

Etude des discussions Wikipédia les schémas d'interaction

Lydia-Mai Ho-Dac (CLLE, UT2J)
Céline Poudat (BCL, Univ. de Nice)
Ludovic Tanguy (CLLE, UT2J)



Plan

1. Introduction: les discussions Wikipédia
2. Vue d'ensemble des données : le corpus EFGwikiCorpus

Quelles approches pour analyser les données, décrire les interactions ?

3. Approche Top-down : Projeter des catégories sur les données
4. Approche Bottom-Up : Faire émerger des profils depuis les données

Les schémas d'interaction, zoom sur :

- 4.1. Monologues
- 4.2. Irruption d'un tiers dans un dialogue en ligne

Conclusion

1. Introduction - Les discussions Wikipédia

Wikipedia

— — —

- Une vingtaine d’années d’existence et un succès concluant (7^e site le plus visité dans le monde et en France – <https://www.similarweb.com/fr/top-websites/> en 2024)
- Des données langagières
 - massives
 - by CC BY-SA 4.0
 - multilingues
- Une source d’information au mode de fonctionnement “wikicratique” inédit
- Un objet web tentaculaire et hétérogène
 - Multitude de pages relevant de différents genres du discours : articles encyclopédiques, pages de discussion, pages d’entre-aide, etc.
 - Dimension génétique avec les historiques des révisions
 - Dimension multilingue avec les liens entre articles et catégories

Wikipedia : 207 000 000 pages

<https://stats.wikimedia.org>

☰ Wikipédia

🌐 303 langues ▾

Article Discussion

Lire Voir le texte source Voir l'historique Outils ▾



 Pour les articles homonymes, voir *WP* et *Wikipedia (homonymie)*.

Wikipédia  [Écouter](#) est une [encyclopédie en ligne collaborative](#) et [multilingue](#) créée par [Jimmy Wales](#) et [Larry Sanger](#) le 15 janvier 2001. Il s'agit d'une [œuvre libre](#), c'est-à-dire que chacun est libre de l'amender et de la rediffuser. Gérée en [wiki](#) dans le [site web](#) *wikipedia.org* grâce au logiciel [MediaWiki](#), elle permet à tous les internautes d'écrire et de modifier des articles, ce qui lui vaut d'être qualifiée d'« [encyclopédie participative](#) ». En quelques années, elle est devenue la plus importante encyclopédie disponible dans le monde et la plus consultée.

La [Fondation Wikimedia](#), [organisation à but non lucratif](#) américaine sise à [San Francisco](#), est dépositaire de la marque *Wikipedia*. Elle finance l'[hébergement web](#) de l'encyclopédie.

Wikipédia



Le [logo de Wikipédia](#) représente un globe

Les pages de discussion : les coulisses des articles

☰ Discussion:Wikipédia

🇫🇷 ▼

[Article](#) [Discussion](#)

[Lire](#) [Voir le texte source](#) [Voir l'historique](#) [Outils](#) ▼

Question à deux balles

Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --[Milegue \(discuter\)](#) 27 août 2014 à 16:46 (CEST) [[répondre](#)]

Personnellement : « la » Wikipédia mais « le » projet Wikipédia. [Like tears in rain](#) {--} 27 août 2014 à 16:58 (CEST) [[répondre](#)]

wikipédia est à la fois *une* encyclopédie, et *un* des projets-frères, so... up to u ;-) J'ai opté pour *la* wikipédia francophone --[Biblorock \(discuter\)](#) 27 août 2014 à 17:55 (CEST) [[répondre](#)]

J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --[Catarella \(discuter\)](#) 27 août 2015 à 09:46 (CEST) [[répondre](#)]

le signale que les Allemands disent « die Wikipedia ». --[Pierrot Lunaire \(discuter\)](#) 14 septembre 2015 à 19:25

Les pages de discussion : des CMR (Communications Médiées par les Réseaux – *CMC Computer Mediated Communication*)

