

La place des projets Wikimedia en français dans l'environnement linguistique mondial de l'internet

Rapport de Wikimedia France

pour

le Ministère de la Culture et de la Communication

31 mai 2011



Table des matières

1 Le français sur les projets Wikimedia.....	3
1.1 Wikipédia en français.....	3
1.2 La place du français sur Wikipédia.....	3
1.3 Les autres projets Wikimedia.....	4
1.3.1 Wiktionnaire.....	5
1.3.2 Wikisource.....	5
1.3.3 Wikinews.....	7
1.3.4 Wikibooks.....	7
1.3.5 Wikiversity.....	8
1.3.6 Wikiquote.....	8
1.3.7 Wikimedia Commons.....	10
1.3.8 Wikispecies.....	10
1.3.9 MediaWiki.....	10
1.4 Les différentes formes de français sur Wikipédia.....	11
2 Communauté des Wikipédiens.....	12
2.1 Nombre de Wikipédiens.....	12
2.2 Chapters.....	14
3 Couverture thématique.....	15
3.1 Répartition des articles.....	15
3.2 Projets.....	17
3.3 Articles de qualité.....	17
3.4 Wikiconcours.....	17
4 Multilinguisme/traductions.....	18
4.1 Niveau de compétences.....	18
4.2 Traduction.....	20
5 Conclusion.....	21

1 Le français sur les projets Wikimedia

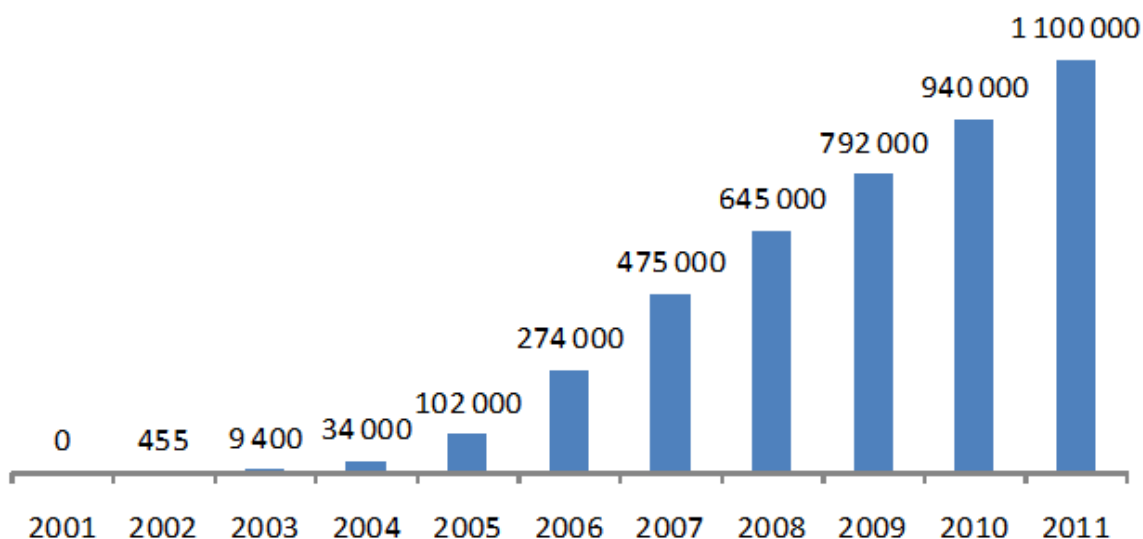
1.1 Wikipédia en français

Wikipédia est un projet d'encyclopédie multilingue sur internet, rédigée par les internautes et publiée sous licence libre.

Consultée par plus de 16 millions de Français chaque mois, et 1,8 million chaque jour, Wikipédia est le 8^{ème} site le plus visité en France¹.

Créée en 2001, la version francophone compte actuellement 1 100 000 articles, ce qui la place au 3^{ème} rang derrière l'anglophone et la germanophone.

Nombre d'articles sur la Wikipédia francophone
(données au 30 avril de chaque année)



1.2 La place du français sur Wikipédia

Wikipédia est une encyclopédie multilingue, initialement créée en anglais en janvier 2001. Elle s'est enrichie quelques mois plus tard de 12 autres langues : le français, l'allemand, le polonais, l'italien, le japonais, le russe, l'espagnol, le portugais, le chinois, le catalan, l'espéranto et l'hébreu.

10 ans après, ce sont 268 langues qui sont représentées, y compris des langues construites (espéranto, volapük, interlingua), des langues mortes (latin), des langues menacées et des langues régionales (alsacien, basque, breton, francoprovençal, occitan, tahitien, etc.).

Cependant, tous les projets linguistiques ne connaissent pas le même développement. Ainsi, si on compte plus de 3 300 000 articles en anglais, plus de la moitié des langues en compte moins de 5 000.

Le français, qui a dépassé le million d'articles en septembre 2010, est la 3^{ème} langue la plus représentée, derrière l'allemand. Il contient actuellement autant d'articles que les 205 langues les moins représentées réunies.

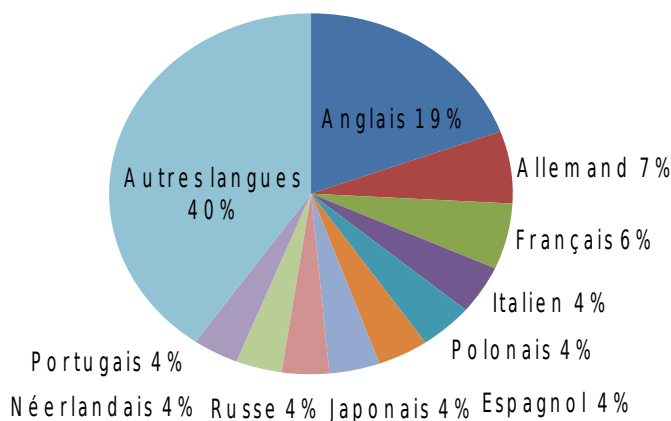
¹ Données Médiamétrie en avril 2011 : <http://www.mediametrie.fr/internet/communiques/telecharger.php?f=428fca9bc1921c25c5121f9da7815cde>

On notera l'absence du chinois parmi les 10 premières langues. Victime de censures régulières de la part des autorités chinoises, Wikipédia a eu du mal à s'y développer. Le chinois est la seule langue où Wikipédia connaît de sérieux concurrents, à savoir Baidu Baïke et Hudong. Il y a d'autres facteurs expliquant la faible croissance du chinois, dont le fait qu'il y a plusieurs « chinois » et de faibles infrastructures Internet en province.

	Langue	Articles	Visites/heure
1	Anglais	3 628 406	9 423 506
2	Allemand	1 241 312	1 374 153
3	Français	1 097 559	867 376
4	Italien	805 753	601 130
5	Polonais	800 410	552 443
6	Espagnol	756 206	1 429 304
7	Japonais	752 373	1 594 864
8	Russe	708 498	869 949
9	Néerlandais	684 105	234 087
10	Portugais	683 153	489 326

Les 10 principales langues de Wikipedia au 30 avril 2011²

Répartition des articles de Wikipédia par langue Avril 2011



La domination des langues « majeures » faiblit. L'anglais qui représentait 31% des articles en avril 2005 est maintenant à moins de 20%. En 5 ans (avril 2006-avril 2011), les 10 principales langues sont passées de 71% à 60% des articles. On notera que le français maintient pourtant une part constante de 6%.

Ce fort développement des langues « secondaires » est amené à se poursuivre, notamment grâce aux actions annoncées par la Wikimedia Foundation pour faciliter la libre diffusion de la connaissance dans les pays du Sud.

