Réguler Wikipédia

Socioinformatique d’un commun du savoir

Léo Joubert

Dysolab, Université de Rouen Normandie · projet ANR COMSAV

Avignon, 2026

Plan

1. Contexte
   une tension au cœur de l'encyclopédie, trois questions pour COMSAV
2. Premiers résultats
   1. activités
   2. reconnaissance
   3. parcours
3. Le matériau
   rejouer l'historique : l'archive, le graphe, les épisodes
4. Enquête et mesures
   archives des règles, pages de discussion, entretiens
5. Conclusion
   un chantier ouvert : l'équipe, les données, la suite

1. Contexte : Une tension au cœur de l'encyclopédie

Une encyclopédie de référence
parmi les sites les plus consultés au monde, citée, enseignée, interrogée par les moteurs de recherche et les IA.

… produite par un processus fragile
l'agrégation de contributions individuelles : chacun peut ajouter, reformuler… ou effacer ce qu'un autre a écrit.

Wikipédia en français, 2001–2026

2,7 M
articles

1,3 M
comptes contributeurs

153 M
versions d'articles

7,3 M
adresses IP

9 %
des versions annulées ou annulantes

1 sur 3
fragments de texte finalement supprimés

Comment un tel collectif **se régule**-t-il ?

1. Contexte : COMSAV : trois questions

Projet ANR JCJC sur la régulation de la Wikipédia francophone.

1 Activités

Comment les **règles éditoriales** s'imposent-elles ?

Qui fait quoi, sur quelle page — et qui en est tenu à l'écart ?

2 Reconnaissance

Comment les contributeurs sont-ils **reconnus ou marginalisés** ?

Quelles places, quels rôles, quels rejets ?

3 Parcours

Comment évoluent les **trajectoires** au sein de l'encyclopédie ?

Quelles ressources pour entrer, pour durer ?

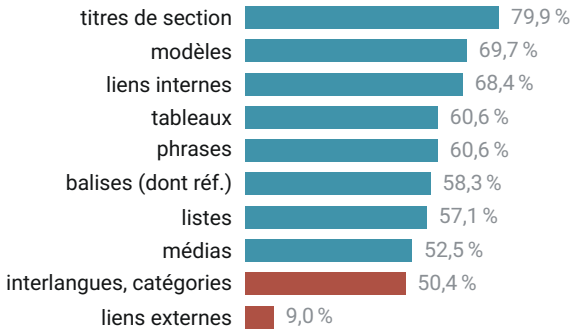
Décrire (Canguilhem) : la régulation comme *ajustement* d'éléments hétérogènes.

Évaluer (Fraser) : la participation fait-elle des contributeurs un ensemble de pairs ?

2. Premiers résultats : Les règles laissent des traces

Part des fragments de texte (tokens) encore présents en avril 2026, par type

moyenne : 65,6 %



Liens externes : 9 %

Presque tous supprimés : la recommandation de ne pas placer de liens externes dans le corps des articles est appliquée, geste après geste.

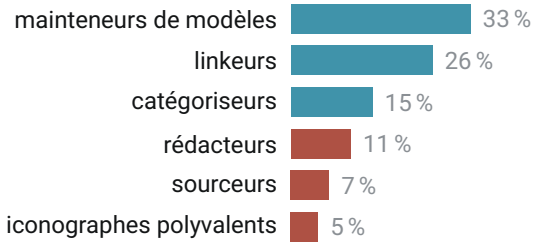
Interlangues : 50 %

Probablement la trace d'une décision collective : le transfert des liens interlangues vers Wikidata en 2013.

2. Premiers résultats : La division du travail en six métiers

Profils des contributeurs à l'échelle de la page

1 000 pages très travaillées, 9 types de contributions → 6 métiers



Un flux de rédaction

Les métiers interviennent à des moments différents de la vie d'une page : on rédige, on source, on lie, on catégorise, on maintient.

Le travail de l'un est **repris** par les autres, comme dans une chaîne éditoriale.

Un profil sur dix seulement est d'abord rédacteur : l'essentiel du travail porte sur la couche de marquage.

2. Premiers résultats : Reconnaissance

1 Activités

Comment les **règles éditoriales** s'imposent-elles ?

Qui fait quoi, sur quelle page — et qui en est tenu à l'écart ?

2 Reconnaissance

Comment les contributeurs sont-ils **reconnus ou marginalisés** ?

Quelles places, quels rôles, quels rejets ?

3 Parcours

Comment évoluent les **trajectoires** au sein de l'encyclopédie ?

Quelles ressources pour entrer, pour durer ?

> **Question de parité** (Fraser) : les nouveaux venus sont-ils traités en pairs ? Quelles manières de contribuer sont reconnues ?

2. Premiers résultats : Trois régimes de production des pages

1. **Maintenance sans construction** (28 %) : éditions mineures, éditeurs de passage.
2. **Construction spécialisée** (48 %) : un petit noyau écrit et révise, peu de conflit.
3. **Effervescence collaborative** (24 %) : aucun éditeur ne domine, beaucoup d'anonymes, beaucoup d'annulations.