— — —

<https://cmc-corpora.org/>

- Communications entre humains réalisées "en réseau" via des "machines" (ordinateurs & smartphones) e.g. clavardages, SMS, forum de discussion, Twitter, Whatsapp, Facebook, commentaires Youtube, etc.
- Une quantité exponentielle de données produites
- De nouvelles formes de communication, nouveaux usages : variantes (ortho)graphiques, écrit spontané, audios enregistrés, conversations sans face à face, discussions asynchrones
- CMC : interactions structurées en fils et messages (posts)
- Challenge : décrire ces nouvelles formes d'interaction
- quels traitements automatiques pour ces données "non standards" ?
 - comment la connaissance est partagée ?
 - comment les échangent s'organisent ?
 - comment les discours se construisent ?

Les pages de discussion : des CMC pour co-écrire une encyclopédie

☰ Discussion:Wikipédia

🇫🇷 ▼

[Article](#) [Discussion](#)

[Lire](#) [Voir le texte source](#) [Voir l'historique](#) [Outils](#) ▼

Question à deux balles

Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --[Milegue \(discuter\)](#) 27 août 2014 à 16:46 (CEST) [[répondre](#)]

Personnellement : « la » Wikipédia mais « le » projet Wikipédia. [Like tears in rain](#) {--} 27 août 2014 à 16:58 (CEST) [[répondre](#)]

wikipédia est à la fois *une* encyclopédie, et *un* des projets-frères, so... up to u ;-) J'ai opté pour *la* wikipédia francophone --[Biblorock \(discuter\)](#) 27 août 2014 à 17:55 (CEST) [[répondre](#)]

J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --[Catarella \(discuter\)](#) 27 août 2015 à 09:46 (CEST) [[répondre](#)]

le signale que les Allemands disent « die Wikipedia ». --[Pierrot Lunaire \(discuter\)](#) 14 septembre 2015 à 19:25

Les pages de discussion : des CMC pour co-écrire une encyclopédie

☰ Discussion:Wikipédia

🇫🇷 ▼

[Article](#) [Discussion](#)

[Lire](#) [Voir le texte source](#) [Voir l'historique](#) [Outils](#) ▼

Sommaire

Début

[article de qualité](#)

[Proposition d'article de qualité refusée le 19 février 2014](#)

[Question à deux balles](#)

[Style reverté ce jour](#)

[Coût](#)

[Explication des abréviations](#)

[Critères d'admissibilité, notoriété & bistrot](#)

[Pourquoi ?](#)

n ? Ce n'est pas écrit dans l'article. Doit-on dire le beau
mais je vous laisse parler d'abord. [Milogue \(discuter\)](#) 27 août

Chaque contributeur peut ouvrir un nouveau fil si le sujet lui semble nouveau en rédigeant un titre + un premier message.

Mais :

- certains fils peuvent parfois donner lieu à des sous-sections
- certains fils font suite à un ancien fil

Wikipédia:Prise de décision/Préparation de la page d'accueil (CEST)

objets-frères so... up to u ;-) J'ai opté pour *la* wikipédia
15:55 (CEST) [répondre](#)

un changement dans l'article. [Catalella \(discuter\)](#) 27 août 2015

edia ». --[Pierrot Lunaire \(discuter\)](#) 14 septembre 2015 à 19:25

Les pages de discussion : une structure en Fils, Posts & TimeStamn

☰ Discussion:Wikipédia

文A ▾

Article Discussion

Titre de fil, auteur A

Voir le texte source Voir l'historique Outils ▾

Question à deux balles

1er message, daté et signé par A

Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --Milegule (discuter) 27 août 2014 à 16:46 (CEST) [répondre]

2e message (indentation), daté et signé par B

Personnellement : « la » Wikipédia mais « le » projet Wikipédia. Like tears in rain {-} 27 août 2014 à 16:58 (CEST) [répondre]

3e message (indentation+), daté et signé par C

wikipédia est à la fois *une* encyclopédie, et *un* des projets-frères, so... up to u ;-) J'ai opté pour *la* wikipédia francophone --Bibliorock (discuter) 27 août 2014 à 17:55 (CEST) [répondre]