² Classement complet des 168 langues sur <http://stats.wikimedia.org/EN/Sitemap.htm#comparisons>

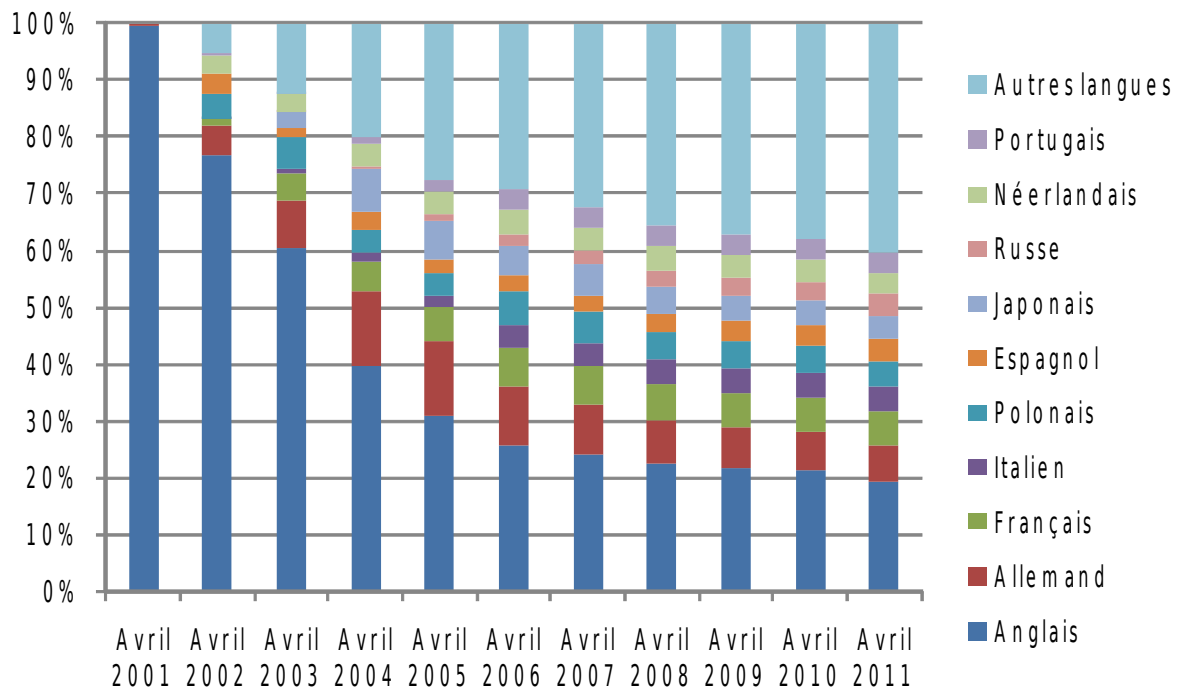


Illustration 1: Évolution de la répartition des articles par langue

1.3 Les autres projets Wikimedia

La Wikimedia Foundation héberge d'autres projets que Wikipédia. Basés sur un fonctionnement proche – publiés sous licence libre, éditables par tous et à vocation culturelle – ils sont également multilingues.

Dans ces projets frères, la version anglophone est quasiment toujours la version principale. Le français est également systématiquement dans le peloton de tête. En revanche, ce n'est pas le cas des autres principales langues de Wikipédia (allemand et espagnol en particulier).

1.3.1 Wiktionnaire

Le Wiktionnaire est un dictionnaire libre et gratuit que chacun peut modifier et améliorer.

La version en français est la deuxième en nombre de termes définis. Elle talonne l'anglais avec 2 millions d'articles contre 2,4 millions. La fréquentation du Wiktionnaire francophone est en revanche 3 fois plus faible (30 000 visites/heure contre 100 000).

Dans les langues les plus représentées, on note la présence de quelques surprises comme le malgache (3^{ème}) ou le lituanien (5^{ème}). Des langues plus « classiques » sont en retrait : l'allemand est 12^{ème} avec 180 000 articles et l'espagnol est 25^{ème}, derrière le lao.

Parmi les 136 langues représentées, on retrouve des langues régionales : le catalan (29 227 articles), le breton (12 937), l'occitan (16 701), le basque (8 484) et le corse (2 095). Le Wiktionnaire apparaît ainsi comme un outil pour préserver ces langues et faciliter leur enseignement.

	Langue	Articles	Visites/heure
1	Anglais	2 441 668	102 102
2	Français	2 021 635	29 908
3	Malgache	1 195 154	379
4	Chinois	1 062 641	5 143
5	Lituanien	558 474	1 028
6	Russe	288 913	22 483
7	Turc	276 803	2 681
8	Vietnamien	229 112	1 193
9	Polonais	218 761	5 603
10	Tamoul	198 331	588

Les 10 principales langues du Wiktionnaire au 30 avril 2011³

1.3.2 Wikisource

Wikisource est une bibliothèque numérique gratuite et librement modifiable ; une des principales spécificités par rapport à d'autres bibliothèques numériques (Gallica, Europeana, Google Books, Internet Archive, etc.) est de comporter une version texte des livres numérisés, cette retranscription étant faite par des internautes, généralement à partir d'une version obtenue à partir de reconnaissance optique de caractères.

Wikimédia France, association de promotion des projets Wikimedia en France, a signé en 2010 un partenariat avec la Bibliothèque nationale de France. Celui-ci comportait l'import de 1300 livres du domaine public sur Wikisource afin de les diffuser et de les corriger en vue d'obtenir une version texte. La correction avance petit à petit, même si la quantité de travail fournie par la communauté des contributeurs sur la Wikisource francophone est très importante.

Wikisource compte au total près de 900 000 pages en 47 langues. Avec 60 000 pages, la version francophone se classe au 4^{ème} rang, derrière l'anglais, le russe et le portugais.

	Langue	Pages de contenu	Visites/heure
1	Anglais	196 391	11 353
2	Russe	116 731	5 801
3	Portugais	86 657	941
4	Français	59 756	7 408
5	Chinois	55 252	1 406
6	Allemand	54 031	2 453
7	Espagnol	46 341	3 898
8	Italien	42 054	2 215
9	Hébreu	32 754	146
10	Arabe	19 854	1 118

Les 10 principales langues de Wikisource au 30 avril 2011⁴

³ Classement complet des 136 langues sur <http://stats.wikimedia.org/wiktionary/EN/Sitemap.htm>

⁴ Classement complet des 47 langues sur <http://stats.wikimedia.org/wikisource/EN/Sitemap.htm>

1.3.3 Wikinews

Wikinews est une source d'informations d'actualité libre, gratuite et modifiable par tous.

La première version de Wikinews est, de loin, celle en serbe avec 63 000 articles. Elle est alimentée par un robot reprenant des dépêches de l'agence Beta News qui a donné son accord pour leur publication sous licence libre.

Plusieurs autres agences de presse publient leurs dépêches sous licence libre, y compris des informations en français. Il s'agit par exemple de Voice of America (États-Unis), Agência Brasil (Brésil) ou RIA Novosti (Russie). À l'heure où les dépêches sont facilement et gratuitement accessibles à tout internaute, cette utilisation des licences libres permet à des particuliers ou à des sites moins importants de reproduire ces informations, et non seulement à des gros sites d'actualité. L'utilisation de licences libres permet aux pays contrôlant ces agences de presse de diffuser massivement leur point de vue sur des événements d'actualité, notamment sur les conflits internationaux.

	Langue	Articles	Visites/heure
1	Serbe	62 996	230
2	Anglais	18 616	6 748
3	Polonais	12 223	825
4	Français	10 113	1 906
5	Italien	7 943	445
6	Espagnol	6 667	870
7	Portugais	6 445	441
8	Chinois	4 113	341
9	Suédois	3 047	92
10	Russe	2 199	348

Les 10 principales langues de Wikinews au 30 avril 2011⁵

1.3.4 Wikibooks

Wikibooks est une bibliothèque de livres pédagogiques, librement modifiable par chacun.

Wikibooks compte près de 130 000 pages en 42 langues.

