

**III Symposium
international sur le
multilinguisme dans le
cyberespace**

**Présentation du
Wiktionnaire**

1) Le projet Wiktionnaire de Wikimedia

Le 12 décembre prochain le projet Wiktionnaire fêtera ses 10 ans. Depuis sa création deux ans après Wikipédia, il a rapidement grandi et existe maintenant dans plus de 150 langues, dont voici les plus importants :

Views/hr	Articles	Σ	All languages (140)
125,251	3,161,337	en	English
46,957	2,244,910	fr	French
24,959	456,489	ru	Russian
19,886	227,811	de	German
11,507	183,757	pt	Portuguese
9,711	301,529	pl	Polish
8,118	118,865	it	Italian
6,566	78,563	es	Spanish
5,803	1,236,688	zh	Chinese
5,327	84,838	ja	Japanese
5,119	235,697	nl	Dutch
4,266	405,003	el	Greek
4,007	326,283	sv	Swedish
3,628	230,627	vi	Vietnamese
3,381	311,564	tr	Turkish
3,184	348,035	ko	Korean
2,837	224,522	fi	Finnish
2,034	609,438	lt	Lithuanian
1,937	190,979	hu	Hungarian
1,852	17,144	bg	Bulgarian
1,704	2,095,408	mg	Malagasy
1,546	40,928	cs	Czech
1,337	121,942	id	Indonesian
1,267	128,617	no	Norwegian
1,223	207,171	io	Ido
1,101	11,911	he	Hebrew
944	49,154	ar	Arabic
903	260,964	ta	Tamil
880	107,516	ku	Kurdish

Source : <http://stats.wikimedia.org/wiktionary/EN/Sitemap.htm>

Chaque version est donc sensée publier à terme les mêmes articles que les autres, seule la langue des définitions et des pages de discussion change.

Comme ils jouissent d'un excellent référencement, on les retrouve au début des recherches web, même pour les mots étrangers. Par exemple dans Google.fr, le mot anglais « multilingual » retourne d'abord deux sites en anglais, suivi du Wiktionnaire en français.

2) Le Wiktionnaire francophone

2.1) Envergure

Le français représente plus de la moitié de ses articles avec 1,2 millions d'entrées. Mais en filtrant les flexions (pluriels et conjugaisons) nous avons inventorié 226 264 mots différents pour cette langue. Ils sont divisés essentiellement en 126 004 noms communs, 47 449 adjectifs, 31 123 noms propres, et 22 162 verbes.

Il est à noter que si le nombre de nouveaux noms communs s'essouffle, le potentiel d'évolution des noms propres reste encore important quand on pense à la façon de prononcer en français toutes les villes du monde.

Classement ↕	Langue ↕	Mots le 30 octobre ↕	Mots le 7 octobre ↕	Évolution ↕	Noms communs ↕	Noms propres ↕	Adjectifs ↕	Verbes ↕	Adverbes ↕	Lemmes (total flexions) ↕	Flexions ↕	Locutions ↕	Ébauches ↕
1	français	1198205	1195884	+2521	128096	31283	47585	22463	2669	227076	971129	26594	11690
2	russe	156936	156970	-34	2716	920	668	743	101	22392	134544	189	71
3	bulgare	156837	156837	0	10550	903	22	2764	221	17014	138623	35	34
4	anglais	120473	120398	+75	19555	1950	6406	5271	1232	33002	87471	5562	1398
5	slovène	80190	80190	0	4903	805	1720	379	116	6979	73311	461	41
6	latin	74755	74661	+94	8835	1188	3265	2365	482	15284	58471	115	51
7	italien	46419	45835	+584	11987	975	2807	2372	495	17276	29143	789	5245
8	espéranto	39800	39786	+14	8464	907	2243	2437	593	15465	24336	567	60
9	tchèque	31599	31580	+19	7149	734	1989	2559	449	13077	18522	197	10
10	malgache	24324	24324	0	4751	35	10	788	2	5594	18730	3	14
11	haut-sorabe	24189	24189	0	41	41	3	346	4	454	23736	0	4
12	néerlandais	23566	23491	+77	14743	952	3018	3575	700	23049	519	447	232
13	chinois	22976	22977	+1	1917	809	259	433	120	22976	2	65	20016
14	japonais	19607	19570	+37	4633	991	287	790	135	19599	8	213	12570
15	allemand	17728	17711	+17	10772	1307	1868	1653	326	16621	1107	370	547
16	islandais	13592	13591	+1	1228	433	220	197	34	5356	8236	87	76
17	espagnol	13521	13502	+19	5630	941	1187	1654	184	10288	3235	1052	250
18	vietnamien	13238	13236	+2	5221	386	2406	3990	342	13237	1	1074	97
19	indonésien	13018	13007	+11	5253	608	1070	3903	233	12230	788	1214	88
20	ido	12948	12948	0	8158	369	1539	2489	226	12873	75	9	48
21	ukrainien	12871	12869	+2	293	525	868	16	2	1757	11114	30	60
22	suédois	12412	12089	+323	4581	832	1244	1504	217	8707	3705	398	116
23	portugais	11393	11404	-11	5617	1157	1279	1953	175	11061	332	950	388
24	roumain	10923	10830	+93	1765	756	281	512	113	3321	7602	163	80
25	coréen (hanja)	8815	8814	+1	193	5	0	2	2	8815	0	1	0
26	polonais	7471	7468	+3	1936	847	303	455	140	4063	3408	244	151
27	catalan	6585	6569	+16	3097	761	708	1284	90	6448	137	394	136
28	finnois	6480	6479	+1	3267	807	521	742	217	5637	843	147	1139
29	ancien français	6420	6009	+411	3800	20	742	1687	180	6118	302	16	58

Source : <http://fr.wiktionary.org/wiki/Wiktionnaire:Statistiques>

2.2) Les contributions

Pour classer les articles dans chaque catégorie de langue, le Wiktionnaire utilise généralement des modèles catégorisant, dont les noms sont basés sur les codes langues ISO (disponibles sur www.SIL.org).

Cette syntaxe peut être saisie semi-automatiquement grâce à différents outils d'assistance à l'édition comme les patrons préremplis personnalisables. Toutefois pour pouvoir bénéficier des gadgets javascripts il faut créer un compte.

Chaque article créé est validé par un utilisateur de confiance sous 24 h, grâce au même système que Wikipédia « patrolled edits », avec pour différence qu'ils sont forcément élus. Il leur appartient de juger si le terme est rare, s'il fait partie d'un idiolecte, si c'est un hapax, ou encore un protologisme à supprimer.

Dans le doute tout le monde peut ouvrir la discussion dans les pages dédiées :

- 1) Proposer un mot.
- 2) Question sur les mots déjà créés.
- 3) Proposition de suppression par un vote.

Enfin, il existe aussi des élections des meilleurs articles dits « de qualité ».

3) Exploitation des données depuis l'extérieur

Toutes ces données sont accessibles depuis les autres sites par différents programmes, comme « Wiktionary pour Android » développé par la fondation, ou « Wiktionary and Google Translate » pour Firefox, qui permet d'afficher rapidement la définition d'un mot si on double-clique dessus.

Enfin, les sauvegardes quotidiennes sont téléchargeables en .xml sur <http://dumps.wikimedia.org/>.