J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --Catarella (discuter) 27 août 2015 à 09:46 (CEST) [répondre]

dernier message (6e), daté et signé par ... E

le signale que les Allemands disent « die Wikipedia ». --Pierrot Lunaire (discuter) 14 septembre 2015 à 19:25

Les pages de discussion : une structure en Fils, Posts & TimeStamm

Discussion:Wikipédia

文A ▾

[Article](#) [Discussion](#)

[Lire](#) [Voir le texte source](#) [Voir l'historique](#) [Outils](#) ▾

Question à deux balles

Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --Milegue (discuter) 27 août 2014 à 16:46 (CEST) [répondre]

- **27/08/2014, 16h46, Milegue : Bonjour...**
- **27/08/2014, 16h58, Like tears in rain : Personnellement...»**
(1 heure plus tard)
- **27/08/2014, 17h55, Bibliorock : wikipédia est.. »**
(1 an plus tard)
- **27/08/2015, 09h46, Catarella : J'ai toujours..u ;-)** J'ai opté pour la wikipédia
(15 jours plus tard)
- **14/09/2015, 19h25, Pierrot Lunaire : Je signale...**
(depuis, plus rien)

Personnellement : « la » Wikipédia mais « le » projet Wikipédia. Like tears in rain {-} 27 août 2014 à 16:58 (CEST) [répondre]

wikipédia est à la fois une encyclopédie et un projet. --Bibliorock (discuter) 27 août 2014 à 17:55 (CEST) [répondre]

J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --Catarella (discuter) 27 août 2015 à 09:46 (CEST) [répondre]

le signale que les Allemands disent « die Wikipedia ». --Pierrot Lunaire (discuter) 14 septembre 2015 à 19:25

Les pages de discussion WP : premiers éléments

— — —

- Tous les articles ne nécessitent pas une page de discussion : moins de 15% des articles WP donnent lieu à une page “avec” discussion
- La page de discussion répond à un besoin originel et est complétée et parfois remplacée par :
 - Commentaires associés aux edits sur l’article
 - Page contributeur
 - Bistrot
- La discussion n’est pas un genre autonome: elle se réfère à l’article dans son état courant (dimension métadiscursive de discours sur le discours)
- La discussion est soumise à un ensemble de normes : recommandations et défense des règles WP e.g. NPOV (neutralité de point de vue) et citation des sources

2. Vue d'ensemble des données : le corpus EFGwikiCorpus

Récupérer les données pour en faire un corpus

1. Extraction des discussions depuis les dump i.e. sauvegarde globale des pages courantes (archive diffusée librement sur la page dumps.wikimedia.org/frwiki/)
2. Sélection des discussions "à garder"
 - Au moins 1 post et 2 mots
 - Discussions courantes, archivées (ou pour WP.fr, portant sur la notion de neutralité)

	Articles dans le <i>dump</i>	Pages de disc. dans les <i>dump</i>	Pages de disc. dans le corpus EFG	% des pages de disc. du <i>Dump</i>
E WikiCorpus	14 856 106	7 903 148	2 025 888	26%
F WikiCorpus	3 729 677	1 852 689	266 699	14%
G WikiCorpus	3 920 295	769 091	713 485	93%
EFG WikiCorpus	22 506 078	10 524 928	3 006 072	29%



WikiCorpus : <https://hdl.handle.net/11403/efg-wikicorpus>

	WP.en	WP.fr	WP.de	Total
Pages de discussion (+ Archives)	2 025 888	266 699	713 485	3 006 072
Fils de discussion	6 636 783	608 857	2 121 852	9 367 493
Messages postés	20 025 945	1 832 416	6 938 168	28 796 529