⁵ Classement complet des 24 langues sur <http://stats.wikimedia.org/wikinews/EN/Sitemap.htm>

	Langue	Articles	Visites/heure
1	Anglais	35 219	16 085
2	Allemand	16 957	5 715
3	Français	10 256	361
4	Hongrois	9 721	421
5	Portugais	6 679	3 364
6	Néerlandais	5 701	849
7	Espagnol	5 425	3 305
8	Polonais	5 127	1 678
9	Italien	4 973	1 239
10	Hébreu	4 164	525

Les 10 principales langues de Wikibooks au 30 avril 2011⁶

1.3.5 Wikiversity

Wikiversity est une communauté pédagogique libre, gratuite et à laquelle chacun peut prendre part.

	Langue	Articles	Visites/heure
1	Anglais	15 718	4 068
2	Français	8 237	2 928
3	Russe	2 243	271
4	Tchèque	1 992	92
5	Allemand	1 827	452
6	Portugais	1 749	305
7	Espagnol	1 017	598
8	Italien	976	401
9	Grec	433	94
10	Finnois	374	57
11	Suédois	323	23
-	Multilingue ⁷	1 917	252

Les 12 langues de la Wikiversité au 30 avril 2011⁸

⁶ Classement complet des 42 langues sur <http://stats.wikimedia.org/wikibooks/EN/Sitemap.htm>

⁷ La version multilingue permet de coordonner les différents projets et sert d'incubateur pour créer des Wikiversités dans d'autres langues

⁸ Données mises à jour régulièrement sur <http://stats.wikimedia.org/wikiversity/EN/Sitemap.htm>

1.3.6 Wikiquote

Wikiquote est un projet libre et gratuit de collecte de citations, modifiable par chacun.

Il compte 98 000 pages en 53 langues.

	Langue	Articles	Visites/heure
1	Anglais	17 418	27 095
2	Polonais	14 158	3 209
3	Italien	12 699	5 425
4	Russe	6 317	2 276
5	Espagnol	4 654	5 285
6	Bulgare	3 731	319
7	Tchèque	3 119	272
8	Français	2 697	3 925
9	Turc	2 614	966
10	Slovène	2 427	108

Les 10 principales langues de Wikiquote au 30 avril 2011⁹

1.3.7 Wikimedia Commons

Wikimedia Commons est une médiathèque détenant aujourd'hui près de 10 250 000 fichiers, librement réutilisables et que chacun peut toujours enrichir. Les images, vidéos et fichiers audio hébergés sur Commons permettent d'illustrer les pages de Wikipédia et des autres projets Wikimedia.

Contrairement à la plupart des autres projets, Commons ne comprend qu'un seul site où cohabitent les différentes langues. Ainsi, une image peut être décrite en plusieurs langues sur la même page, et l'interface linguistique peut être déclinée pour s'adapter à la langue du lecteur. Si l'anglais est la langue la plus utilisée pour communiquer entre les contributeurs, ceux-ci disposent d'espaces de discussions dans une quarantaine de langues¹⁰. En avril 2011, la page de discussion en anglais était la plus utilisée, devant la germanophone et la francophone.

1.3.8 Wikispecies

Wikispecies est un inventaire du vivant. Il recense actuellement près de 270 000 espèces vivantes. Plutôt orienté vers les spécialistes, il s'agit d'un des deux seuls projets (avec Commons) n'ayant pas de déclinaisons par langue, le contenu étant en latin (noms des espèces) et dans les autres langues pour lesquelles l'espèce a un nom vernaculaire, la langue de communication du site étant généralement l'anglais.

1.3.9 MediaWiki

MediaWiki est le logiciel faisant fonctionner Wikipédia et tous les sites internet de type « wiki ». Son interface existe dans 348 langues. Ces langues comprennent parfois plusieurs dialectes d'une

⁹ Classement complet des 53 langues sur <http://stats.wikimedia.org/wikiquote/EN/Sitemap.htm>

¹⁰ Accéder au « Bistro » francophone : <http://commons.wikimedia.org/wiki/Commons:Bistro>

même langue : en plus de l'arabe « classique » il y a par exemple l'arabe algérien, l'arabe marocain et l'arabe égyptien. En revanche, il n'y a qu'une seule version de français.

En moyenne, 47% des messages de la version de base de MediaWiki (hors extensions) sont traduits. Lorsque le message n'est pas traduit dans la langue choisie, l'utilisateur voit alors le message s'afficher dans une autre langue, généralement l'anglais.

181 langues, soit 52%, ont plus de 99% de leurs messages les plus utilisés traduits. En revanche, les traductions sont beaucoup moins nombreuses pour les messages moins utilisés. Ainsi, seules 6 langues voient plus de 99% de l'intégralité de leurs messages traduits, extensions incluses. Il s'agit de l'allemand, du macédonien, de l'anglais, du néerlandais, de l'interlingua et du français.

Part de messages traduits	Nombre de langues			
	Messages les plus utilisés	Version de base de MediaWiki	Extensions utilisées sur les projets Wikimedia	Toutes les extensions
0,00%	3	3	42	25
>0%	345	345	306	323
>25%	285	216	89	74
>50%	225	149	71	52
>75%	203	111	59	30
>99%	181	39	19	6

1.4 Les différentes formes de français sur Wikipédia

Différentes règles ont progressivement été mises en place en termes de forme du français.

Ainsi pour les titres d'articles par exemple, le principe dit « de moindre surprise » privilégie le terme le plus connu et répandu dans la société, sauf exception pour les sigles où le nom dans son entier est alors retenu.

Pour les introductions des biographies, le terme « mort » est privilégié plutôt que « décédé ». Le doute subsiste toutefois pour le temps de conjugaison utilisé pour décrire la profession d'une personne décédée. Doit-on ainsi écrire: « X est un auteur », « X était un auteur » ou « X fut un auteur » ? Actuellement, et faute de consensus trouvé, ces trois types de phrases se retrouvent sur Wikipédia selon les articles.

Des divergences existent également selon les régionalismes, parmi lesquels par exemple « endive / chicon », ou encore « pain au chocolat / chocolatine ». Ces divergences peuvent même se complexifier lorsqu'un terme est couramment employé dans d'autres pays francophones. C'est le cas du terme « couque au chocolat » qui s'ajoute encore aux deux précédents, terme utilisé principalement en Belgique. Il en va de même pour certains mots beaucoup plus utilisés dans la province de Québec au Canada. Les Suisses francophones écrivent pour leur part « nonante » plutôt que « quatre-vingt-dix ». L'on tente dans ces cas d'appliquer le principe de moindre surprise.

Les avis diffèrent également quant à l'utilisation de l'orthographe didactique ou courante dans les titres lors des transcriptions des termes issus de langues non-occidentales, par exemple le japonais

avec « Tōkyō » ou « Tokyo » ? (actuellement la première orthographe est choisie en titre de l'article).

Parmi les autres points d'achoppement, on peut encore citer l'application de la réforme de l'orthographe de 1990, inégalement acceptée et promue au sein de la francophonie ou la féminisation des noms de métiers.

Enfin une page nommée « conventions typographiques »¹¹ regroupe sur Wikipédia toutes les conventions, règles et/ou recommandations à suivre dans la rédaction des articles. Elles traitent de divers points parmi lesquels les apostrophes (*apostrophe typographique déconseillée au profit de l'apostrophe dactylographique*), les types de guillemets (« » plutôt que " "), les heures, les cas d'utilisation de l'emphase (gras) ou de l'italique, etc.

À noter également que l'accentuation des majuscules et lettres capitales est obligatoire.

2 Communauté des Wikipédiens

2.1 Nombre de Wikipédiens

Chaque mois, près de 5 000 personnes effectuent plus de 5 contributions sur la Wikipédia en français. La version francophone se situe ainsi au 3e rang des projets ayant le plus de contributeurs actifs, derrière l'anglais (40 000) et l'allemand (7 000).

En avril 2011, 83% des contributeurs actifs sur les différentes Wikipédia se concentrent sur les 10 principales langues.

