">incubator.Wikimédia.org/wiki/Wp/chp</a>	4	325	0,00		0,00	1904
Hän	déné	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/haa">incubator.Wikimédia.org/wiki/Wp/haa</a>	1	20	0,05	1	1,00	1970
Tlingit	déné	a	<a href="http://incubator.Wikimédia.org/wiki/Wp/tli">incubator.Wikimédia.org/wiki/Wp/tli</a>	1	1360	0,00	9	9,00	1973
Meänkieli	ou	n	<a href="http://incubator.Wikimédia.org/wiki/Wp/fit">incubator.Wikimédia.org/wiki/Wp/fit</a>	1841	30000	0,06	50167	27,25	1985
Kvène	ou	n	<a href="http://incubator.Wikimédia.org/wiki/Wp/fkv">incubator.Wikimédia.org/wiki/Wp/fkv</a>	610	5000	0,12	16191	26,54	2008
Same du Sud	ou	n	<a href="http://incubator.Wikimédia.org/wiki/Wp/sma">incubator.Wikimédia.org/wiki/Wp/sma</a>	160	600	0,27	1031	6,44	1974

Skolt	ou	n	<a href="https://incubator.wikimedia.org/wiki/Wp/sms">incubator.Wikimédia.org/wiki/Wp/sms</a>	<b>90</b>	320	<b>0,28</b>	<b>1149</b>	12,77	1973
Same de Lule	ou	n	<a href="https://incubator.wikimedia.org/wiki/Wp/smj">incubator.Wikimédia.org/wiki/Wp/smj</a>	<b>43</b>	650	<b>0,07</b>	<b>379</b>	8,81	1983
Carélien	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/krj">incubator.Wikimédia.org/wiki/Wp/krj</a>	<b>1808</b>	36000	<b>0,05</b>	<b>48383</b>	26,76	2007
Khakasse	tur	r	<a href="https://incubator.wikimedia.org/wiki/Wp/kjh">incubator.Wikimédia.org/wiki/Wp/kjh</a>	<b>1445</b>	42604	<b>0,03</b>	<b>21745</b>	15,05	1939
Ingrien	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/izh">incubator.Wikimédia.org/wiki/Wp/izh</a>	<b>677</b>	820	<b>0,83</b>	<b>31391</b>	46,37	2014
Tchouktche	Ps	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ckt">incubator.Wikimédia.org/wiki/Wp/ckt</a>	<b>415</b>	5095	<b>0,08</b>	<b>3923</b>	9,45	1937
Énètse	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/enf">incubator.Wikimédia.org/wiki/Wp/enf</a>	<b>197</b>	43	<b>4,58</b>	<b>654</b>	3,32	1980
Same de Kildin	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/sjd">incubator.Wikimédia.org/wiki/Wp/sjd</a>	<b>143</b>	340	<b>0,42</b>	<b>5165</b>	36,12	1985
Mansi	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/mns">incubator.Wikimédia.org/wiki/Wp/mns</a>	<b>118</b>	940	<b>0,13</b>	<b>4053</b>	34,35	1938
Ket	ps	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ket">incubator.Wikimédia.org/wiki/Wp/ket</a>	<b>71</b>	213	<b>0,33</b>	<b>734</b>	10,34	1980
Nénètse de la Toundra	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/yrk">incubator.Wikimédia.org/wiki/Wp/yrk</a>	<b>66</b>	926	<b>0,00</b>	<b>3555</b>	53,86	1937
Ludic	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/lud">incubator.Wikimédia.org/wiki/Wp/lud</a>	<b>43</b>	3000	<b>0,01</b>	<b>393</b>	9,14	1993
Youkaguir de la Toundra	ps	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ykg">incubator.Wikimédia.org/wiki/Wp/ykg</a>	<b>41</b>	370			0,00	1983
Nanaï	ps	r	<a href="https://incubator.wikimedia.org/wiki/Wp/gld">incubator.Wikimédia.org/wiki/Wp/gld</a>	<b>40</b>	1400	<b>0,03</b>	<b>1628</b>	40,70	1937
Evenki	tun	r	<a href="https://incubator.wikimedia.org/wiki/Wp/evn">incubator.Wikimédia.org/wiki/Wp/evn</a>	<b>26</b>	26580	<b>0,00</b>	<b>2526</b>	97,15	1937
Yupik sibérien	par	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ess">incubator.Wikimédia.org/wiki/Wp/ess</a>	<b>26</b>	1300	<b>0,02</b>	<b>183</b>	7,04	1937
Dolgane	tur	r	<a href="https://incubator.wikimedia.org/wiki/Wp/dlg">incubator.Wikimédia.org/wiki/Wp/dlg</a>	<b>4</b>	1100	<b>0,00</b>	<b>135</b>	33,75	1961
Khanty	ou	r	<a href="https://incubator.wikimedia.org/wiki/Wp/kca">incubator.Wikimédia.org/wiki/Wp/kca</a>	<b>4</b>	9600	<b>0,00</b>	<b>159</b>	39,75	1937
Oudihé	tun	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ude">incubator.Wikimédia.org/wiki/Wp/ude</a>	<b>4</b>	100	<b>0,04</b>	<b>133</b>	33,25	1989
Orotche	tun	r	<a href="https://incubator.wikimedia.org/wiki/Wp/oac">incubator.Wikimédia.org/wiki/Wp/oac</a>	<b>3</b>	8	<b>0,38</b>	<b>123</b>	41,00	1994
Oultch	tun	r	<a href="https://incubator.wikimedia.org/wiki/Wp/ulc">incubator.Wikimédia.org/wiki/Wp/ulc</a>	<b>3</b>	150	<b>0,02</b>	<b>116</b>	38,67	1989
Négidal	ps	r	<a href="https://incubator.wikimedia.org/wiki/Wp/neg">incubator.Wikimédia.org/wiki/Wp/neg</a>	<b>2</b>	75	<b>0,03</b>	<b>117</b>	58,50	1992

Les langues circumpolaires sans édition de Wikipédia sont :

l'ahtna, l'alutiïq, le tsimshian de la côte, le deg hit'an, le dena'ina, le nénètse des forêts, le gitksan, le haïda, l'inuktitut, l'inuvialuktun, le kaska, le youkaguir de la Kolima, le koyukon, le kolchan, le bas tanana, le naukan, le nganasanne, le nisgha, le slavey du Nord, le tutchoni du Nord, le same de Pite, le slavey du Sud, le tutchoni du Sud, le t̥t̥ç̥ç̥ [dogrib], le tanacross, le same d'Ume et le haut tanana.