Liens multilingues entre articles

	Aucun lien	Liens WP.en	Liens WP.fr	Liens WP.de	2 Liens WP
E WikiCorpus	19 396		461 768	417 38	270 717 (13%)
F WikiCorpus	1 751	135 480 (51%)		97 439	87 610 (33%)
G WikiCorpus	2 242	369 582 (52%)	268 029		240 228 (34%)

Récupérer les données pour en faire un corpus

— — —

1. Extraction des discussions depuis les dump i.e. sauvegarde globale des pages courantes (archive diffusée librement sur la page dumps.wikimedia.org/frwiki/)
2. Sélection des discussions "à garder"
 - Au moins 1 post et 2 mots
 - Discussions courantes, archivées (ou pour WP.fr, portant sur la notion de neutralité)
3. Détection des éléments structurant chaque discussion : article lié, catégorie et portail thématique, section (fil), message, interaction, signature (nom de l'auteur + date), liens multilingues

Structure des données

TEI-CMC, (Ho-Dac, à paraître) avec
idéalement pour chaque page :

- le titre de la page
- le titre de la discussion
- la liste structurée des messages
avec pour chacun :
 - l'auteur
 - la date de publication
 - la position (indentation)
 - le texte intégral

```
<?xml version='1.0' encoding='UTF-8' ?>
<?xml-model href='../cmc-wikitalk.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStm>
        <title>Discussion:Wikipédia</title>
      </titleStm>
      <publicationStm>
      </publicationStm>
    </fileDesc>
    <profileDesc>
    </profileDesc>
  </teiHeader>
  <text>
    <body>
      <div type="talk">
        <head>Wikipédia</head>
        <div type="thread" xml:id="i.7521136_1">
          <div type="thread" xml:id="i.7521136_2">
            <div type="thread" xml:id="i.7521136_3">
              <head>Question à deux balles</head>
              <post indentLevel="0" mode="written" when-iso="2014-08-27T16:46+02" who="WU00001011" xml:id="i.7521136_3_1">
                <p> Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Milegue" ><name full="yes">Milegue</name></ref><date>27 août 2014 à 16:46 (CEST)</date></signed></p>
              </post>
              <post indentLevel="1" mode="written" when-iso="2014-08-27T16:58+02" who="WU00002889" xml:id="i.7521136_3_2">
                <p>Personnellement : « la » Wikipédia mais « le » projet Wikipédia. "'<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Like%20tears%20in%20rain"><name full="yes">Like tears in rain </name></ref>"' <date>27 août 2014 à 16:58 (CEST)</date></signed></p>
              </post>
              <post indentLevel="2" mode="written" when-iso="2014-08-27T17:55+02" who="WU00003418" xml:id="i.7521136_3_3">
                <p>wikipédia est à la fois "une" encyclopédie, et "un" des projets-frères, so... up to u ;- ) J'ai opté pour "la" wikipédia francophone --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Bibliorock"><name full="yes">Bibliorock</name></ref><date>27 août 2014 à 17:55 (CEST)</date></signed></p>
              </post>
              <post indentLevel="3" mode="written" when-iso="2015-08-27T09:46+02" who="WU00005878" xml:id="i.7521136_3_4">
                <p>J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Agatino%20Catarella"><name full="yes">Agatino Catarella</name></ref> Catarella <date>27 août 2015 à 09:46 (CEST)</date></signed></p>
              </post>
              <post indentLevel="4" mode="written" when-iso="2015-09-14T19:25+02" who="WU00000040" xml:id="i.7521136_3_5">
                <p>Je signale que les Allemands disent « die Wikipedia ». --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Pierrot%20Lunaire"><name full="yes">Pierrot Lunaire </name></ref><date>14 septembre 2015 à 19:25 (CEST)</date></signed></p>
              </post>
            </div>
          </div>
        </div>
      </div>
    </body>
  </text>
</TEI>
```