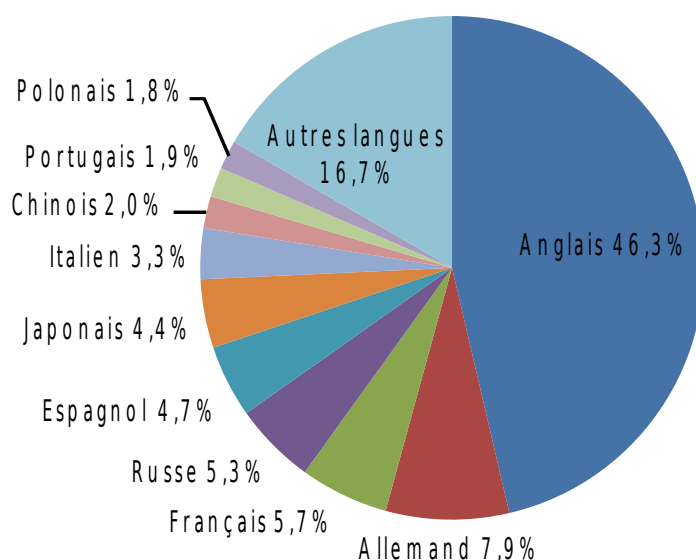


Illustration 2: Répartition par langue des contributeurs actifs (plus de 5 modifications durant le mois d'avril 2011)

Cette tendance est légèrement plus faible parmi les contributeurs très actifs : 77% des personnes effectuant plus de 100 modifications par mois se concentrent dans les 10 premières langues.

¹¹ Contenu visible sur http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Conventions_typographiques

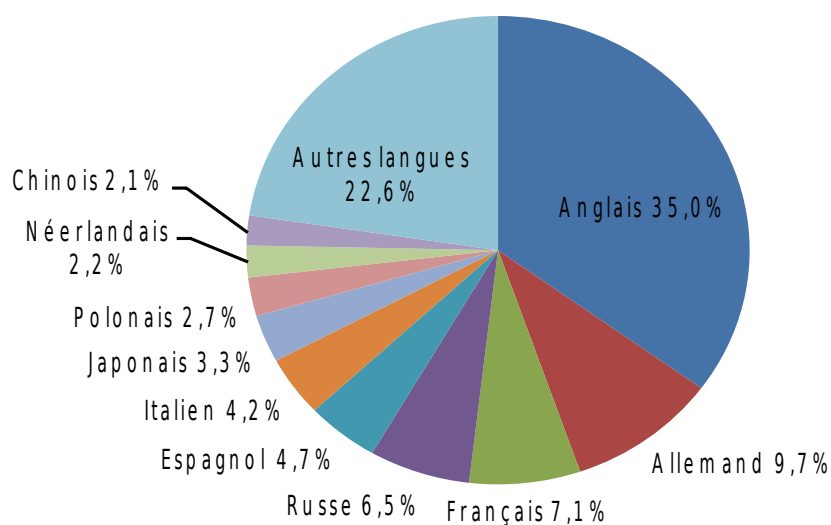


Illustration 3: Répartition par langue des contributeurs très actifs (plus de 100 modifications durant le mois d'avril 2011)

Sur la Wikipédia francophone, le nombre de contributeurs est stable depuis 2009. On compte ainsi 4 900 contributeurs actifs (5 modifications par mois) et 760 contributeurs très actifs (100 modifications par mois).

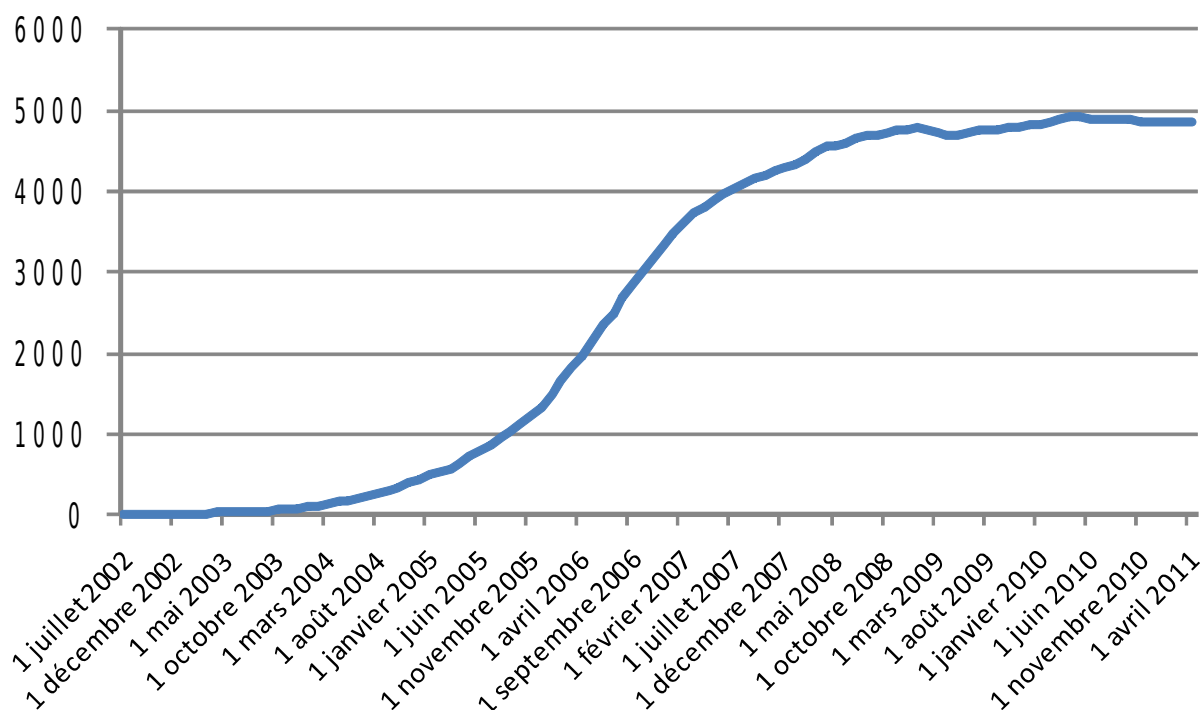


Illustration 4: Évolution du nombre de contributeurs actifs sur la Wikipédia francophone (5 modifications par mois – moyenne annuelle mobile)

