

Wikidata et Wiktionnaire : retour sur un échec annoncé



WIKIDATA



Wiktionnaire
Le dictionnaire libre

CC by-sa 3.0,
<https://commons.wikimedia.org/wiki/File:WiktionaryFr.svg>

Plan

Rapide historique

Avis légal de la Fondation sur les données lexicographiques

Licence CC0 pour les données lexicographiques ?

Wikimedia Commons

Statistiques

Omegawiki ?

Rapide historique

- 22 mars 2004 : création du Wiktionnaire en français
- 30 octobre 2012 : création de Wikidata
- 22 février 2018 : il est demandé à la communauté Wikidata d'approuver la licence CC0 pour les données lexicographiques.
- 23 mai 2018 : les données lexicographiques sont activées sur Wikidata (seuls les lexèmes et les formes sont disponibles)
- 18 octobre 2018 : les sens sont activés

Avant l'arrivée des données lexicographiques

Wikimania 2016 : rencontre entre des contributeurs au Wiktionnaire et des développeurs Wikidata



« Il y avait de leur part (dév. Wikidata) une véritable motivation à **faire évoluer techniquement** les Wiktionnaires et à enfin faire avancer la discussion avec les communautés de contributeurs, qui n'ont pas été particulièrement attirés par la question jusqu'ici. »

Création de [wikt:Projet:Coopération/Wikidata](#)

Avant l'arrivée des données lexicographiques

De nombreuses propositions de modèles de données

Date de début	Auteur(s) principal(aux)
2013-02	JAn Dudík (cswikt), This, that and the other (enwikt), Darkdadaah (frwikt)
2013-06	Denny (WMDE)
2013-07	Micru (ca, wikidata), Francis Tyers (iswikt)
2013-08	Denny (WMDE)
2013-09	Ivadon (dewikt)
2013-10	Bigbossfarin (enwikt)
2014-10	GPHemsley (enwiki, wikidata)
2015-05	Denny (WMDE)

Toutes les propositions ont été traduites en français (sauf 2013-09, en allemand) pour pouvoir en débattre plus facilement (~63 ko de texte)

Avis légal de la Fondation Wikimedia (17/02/2018)

m:Wikilegal/Lexicographical_Data/fr

Informations lexicographiques **non protégées** par le droit d'auteur :

- Lemmes (le mot lui-même)
- Définitions de jargon ou de mots avec des expressions figées
- Prononciations
- Informations grammaticales (masculin/féminin, nom/verbe/..., ...)
- Collocations et expressions figées (exemple : « lourd silence », « pluie battante » ne sont pas protégeables)

Avis légal de la Fondation Wikimedia (17/02/2018)

[m:Wikilegal/Lexicographical_Data/fr](https://m.wikilegal/lexicographical_data/fr)

Informations lexicographiques **protégées** par le droit d'auteur :

- Microstructure et macrostructure (l'organisation du dictionnaire).
Est-ce qu'importer toutes les pages de catégories du Wiktionnaire (élément de la macrostructure) en tant qu'élément dans Wikidata est autorisé ?
- Définitions
- Étymologies
- Informations pragmatiques (façon dont le mot est utilisée, s'il est archaïque, etc)
- Informations encyclopédiques
- Exemples de phrases (dans le cas d'exemples originaux)

Avis légal de la Fondation Wikimedia (17/02/2018)

[m:Talk:Wikilegal/Lexicographical Data](https://m:talk.wikilegal/lexicographical_data)

Questions ouvertes. Quid de ?

- Choix d'une illustration associée à une définition
- Liste de tous les mots d'un dictionnaire
- Thésaurus



Un prout

Licence CC0 pour les données lexicographiques (22/02/2018)

[Wikidata:Project_chat#Adding_the_Lexeme_namespace_to_the_licensing_footer_text](#)

Discussions vives entre les partisans de la licence CC0 permettant une réutilisation la plus large possible et des contributeurs aux Wiktionnaires (principalement) qui voient Wikidata comme un projet concurrent qui va se faire en parallèle des Wiktionnaires

La discussion apparaît comme un vote alors qu'il est très probable que les développeurs de Wikidata aient prévus d'appliquer la licence CC0 quel que soit le résultat

Licence CC0 pour les données lexicographiques (22/02/2018)

[Wikidata:Project_chat#Adding_the_Lexeme_namespace_to_the_licensing_footer_text](#)

Je considère que vous faites un fork
du Wiktionnaire dans Wikidata
avec votre propre calendrier.