Structure des données

```
<?xml version='1.0' encoding='UTF-8' ?>
<?xml-model href="http://www.tei-c.org/ns/1.0" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0">
  <teiHeader>
    <fileDesc>
      <titleStm>
        <title>Discussion:Wikipédia</title>
      </titleStm>
      <publicationStm>
        <sourceDesc>
          </fileDesc>
        </sourceDesc>
      </publicationStm>
    </fileDesc>
    <profileDesc>
      </profileDesc>
    </teiHeader>
    <text>
      <body>
        <div type="talk">
          <head>Wikipédia</head>
          <div type="thread" xml:id="i.7521136_1">
            <div type="thread" xml:id="i.7521136_2">
              <div type="thread" xml:id="i.7521136_3">
                <head>Question à deux balles</head>
                <post indentLevel="0" mode="written" when-iso="2014-08-27T16:46+02" who="WU00001011" xml:id="i.7521136_3_1">
                  <p>Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Milegue"><name full="yes">Milegue</name></ref><date>27 août 2014 à 16:46 (CEST)</date></signed></p>
                </post>
                <post indentLevel="1" mode="written" when-iso="2014-08-27T16:58+02" who="WU00002889" xml:id="i.7521136_3_2">
                  <p>Personnellement : « la » Wikipédia mais « le » projet Wikipédia. --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Like%20tears%20in%20rain"><name full="yes">Like tears in rain</name></ref> --<date>27 août 2014 à 16:58 (CEST)</date></signed></p>
                </post>
                <post indentLevel="2" mode="written" when-iso="2014-08-27T17:55+02" who="WU00003418" xml:id="i.7521136_3_3">
                  <p>wikipédia est à la fois "une" encyclopédie, et "un" des projets-frères, so... up to u ;- ) J'ai opté pour "la" wikipédia francophone --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Bibliorock"><name full="yes">Bibliorock</name></ref><date>27 août 2014 à 17:55 (CEST)</date></signed></p>
                </post>
                <post indentLevel="3" mode="written" when-iso="2015-08-27T09:46+02" who="WU00005878" xml:id="i.7521136_3_4">
                  <p>J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Agatino%20Catarella"><name full="yes">Agatino Catarella</name></ref> Catarella <date>27 août 2015 à 09:46 (CEST)</date></signed></p>
                </post>
                <post indentLevel="4" mode="written" when-iso="2015-09-14T19:25+02" who="WU00000040" xml:id="i.7521136_3_5">
                  <p>Je signale que les Allemands disent « die Wikipedia ». --<signed type="signed"><ref target="https://fr.wikipedia.org/wiki/Utilisateur:Pierrot%20Lunaire"><name full="yes">Pierrot Lunaire</name></ref><date>14 septembre 2015 à 19:25 (CEST)</date></signed></p>
                </post>
              </div>
            </div>
          </div>
        </div>
      </body>
    </text>
  </div>
</div>
<div type="thread" xml:id="i.7521136_4">
```

Discussion:Wikipédia

[Article](#) [Discussion](#)

Question à deux balles

Bonjour, quel est le genre de Wikipédia, masculin ou féminin ? Ce n'est pas écrit dans l'article. Doit-on dire le beau Wikipédia ou la belle Wikipédia ? J'ai un avis sur la question mais je vous laisse parler d'abord. --Milegue (discuter) 27 août 2014 à 16:46 (CEST) [[répondre](#)]

Personnellement : « la » Wikipédia mais « le » projet Wikipédia. --Like tears in rain (discuter) 27 août 2014 à 16:58 (CEST) [[répondre](#)]

wikipédia est à la fois une encyclopédie, et un des projets-frères, so... up to u ;-) J'ai opté pour "la" wikipédia francophone --Bibliorock (discuter) 27 août 2014 à 17:55 (CEST) [[répondre](#)]

J'ai toujours vu le féminin. Je ne serais pas contre un changement dans l'article. --Agatino Catarella (discuter) 27 août 2015 à 09:46 (CEST) [[répondre](#)]

Je signale que les Allemands disent « die Wikipedia ». --Pierrot Lunaire (discuter) 14 septembre 2015 à 19:25 (CEST) [[répondre](#)]

Quelles approches pour analyser les données, décrire les interactions ?