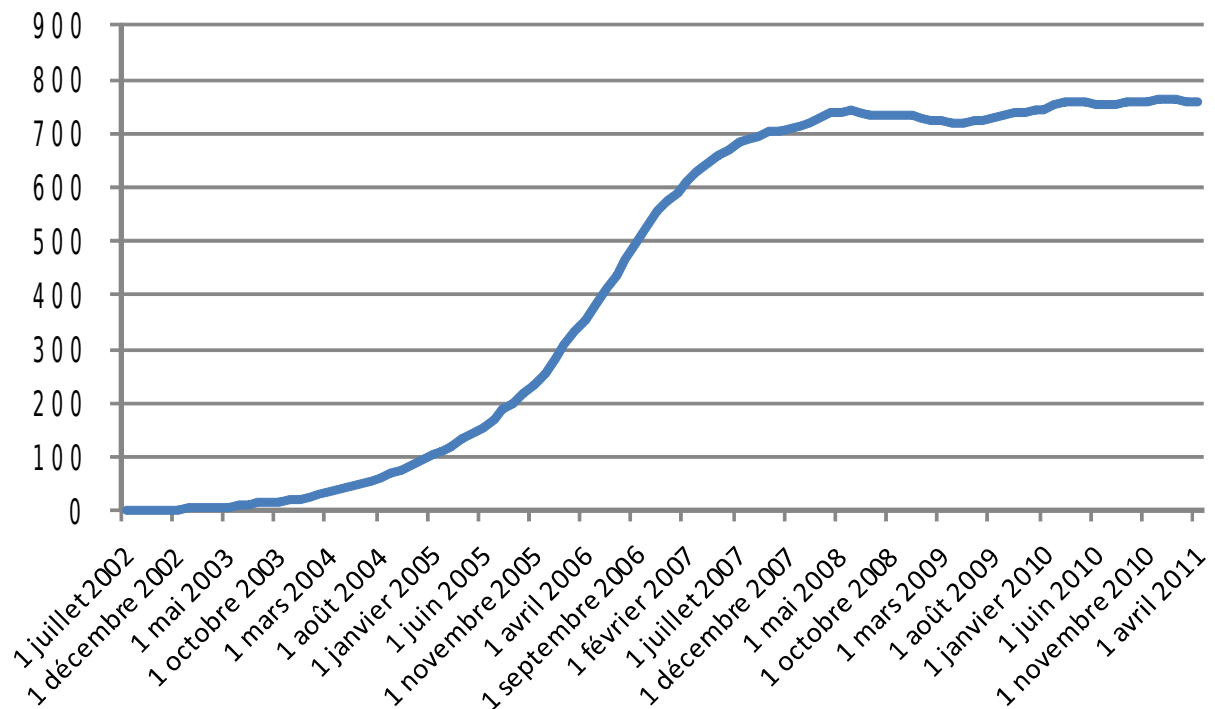


Illustration 5: Évolution du nombre de contributeurs très actifs sur la Wikipédia francophone (100 modifications par mois – moyenne annuelle mobile)

2.2 Profil des Wikipédiens

La récente étude « Enquête sur les utilisateurs de Wikipédia » menée par l'observatoire M@rsouin¹² a recueilli les réponses de 16 000 internautes, en majorité des lecteurs puisque les « contributeurs réguliers » et « gros contributeurs » ne représentent que 20% des répondants (~ 3200 personnes).

L'étude révèle que les contributeurs à Wikipédia sont plus jeunes que la moyenne des internautes, avec près d'un tiers des répondants âgés de 20 à 30 ans et plus d'un quart de moins de 20 ans. Cet état induit un nombre important de scolaires et étudiants contribuant au projet encyclopédique. De plus, pour les contributeurs détenant un emploi, près de la moitié d'entre eux occupent des positions d'encadrement. Étudiants/lycéens et cadres supérieurs impliquent pour ces contributeurs des activités où la recherche et l'analyse et le traitement d'informations est important.

Il apparaît également que les femmes sont beaucoup moins présentes que les hommes, avec une proportion de l'ordre de 20 %. Cette information n'est toutefois pas nouvelle pour Wikipédia, la Wikimedia Foundation elle-même tentant déjà depuis peu d'inciter les femmes à contribuer¹³.

Origine des contributeurs

Sur l'ensemble des Wikipédia, 82,5% des modifications sont effectuées depuis un pays du Nord¹⁴. Ces pays ne représentent que 19 % de la population mondiale, mais 62 % des utilisateurs d'internet. Leur représentation parmi les Wikipédiens est donc encore plus forte.

¹² Résultats d'étude consultables sur <http://www.marsouin.org/spip.php?article420>

¹³ Le New York Times traite de ce sujet sur http://www.nytimes.com/2011/01/31/business/media/31link.html?_r=1

¹⁴ Statistiques détaillées sur

<http://stats.wikimedia.org/wikimedia/squids/SquidReportPageEditsPerCountryOverview.htm>

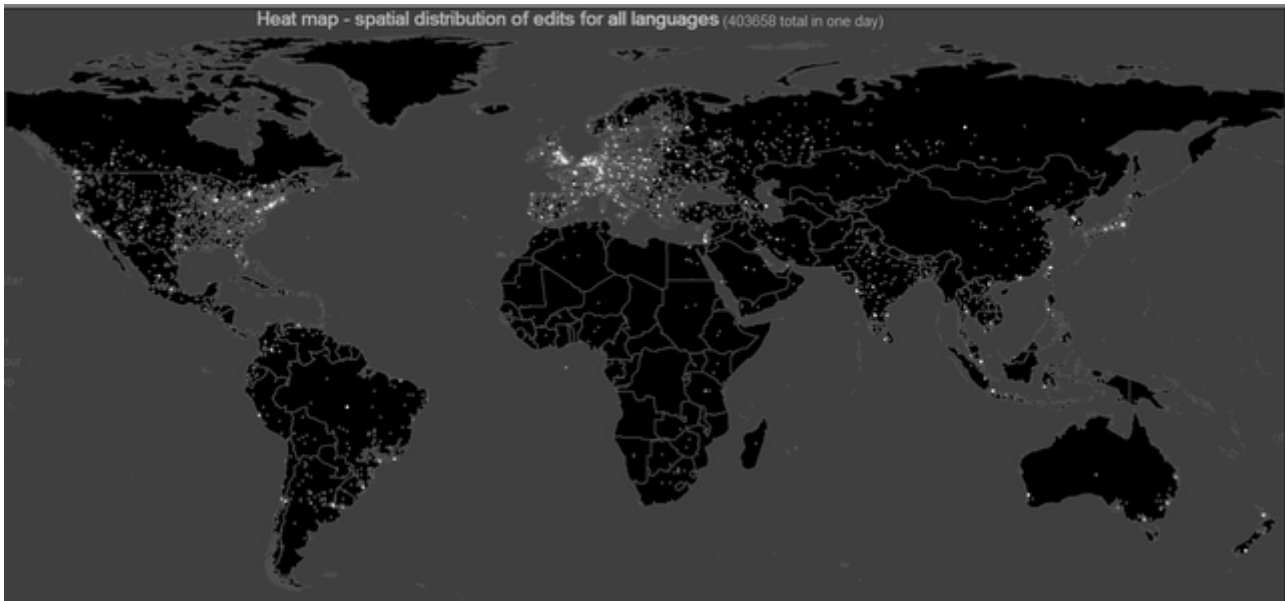


Illustration 6: Répartition spatiale des contributions dans toutes les langues de Wikipédia

Wikipédia en français ne déroge pas à cette règle puisque 79 % de ses contributeurs sont en France, 6 % en Belgique, 5 % au Canada, 2,5 % en Suisse et 1 % aux États-Unis.

Les pays du Nord sont un peu moins bien représentés parmi les lecteurs. Les pays du Maghreb représentent ainsi un peu plus de 5 % des consultations de Wikipédia en français. Un phénomène similaire est observé sur la Wikipédia en espagnol : les Mexicains y effectuent 13,5 % des modifications et 26,2 % des consultations.

Pays	Part des modifications	Part des consultations
France	78,6 %	72,2 %
Belgique	6,2 %	5,6 %
Canada	4,8 %	6,9 %
Suisse	2,5 %	3,2 %
États-Unis	1,0 %	0,9 %
Allemagne	0,8 %	0,7 %
Italie	0,7 %	ND
Royaume-Uni	0,6 %	0,5 %
Maroc	0,6 %	2,1 %
Espagne	0,5 %	ND
Algérie	ND	1,5 %
Tunisie	ND	1,1 %

Répartition des modifications et des consultations par pays

2.2 Associations locales

Les chapters (en français "associations locales") sont des organisations indépendantes de l'organisation qui héberge Wikipédia et les projets Wikimedia, la Wikimedia Foundation. Leur rôle est de promouvoir localement les projets Wikimedia, en expliquant leur fonctionnement et en favorisant la création de contenus libres pouvant enrichir les projets.

Il y a actuellement 32 chapters officiels, dont 3 concernant des pays francophones¹⁵.

Wikimedia France, a été créée en 2004 et comporte 250 membres en 2011. Si la totalité des discussions ont lieu en français, l'association soutient également des projets dans les langues régionales.

Wikimedia CH a été créée en 2006 et comporte 90 membres en 2011. À l'instar de la Suisse, l'association est fortement multilingue, les discussions internes ayant lieu en français, en allemand et en anglais et tente de créer des projets pour chacune des communautés linguistiques nationales (francophone, germanophone et italophone). Le romanche est assez peu développé du fait du faible nombre de locuteurs, mais Wikimedia CH essaye de le prendre en compte sur ses supports de communication, notamment le site web.