Noé

Licence CC0 pour les données structurées

[Wikidata:Project_chat#Adding_the](#)

Je considère que vous faites partie du Wiktionnaire dans Wikidata, et non pas avec votre propre calendrier.

Noé

Ce projet a commencé sous le nom de « Données structurées pour le Wiktionnaire ». Quelque part en chemin, le but a changé, et il est devenu « Wikidata pour les données lexicographiques ».

Je pense que ce changement était une erreur. Je ne pense pas qu'il puisse y avoir deux projets lexicographiques rivaux chez Wikimedia.

Jheald

Licence CC0 pour les données structurées

[Wikidata:Project_chat#Adding_the](#)

Ce projet a commencé sous le nom de « Données structurées pour le Wiktionnaire ». Quelque part en chemin, le but a changé, et il est devenu « Wikidata pour les données lexicographiques ».

Je considère que vous êtes le début sur le Wiktionnaire : que Wikidata commence à traiter les données lexicographiques sans même se soucier de consulter les personnes qui créent et gèrent déjà des données lexicographiques sur Wikimedia chaque jour. Compte tenu de la situation des licences et du manque flagrant de communication, deux projets parallèles vont travailler sur les mêmes problèmes, mais séparément.

C'est exactement ce que nous craignons depuis le début sur le Wiktionnaire : que Wikidata commence à traiter les données lexicographiques sans même se soucier de consulter les personnes qui créent et gèrent déjà des données lexicographiques sur Wikimedia chaque jour. Compte tenu de la situation des licences et du manque flagrant de communication, deux projets parallèles vont travailler sur les mêmes problèmes, mais séparément.

Metaknowledge

Licence CC0 pour les données lexicographiques (22/02/2018)

Au bout d'une semaine de discussions, strakhov a fait les comptes

Pour la licence CC0 : 26 personnes (~5910 modifications sur les Wiktionnaires, dont 5055 par un seul contributeur)

Contre : 9 personnes (~158 027 modifications sur les Wiktionnaires)

Abstention : 3 personnes (~127 000 modifications sur les Wiktionnaires)

Licence CC0

géographiques

Au

Con

Abstention

Je voulais juste dire que le seul contributeur significatif dans Wiktionary soutenant cette proposition est apparemment VIGNERON. Je ne vois pas en quoi l'antagonisme d'un projet entier est bon pour nous, même s'il attire d'autres personnes de l'extérieur. Je ne m'oppose pas à la « structuration » de Wiktionary, car je trouve le travail effectué là aussi assez inefficace, mais j'essaierais d'impliquer ces communautés au lieu de les contourner ici par la force brute. Leur demander et leur donner ce dont ils ont besoin. Si une partie importante des communautés Wiktionary pense que leur travail est plagié ou utilise une mauvaise licence par cette approche des lexèmes de Wikidata et de la licence CC0, il est peut-être temps de repenser la proposition

strakhov

(Wiktionnaires)

Intermède : comment ça se passe avec Commons ?

[c:Commons:Structured_data](#)



2017-2019 : projet « Données structurées » sur Wikimedia Commons

Stockage des (méta)données relatives aux fichiers multimédias de manière structurée

Le *backend* de Commons est migré sur Wikibase

Intermède : comment ça se passe avec Commons ?

[c:Commons:Structured_data](#)



2017-2019 : projet « Données structurées » sur Wikimedia Commons

Stockage des (méta)données relatives aux fichiers multimédias de manière structurée