Régime de production selon le label de qualité

	1	2	3
Ébauche	**94 %**	5 %	0 %
Bon début	**99 %**	1 %	0 %
Avancement B	24 %	**59 %**	17 %
Avancement A	5 %	47 %	**48 %**
Bon article	3 %	**73 %**	24 %
Article de qualité	2 %	**62 %**	36 %

Unité des ébauches, pluralité des qualités : les labels reconnaissent au moins deux manières de bien écrire un article — l'œuvre d'un petit noyau, ou la délibération d'une foule.

2. Premiers résultats : Le revert comme signal social

Probabilité qu'une contribution soit annulée (modèle sur ≈ 60 000 contributions, AUC = 0,90)

Ce qui compte	Rejet
Être **nouveau** dans la communauté	↗ **très fort**
Avoir **déjà été annulé**	↗ **très fort**
Contribuer à une page **déjà aboutie**	↗ fort
Respecter les **règles** wikipédiennes	↘ fort
Être administrateur, être anonyme	faible

En une phrase

À conformité aux règles égale, on est d'autant plus rejeté qu'on est nouveau ou qu'on a déjà été rejeté.

Le revert n'est pas seulement un contrôle qualité : c'est un **signal social**.

Joubert, WikiWorkshop. Ambivalence du revert : sanction *ou* amélioration — aucune métrique ne tranche seule.

2. Premiers résultats : Parcours

1 Activités

Comment les **règles éditoriales** s'imposent-elles ?

Qui fait quoi, sur quelle page — et qui en est tenu à l'écart ?

2 Reconnaissance

Comment les contributeurs sont-ils **reconnus ou marginalisés** ?

Quelles places, quels rôles, quels rejets ?

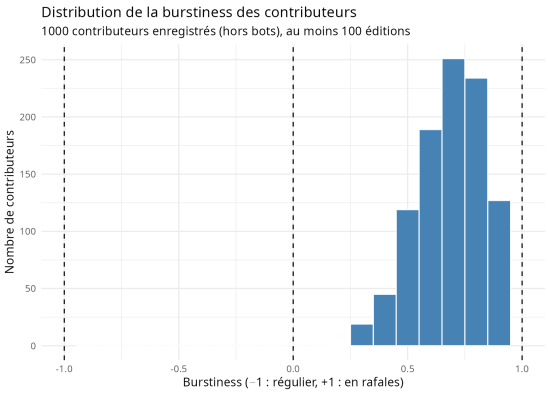
3 Parcours

Comment évoluent les **trajectoires** au sein de l'encyclopédie ?

Quelles ressources pour entrer, pour durer ?

> **Question de parité** (Fraser) : quelles vies permettent de contribuer, et de durer ?

2. Premiers résultats : Contribuer en rafales



1. **Indicateur**
   la *burstiness* mesure si l'on édite régulièrement (-1) ou en rafales ($+1$) : presque tout le monde contribue en rafales.
2. **Enquête**
   les rafales correspondent à des périodes où contribuer est *compatible avec la vie quotidienne*.
3. **Nouvel indicateur**
   distinguer les rafales entre elles : par leur contenu, leur page, leur issue.

3. Le matériau : Ce que l'archive enregistre

Ce que le lecteur voit

Le **féminisme** est un ensemble de mouvements et d'idées politiques, sociales et culturelles.
La définition globale de ce terme lui donne pour objectif de promouvoir l'égalité entre les femmes et les hommes en militant pour les droits des femmes[1,2].

ce que l'archive enregistre

Le wikitext : deux couches entremêlées

```
Le '''féminisme''' est un ensemble de [[Mouvement social|mouvements]] et
d'[[doctrine|idées]] politiques, sociales et culturelles. La définition
globale de ce terme lui donne pour objectif de promouvoir
l'[[égalité des sexes|égalité]] entre les femmes et les hommes en
militant pour les [[droits des femmes]]<ref>{{Ouvrage
|auteur1=Angélique Gérard |titre=Pour la fin du sexisme !
|année=2019 |passage=57 |lire en ligne=https://books.google.fr/…}}</ref>,
et ce, sur le principe fondamental que les hommes et les femmes…
```

Couche 1 — le texte : les mots que lit le lecteur

Couche 2 — le marquage : liens [[…]] modèles {{…}} références <ref> style '''…'''

Extrait réel de l'introduction de la page « Féminisme » (fr.wikipedia.org), abrégé.

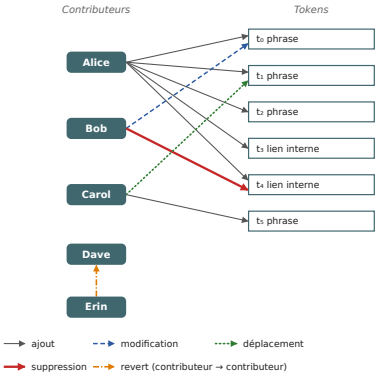
3. Le matériau : Le graphe contributeurs--tokens

Pour chaque article

Un graphe qui relie des **contributeurs** à des **fragments de texte** (tokens) :

- une arête par geste sur un token, **typée** (ajout, modification, déplacement, suppression) et **datée** ;
- une arête revert quand une version en annule une autre.