Wikimedia Canada est le dernier-né des chapitres Wikimédia, reconnu le 24 mai 2011. En raison de la taille et de l'organisation du pays, ainsi que des deux langues officielles, un sous-groupe Québec devrait rapidement se mettre en place.

D'autres associations nationales tentent de se créer ou d'être reconnues comme *chapters*. C'est le cas de la Belgique et du Mali qui comporte une association ayant déjà mené quelques projets.

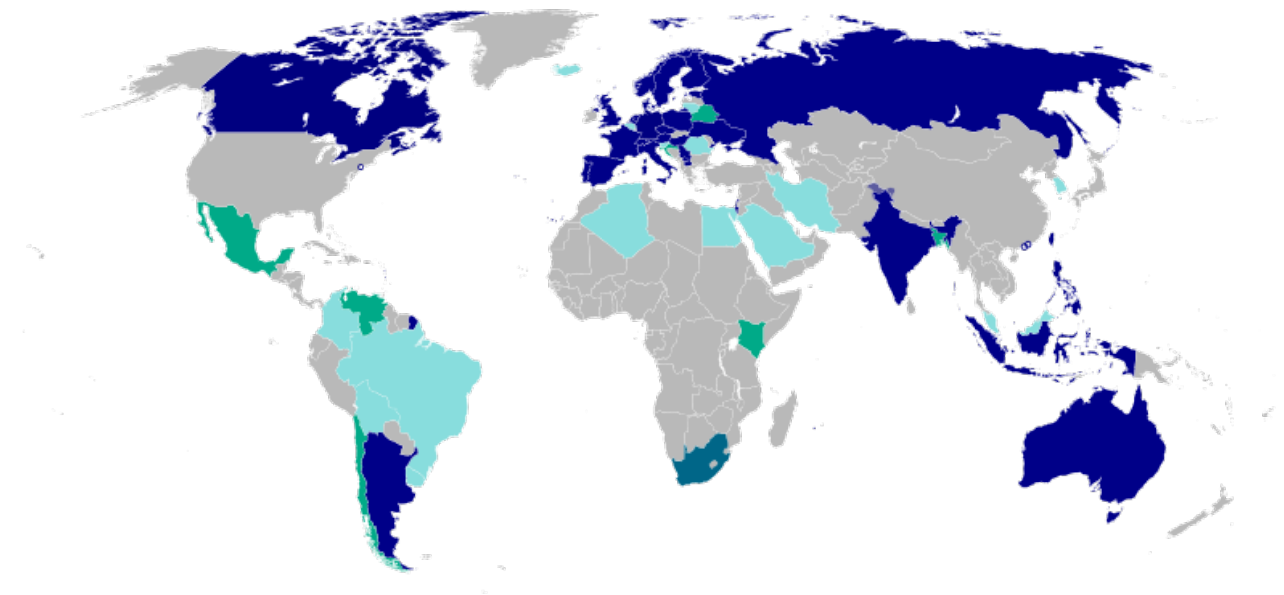


Illustration 7: Répartition des chapters Wikimedia

■	Associations existantes
■	Approuvées, mais pas encore fondées
■	Associations en cours de discussions avancées avec la fondation
■	Associations en cours de création ou demandant à être reconnues comme chapter

¹⁵ Liste complète et actualisée sur http://meta.wikimedia.org/wiki/Wikimedia_chapters

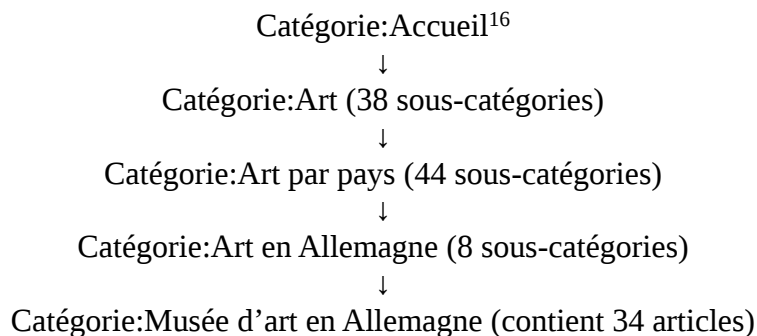
3 Couverture thématique

3.1 Répartition des articles

Répartition thématique

Tous les articles de Wikipédia doivent, dans la mesure du possible, être liés à un portail et son projet équivalent (cf. Projets *infra*).

Ils doivent également être « catégorisés ». Encore plus que pour les projets, les catégories doivent pouvoir être les plus précises possibles tout en restant liées entre elles. Ces liens créent l'« arborescence » des articles. Ainsi l'on peut retrouver l'arborescence suivante :



Tout article doit pouvoir avoir un ou plusieurs portails (et projets) et une ou plusieurs catégories. Pour les nouveaux articles, créés par des internautes pas toujours sensibilisés à leur existence, plusieurs opérations de maintenance se mettent en place afin de les catégoriser au plus vite.

L'analyse des catégories utilisées permet d'estimer la répartition des articles par thème :

Thème	Part des articles
Géographie et lieux	22 %
Culture et arts	17 %
Personnalités	15 %
Société et sciences sociales	12 %
Techniques et sciences appliquées	10 %
Histoire	10 %
Sciences physiques et naturelles	7 %
Mathématiques et logique	3 %
Santé	3 %
Religions et croyances	1 %

Répartition des articles de Wikipédia par thème

¹⁶ Sommet de l'arborescence sur <http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Cat%C3%A9gories>

Répartition géographique

56% des articles sont reliés à un portail géographique. Il s'agit à la fois d'articles « géographiques », traitant d'un pays ou d'une ville, et d'articles plus généraux traitant par exemple d'une personne ayant la nationalité du pays indiqué. Les articles non localisés concernent des sujets plus universels comme la *pomme* ou le *football* par exemple.

60% des articles ainsi localisés concernent l'Europe, la moitié d'entre eux étant relative à la France. Les autres pays les plus représentés sont les États-Unis, l'Italie et le Canada qui doit 55 % de ses articles au Québec.

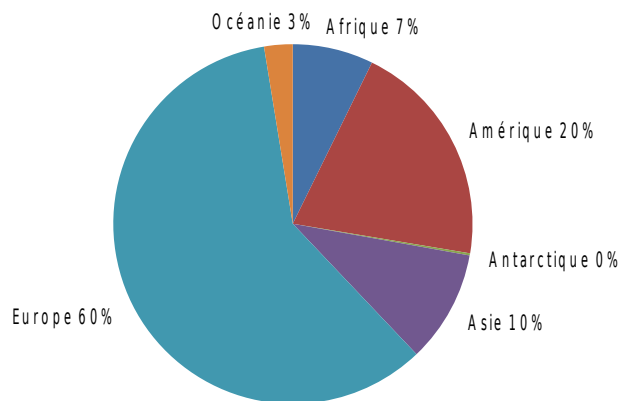


Illustration 8: Répartition des sujets des articles de la Wikipédia francophone par continent

	Pays	Articles	Part des articles de Wikipédia
1	France	165 975	15,0 %
2	États-Unis	58 708	5,3 %
3	Italie	40 606	3,7 %
4	Canada	33 446	3,0 %
5	Allemagne	26 754	2,4 %
6	Royaume-Uni	23 104	2,1 %
7	Belgique	18 628	1,7 %
8	Espagne	16 027	1,4 %
9	Suisse	14 166	1,3 %
10	Japon	13 481	1,2 %

Répartition des sujets des articles par pays concerné

3.2 Projets

La communauté des Wikipédiens est organisée autour de 560 projets et sous-projets. S'il existe une cinquantaine de projets techniques (corrections orthographiques, accessibilité, etc.), la grande majorité concerne des thématiques, parfois divisées en sous-thèmes (exemple : projet sport regroupant les projets football, hockey sur glace, tennis, jeux olympiques, etc.). Ces projets permettent de coordonner les actions des contributeurs à la fois pour définir des actions prioritaires (exemple : création d'articles sur les champions olympiques d'hiver) et pour garantir une certaine homogénéité entre les articles. Ces projets permettent également d'échanger tant sur le fond que sur la forme des articles.