Le *backend* de Commons est migré sur Wikibase

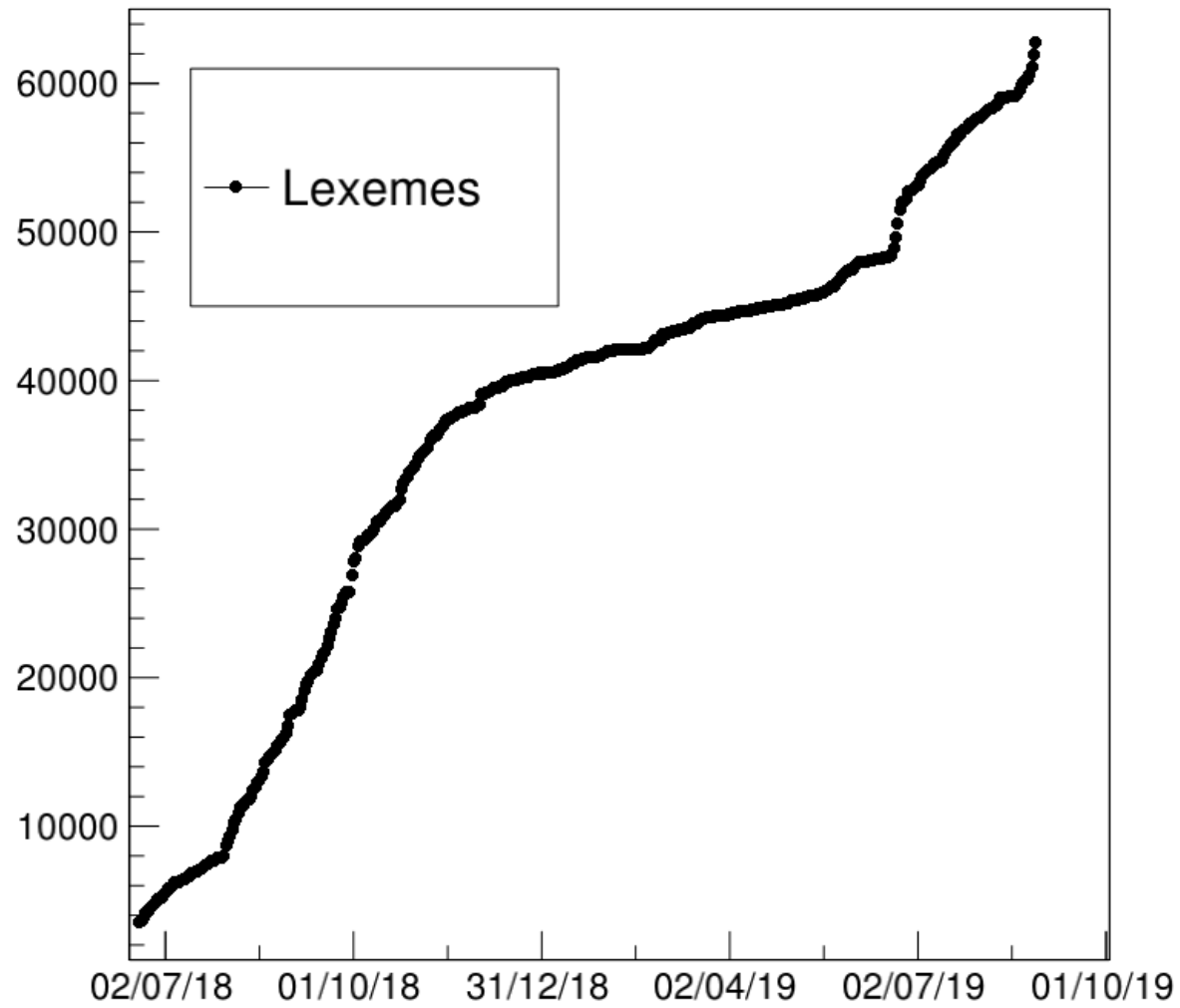
Collaboration entre les développeurs de Wikidata et la communauté de Commons
Pourquoi pas aussi avec le Wiktionnaire ?