> La suite des arêtes d'un token raconte sa vie : *ajouté par l'une, reformulé par un deuxième, supprimé par une troisième*.



3. Le matériau : Du graphe aux épisodes

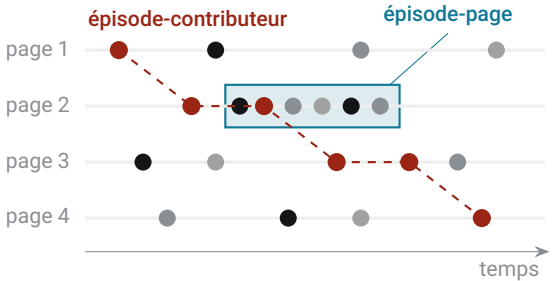


Schéma illustratif : un point = une contribution.

Question ouverte : qu'est-ce qui rend un épisode saillant ? Une rafale, une suppression subie, une page en conflit, un long silence…

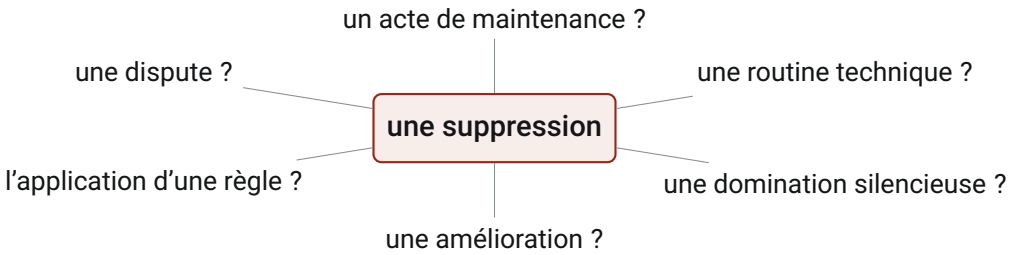
Épisode-page

On suit **une page** à travers ceux qui l'écrivent : plusieurs récits croisés d'un même moment.

Épisode-contributeur

On suit **une personne** d'une page à l'autre : du pseudonyme aux moments saillants d'un parcours.

4. Enquête et mesures : Que mesure-t-on quand on mesure une suppression ?



L'informatique donne l'échelle : des millions d'événements, mais des indicateurs *ambivalents*.

La sociologie donne le sens : ce que les acteurs font et disent, mais *sans prise* sur des collectifs gigantesques.

4. Enquête et mesures : Trois sources pour interpréter le graphe

1 Archives des règles

L'histoire des **recommandations** : quand, contre quoi et par qui une règle a-t-elle été écrite ?

Les règles sont elles-mêmes des pages, avec leur historique.

2 Pages de discussion

Ce que les contributeurs **se disent** pendant qu'ils écrivent : arguments, invocations des règles, menaces, compromis.

ex. controverses de santé : vaccination, allaitement.

3 Entretiens

Explicitation sur épisodes : on retrouve l'enquêté·e dans les traces et on lui fait raconter quelques épisodes saillants.

Une fiche par enquêté·e, lisible sans savoir coder.

> Le graphe sert alors à **choisir** quoi lire : quels contributeurs, quelles pages, quels épisodes.

4. Enquête et mesures : La neutralité, un résultat en construction

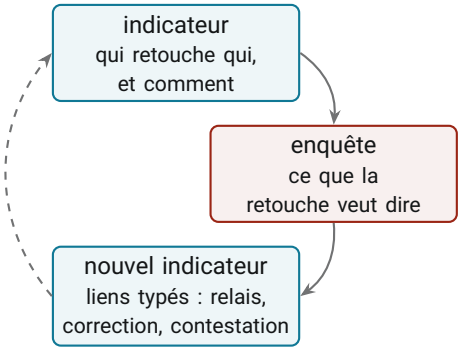
Durer : ce que mesure le questionnaire

Enquête auprès de ≈ 8 000 contributeurs (8 langues) : l'**adhésion aux principes** de Wikipédia prédit la contribution régulière.

Mais comment cette adhésion se construit-elle ?

Apprendre : ce que montrent les débats

Dans les controverses, la **neutralité de point de vue** s'apprend, se performe et se dispute : elle devient le baromètre avec lequel on juge les retouches des autres.



La règle est structurellement incomplète : il n'existe pas de règle sur la manière d'appliquer la règle. C'est là que se logent apprentissage, jeu et conflit.

5. Conclusion : Un chantier ouvert

L'équipe

Sociologie (Dysolab, Rouen), informatique (IMT Atlantique, Avignon), économie.

2 post-doctorats, 1 ingénieur d'études, 1 post-doctorat à venir.

Les données

La base WPTokens et le package R `comsav` sont en place.

Pour proposer un projet sur la base : rejoindre l'équipe.

La suite

Guide d'entretien (hiver 2026), entretiens (2027), typage des liens, extension à d'autres communs.

Merci !

leo.joubert@univ-rouen.fr

COMSAV (ANR JCJC) · Data2Laws et CONGRATS (PEPR eNSEMBLE)