3.3 Articles de qualité

Les articles peuvent obtenir le statut « Article de qualité » ou « Bon article » lorsqu'ils atteignent un niveau de qualité jugé satisfaisant. Ces labels sont obtenus après un vote des contributeurs évaluant l'exhaustivité, la neutralité, la qualité des sources, la rédaction et l'illustration de l'article. Actuellement, plus de 2 200 articles sont ainsi labellisés¹⁷.

Les articles labellisés concernent principalement les sciences humaines (35 %) et les arts et culture (21 %).

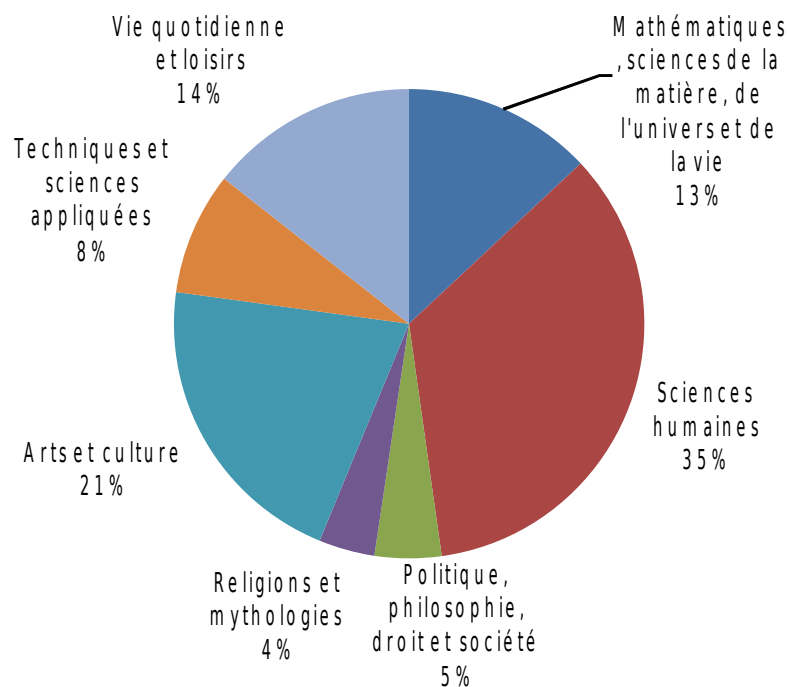
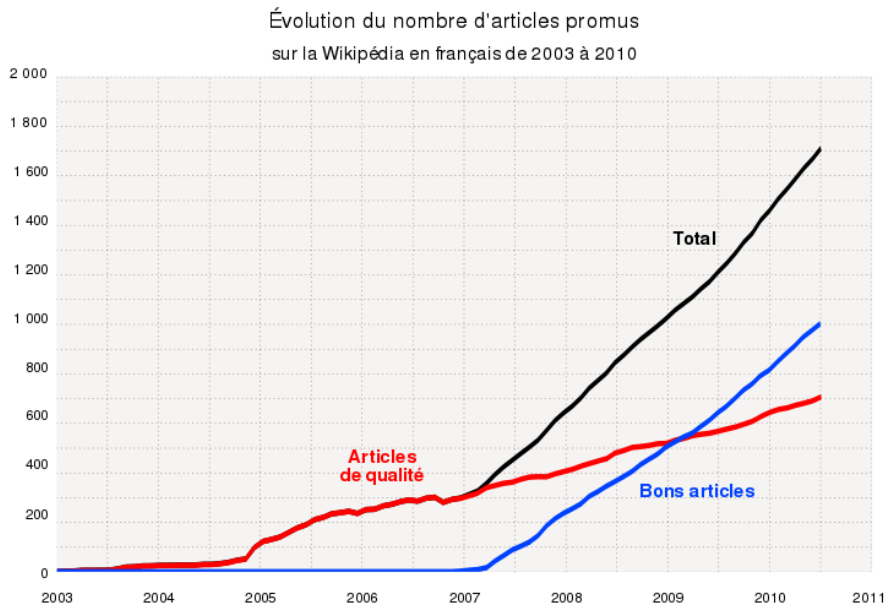


Illustration 9: Répartition des articles labellisés par thèmes

¹⁷ Liste complète et actualisée des articles labellisés sur http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Contenus_de_qualit%C3%A9

3.4 Wikiconcours

Depuis 2006 s'organisent en mars et en septembre des wikiconcours. Il s'agit de faire intervenir en équipe plusieurs Wikipédiens, qui concourent sur un certain nombre d'articles selon deux formules :

- améliorer leur qualité en vue de leur faire obtenir le label « article de qualité »
- améliorer la quantité d'informations sur des articles considérés à l'état d' « ébauches ».

D'une durée de deux semaines à deux mois selon la formule, le wikiconcours permet à chaque équipe de travailler sur plusieurs articles associés par un thème libre. L'avancement des équipes est ensuite laissé à l'appréciation d'un jury composé d'autres contributeurs, chargés à l'issue de leur délibération, de remettre différents prix, récompensant à la fois équipes et articles.

Le jury choisit également des thèmes pour lesquels ils remettent des prix dits « spéciaux ». Ceux-ci diffèrent d'un concours à l'autre mais l'on peut retrouver à titre d'exemple :

- l'article le mieux illustré
- le meilleur sourçage
- l'équipe meilleur espoir.

Le dernier wikiconcours, formule « désébauchage », s'est déroulé du 1^{er} au 15 mars 2011. Plusieurs prix ont été remis, récompensant la quantité et la qualité du contenu rajouté aux articles désébauchés.

Édition	Équipes inscrites	Participants	Meilleur thème
2006	40	40	Élection présidentielle française de 1974 ¹⁸
Été 2007	24	95	New-York
Printemps 2008	42	122	Sport
Automne 2008	36	91	Tunisie Archéologie
Printemps 2009	55	123	Les États-Unis
Automne 2009	38	77	J.R.R. Tolkien et son œuvre
Printemps 2010	51	122	Merveilles naturelles de l'état de Washington
Automne 2010	52	109	Paranormal
Printemps 2011	45	96	Bassin minier du Nord-Pas-de-Calais

Les dernières éditions du Wikiconcours ont enregistré une centaine de participants. Ils représentent entre 10 et 15% des contributeurs les plus actifs (plus de 100 modifications par mois).

Sur les 10 thèmes récompensés, 3 concernent directement les États-Unis, 2 la France, 1 la Tunisie et 1 la Grande-Bretagne. La part importante de thèmes anglo-saxons peut s'expliquer de trois façons :

- l'influence de la culture américaine de façon générale incite les contributeurs à écrire sur ces thèmes ;
- la taille importante de la Wikipédia anglophone permet de rédiger plus facilement des articles par traduction ;

¹⁸ Chaque participant au premier Wikiconcours ne pouvait choisir qu'un seul article, et pas un thème.

– la plus forte utilisation des licences libres aux États-Unis permet d'illustrer plus facilement les articles, critère indispensable à l'obtention d'un article de qualité. En effet, tous les documents produits par l'administration américaine sont systématiquement placés dans le domaine public, permettant à n'importe qui de les réutiliser. Ainsi, les articles francophones de Léon Blum, Gaston Doumergue et Valéry Giscard d'Estaing sont illustrés par des photographies réalisées par l'administration américaine.

4 Multilinguisme/traductions

4.1 Niveau de compétences

Les contributeurs enregistrés disposent d'une page de présentation où ils peuvent indiquer leurs centres d'intérêts et leurs compétences, notamment en langues.

Ces informations, facultatives et relevant de l'auto-évaluation, sont à manier avec précaution.