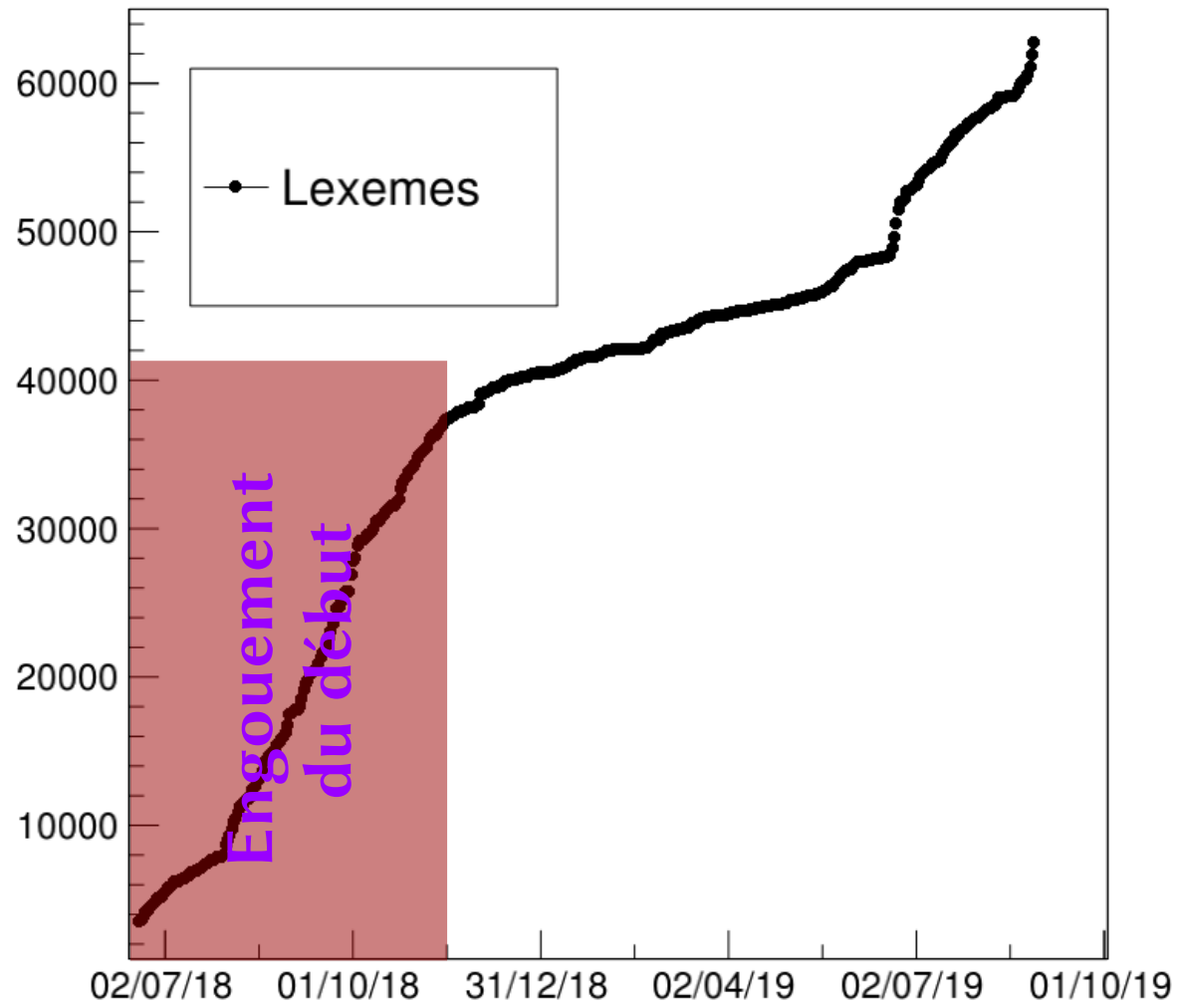
Finalement, les lexèmes arrivent

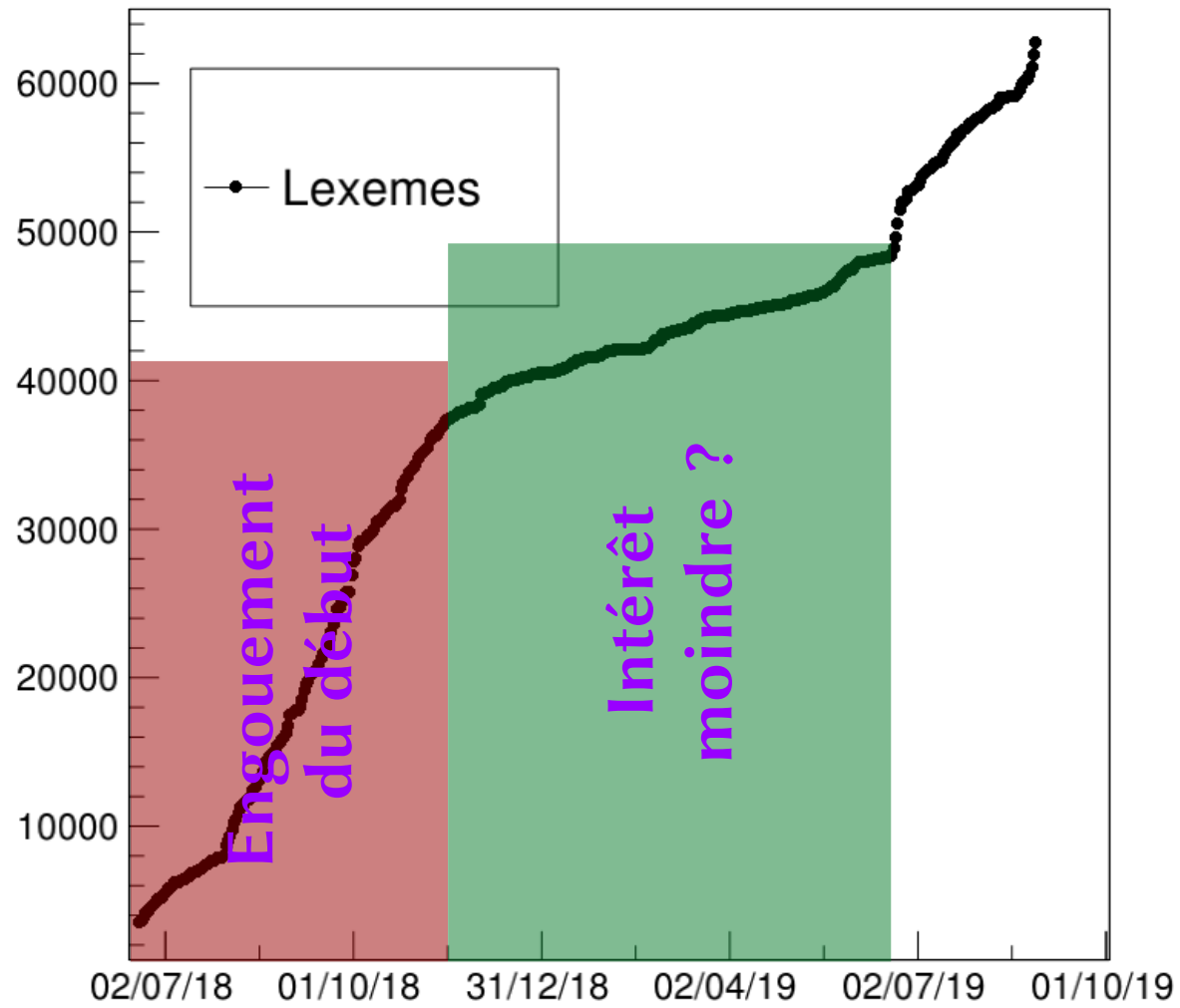
22 mai 2018 : activation des lexèmes et des formes
il est demandé de ne pas importer de données des
Wiktionnaires