Sur les 16 000 contributeurs ayant mentionné leur langue maternelle, 5 300, soit un tiers, indiquent une autre langue que le français. Ils représentent deux types de personnes :

- des contributeurs aux autres projets linguistiques de Wikipédia et qui interviennent ponctuellement pour effectuer des modifications techniques (par exemple pour rajouter des liens vers les articles qu'ils ont rédigé dans une autre langue);
- des contributeurs maîtrisant également le français et participant à la rédaction des articles en français.

12% des contributeurs indiquant une autre langue maternelle ont également le français comme langue maternelle. Ils maîtrisent généralement l'anglais (26%), l'espagnol (11%), l'arabe (10%), l'italien (8%) ou l'allemand (8%).

Le nombre de contributeurs indiquant son niveau dans une langue donnée, quel qu'il soit, est révélateur du niveau moyen des Wikipédiens dans cette langue et de l'importance qu'ils lui accordent. En effet, un utilisateur qui ne parlerait ni anglais ni khmer aura tendance à n'indiquer son niveau qu'en anglais, cette information lui paraissant plus importante pour les autres contributeurs.

Le classement des langues les plus renseignées est proche du classement des Wikipédia par taille. Seules différences notables : la présence du latin (49e Wikipédia, hors français) et de l'arabe (26e) dans le Top 10. Le polonais, 4e Wikipédia n'est que 12e langue renseignée.

Parmi les Wikipédiens ayant indiqué leur niveau d'anglais, 55 % le maîtrisent bien ou très bien¹⁹. Pour les autres langues, ce ratio chute aux alentours de 17%. Il est encore plus faible pour le latin (10%) et le japonais (9%). Ces derniers chiffres s'expliquent d'une part par le nombre élevé de contributeurs ayant étudié le latin sans le maîtriser parfaitement, et d'autre part par les amateurs de culture nipponne ayant commencé à apprendre le japonais. Les autres langues sont plus susceptibles d'être pratiquées régulièrement dans le cadre professionnel ou privé et donc mieux maîtrisées.

¹⁹ Niveau babel 3 à 5, hors langue maternelle.

Langue	Information sur le niveau de compétence		Part des utilisateurs maîtrisant bien la langue
	Rang	Nombre d'indications	
Anglais	1	13 885	55,0%
Allemand	2	5 204	16,5%
Espagnol	3	4 532	17,7%
Italien	4	2 046	17,3%
Latin	5	1 031	10,3%
Russe	6	997	17,5%
Néerlandais	7	930	17,5%
Japonais	8	871	9,4%
Portugais	9	709	21,0%
Arabe	10	477	19,3%

Classement des langues suivant le nombre de contributeurs indiquant son niveau de compétences

4.2 Traduction

Chaque version linguistique de Wikipédia est indépendante des autres. Ainsi, un même sujet peut être traité de manière totalement différente en français en anglais ou en tchèque. Il arrive cependant que des articles soit des traductions des pages écrites dans d'autres langues.

Origine des articles traduits

Sur la version francophone de Wikipédia, on estime ainsi à 5% la proportion d'articles issus de traductions. Ce nombre comprend aussi bien des articles traduits entièrement que des articles dont seuls quelques paragraphes sont issus de traductions. Il convient également de noter qu'un article traduit évolue de façon totalement indépendante de l'article d'origine : des parties peuvent être rajoutées, supprimées ou largement modifiées.

Sans surprise, la majorité des articles traduits proviennent de l'anglais. Le classement des langues d'origine des traductions est par ailleurs similaire au classement des Wikipédia par taille. Seules différences dans le Top 10 : le polonais et le japonais sont moins bien classés dans les traductions, ces langues étant moins bien maîtrisées par les francophones que l'anglais ou l'espagnol par exemple.

Langue	Articles traduits en français		Nombre d'articles de la Wikipédia	
	Rang	Nombre	Nombre	Rang ²⁰
Anglais	1	35 987	3 543 418	1
Allemand	2	4 629	1 185 247	2
Italien	3	3 512	768 688	3
Espagnol	4	3 024	694 264	6
Néerlandais	5	926	662 840	8
Russe	6	562	644 638	9

²⁰ Classement par nombre d'articles, une fois retiré le français.

Portugais	7	462	664 190	7
Catalan	8	171	301 755	12
Polonais	9	146	765 044	4
Basque	10	139	62 705	42

Outils de traduction

S'il existe quelques outils techniques d'aide à la traduction (mise en page, possibilité de visualiser facilement les différentes versions linguistiques), l'usage de traducteurs automatiques est très fortement déconseillé à cause de la mauvaise qualité de la traduction généralement constatée. Tout ajout réalisé à l'aide de tels outils est immédiatement supprimé, surtout s'il n'est pas retravaillé après l'utilisation de l'outil.

Projet traduction

Les traducteurs ont tout de même la possibilité de s'inscrire à un projet traduction où ils peuvent demander l'aide des autres Wikipédiens et signaler leurs travaux. Ils peuvent ainsi lister les pages qu'ils sont en train de traduire pour signaler des changements en profondeur dans l'article, demander de l'aide, et suivre l'avancement de la traduction. Ces signalements sont cependant souvent restreints aux travaux importants (traduction complète d'articles longs) et n'ont été utilisés que dans 9 % des traductions effectuées. Les Wikipédiens préfèrent ainsi travailler seuls sur leurs traductions.

5 Conclusion

Le français est l'une des langues majeures sur les projets Wikimedia. 3^{ème} langue sur Wikipédia, il s'agit également d'une des langues les plus utilisées dans les autres projets : Commons, Wiktionnaire, Wikisource, etc. ainsi que dans l'interface MediaWiki. Les thèmes concernant la France sont également fortement représentés dans l'encyclopédie.

Trois défis restent cependant à relever :

- soutenir les langues régionales, peu représentées sur internet ;
- développer la participation des contributeurs originaires des pays francophones du Sud ;
- faire face à l'utilisation massive de l'anglais sur internet et à la diffusion massive de la culture américaine, au détriment parfois des cultures locales.

Dans ce contexte, plusieurs pistes de réflexions nous paraissent indispensables :

- Il apparaît primordial de développer l'usage des licences libres en France pour permettre d'assurer à la fois l'usage du français dans le monde et une communication efficace sur la France, notamment en matière de tourisme.

- Les langues régionales sont peu présentes sur Internet, les fournisseurs de contenus préférant utiliser une langue comprise par le plus grand nombre. Les projets Wikimedia font donc exception en produisant une encyclopédie dans les langues régionales, mais aussi en réalisant des dictionnaires (Wiktionnaire) et en mettant à disposition des textes anciens (sur Wikisource). Afin de préserver ces langues, il apparaît nécessaire de développer ces projets et de favoriser la diffusion de documents librement réutilisables.

Tout financement public de documents en langue régionale devrait être conditionné à la mise sous licence libre afin d'ôter tout frein à leur diffusion.

– Différentes institutions prestigieuses françaises (BnF, Château de Versailles, Museum de Toulouse, etc.) et étrangères (British Museum, etc.) ont déjà réalisé des partenariats avec Wikimedia France et ses équivalents étrangers. Ces projets permettent au plus grand nombre d'avoir accès à leurs collections.

Ces initiatives sont à développer afin de permettre la diffusion de la culture française auprès de tous les Français mais aussi auprès des internautes du monde entier.

– Plusieurs pays étrangers (États-Unis, Russie, Brésil, etc.) ont déjà saisi l'intérêt de produire le maximum de documents sous licence libre. Cela leur permet à la fois de mieux informer leurs citoyens, de soutenir leurs entreprises et de permettre la diffusion de leurs cultures et points de vue dans le monde entier.

Pour maintenir l'influence de la culture et de la langue françaises dans le monde, il apparaît important de suivre ces exemples et de mettre systématiquement sous licence libre (ou dans le domaine public) les documents publics produits par l'administration française.

Dossier créé par Moyg, Floflo, Léna, (:Julien:), Trizek, Ludo29, Seb35 (pseudonymes de Wikipédia).

Diffusé sous licence CC-BY-SA.