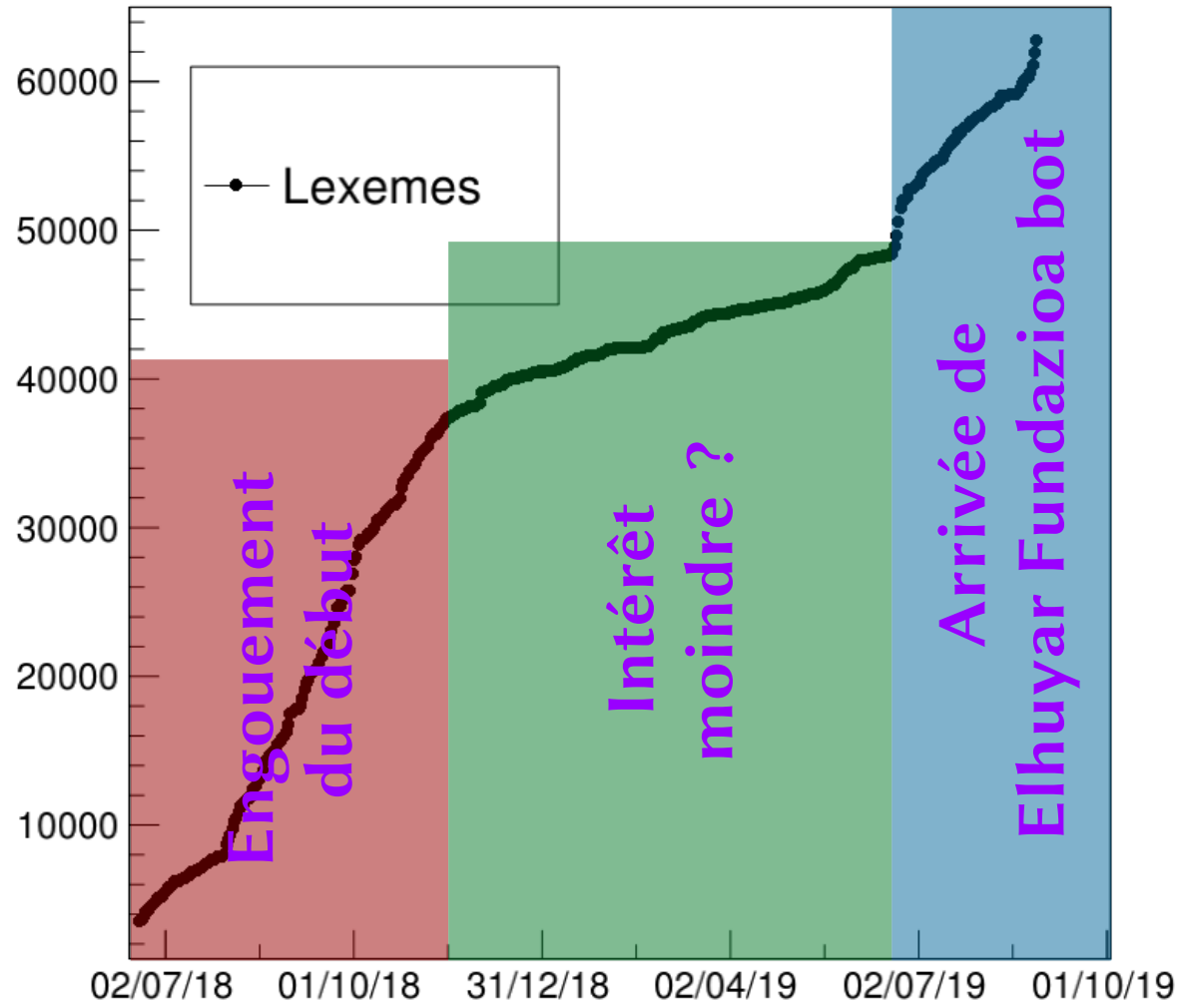
22 septembre : l'import de données par bot est autorisé par
l'équipe Wikidata

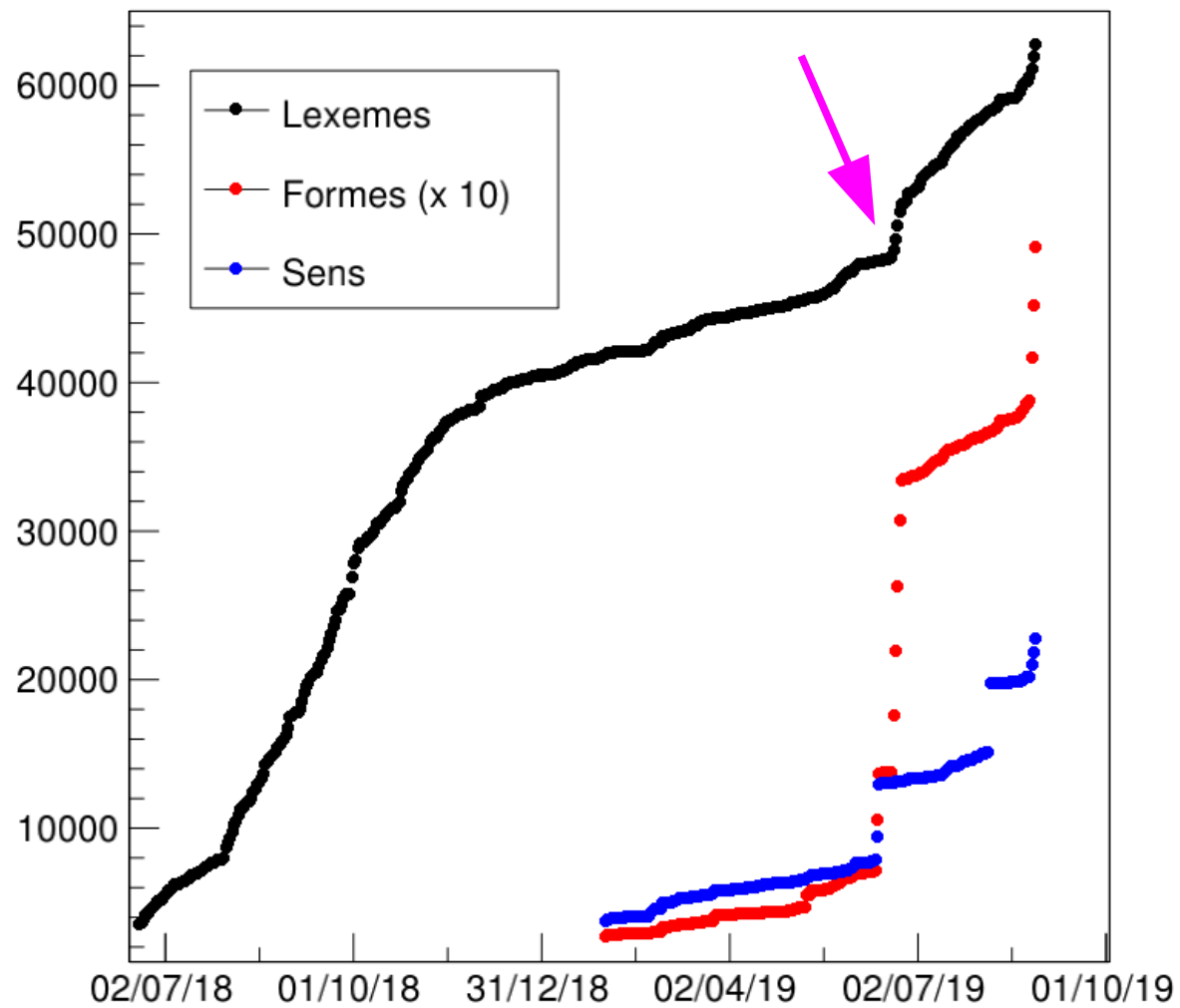
18 octobre : les sens sont activés











Des personnes des Wiktionnaires ?

Au 31 août :

- **80** utilisateurs ont fait **plus de 100 contributions**
 - **~20 %** de contributeurs au Wiktionnaire
- **22** ont fait **plus de 1000 contributions** (dont 3 bots)
- Parmi les **20 plus gros contributeurs** : **8** ont contribué significativement (>1000 edits) à un Wiktionnaire (KaMan (pl), Vesihisi (sv), Njardarlogar (en, nn), VIGNERON (fr, br), Infovarius (ru), Thamizhpparithi Maari (ta), Rua (en), -xfi- (cs))
- Et **5** ont contribué **ces 6 derniers mois**

Wikidata : un Omegawiki bis ?

Omegawiki avait le but de fournir un dépôt central de données lexicographiques. Ce projet est objectivement un échec.

- Trop technique : utile pour ajouter/réutiliser des informations purement techniques (déclinaisons, conjugaison, ...)
- Attention aux discussions. Une seule langue (anglais) (ou presque) conduit à une vision anglo-centrée et exclut *de facto* des contributeurs. Comment assurer et encoder la diversité linguistique ?

Conclusion ?

Wikidata est vu comme un projet concurrent aux Wiktionnaires

Wikidata ne profite pas de l'expérience des contributeurs aux Wiktionnaires

Contributions isolées sur les données lexicographiques – pas encore de communauté structurée sur Wikidata

Mon avis : Wikidata sera rempli de données purement grammaticale (déclinaison, conjugaison, ...) qui pourront être réutilisées par les Wiktionnaire. Je ne crois pas au développement des sens.



Merci pour votre attention

- Questions ?
- Commentaires ?
- Discussion ?