3. Approche Top-down

Actes de paroles

Postures des locuteurs

Discussions conflictuelles

4. Approche Bottom-Up

Profilage des discussions

Schémas d'interaction

3. Approche Top-Down

Projeter des catégories sur les données

Caractériser les discussions

— — —

- Selon quelles catégories ?
- A quel niveau ?
 - page
 - fil
 - post
 - autre unité de discours (cf. [CMC Act Taxonomy](#), Herring et al. 2005)
- Annoter par qui ? utilisateurs, sociologues, psychologues, linguistes ...

Annotation des actes de paroles dans 4525 messages extraits de 1864 fils de la E_WP (Ferschke, 2014 – PhD thesis, Darmstad)

Scope	Label	Description
Topic	ERROR	Segmentation Errors
	REFOBJ-PART	Comment about specific section of the article
	REFOBJ-WHOLE	Comment about the whole article
	REFOBJ-META	Meta comment not directly referring to article
Article Criticism		
Turn	CRITCOMPL	Information is incomplete or lacks detail
	CRITACC	Lack of accuracy, correctness or neutrality
	CRITLANG	Deficiencies in language and style
	CRITSUIT	Content not suitable for an encyclopedia
	CRITSTRUCT	Deficiencies in structure, organization or visual appearance
	CRITAUTH	Lack of authority
Self Commitment		
Turn	ACTF	Commitment to action in the future
	ACTP	Report of past action
Requests		
Turn	REQEDIT	Request for article edit
	REQMAINT	Request for admin or maintenance action
Interpersonal		
Turn	ATTPOS	Positive attitude
	ATTNEG	Negative attitude

Table 6.5: Revised annotation scheme for the dialog act classification in the EWD Corpus. Topic level labels are assigned to whole discussion threads while turn level labels are assigned to individual turns.

45% des messages hors catégories

- 2729 messages catégorisés
 - 49% critique p.r. à l'article
 - 21% "rapport d'activité"
 - 18% interaction perso.
 - 12% requête
- Classifieur Naive Bayes (n-grams et traits structuraux)
 - F1 = 0,78

Détection des messages haineux (Wulczyn et al. 2017)

— — —

Annotation par les foules de 1000 messages pris à la fois au hasard et parmi les messages écrits par des contributeurs bannis

Does the comment contain a personal attack or harassment?

- Targeted at the recipient of the message (i.e. you suck).
- Targeted at a third party (i.e. Bob sucks).
- Being reported or quoted (i.e. Bob said Henri sucks).
- Another kind of attack or harassment.
- This is not an attack or harassment.

- 11,7% de posts agressifs
- Classifiseur Perceptrons multi-couche (n-gram de caractères)
 - attaque oui/non
 - $P = 0,97$

Figure 2: The question posed to our Crowdfower annotators.

Détection des conflits et des désaccords (Poudat & Ho-Dac 2019)

— — —

- **Désaccord** = acte de langage réactif qui exprime une réaction négative relative à une assertion, une prise de position préalablement exprimée ou une action (*e.g.* la modification d'un article)
- **Conflit** = désaccords persistants qui impactent le processus d'édition de l'article pouvant amener au blocage du processus d'écriture collaborative (règle des 3Reverts)

⇒ Quelques désaccords directs (20% des désaccords annotés dans un échantillon de pages "bloquées" de la WP.fr)

⇒ Une majorité de désaccords indirects

Expression indirecte du désaccord

Accord sur fond de désaccord	Concession Forme fréquente 26%	Seuls les accords directs avec une concession (oui mais...) ont été annotés <i>D'accord pour refuser le paragraphe ajouté à partir d'arkiv ; en revanche la suppression de la participation d'AR à la mission ne me semblait pas déraisonnable</i>
Se poser comme source évaluative	Incompréhension Forme fréquente 42%	<i>je ne comprends pas très bien pourquoi il faudrait mentionner leur doctorat dès l'introduction.</i>
	Opinion Forme fréquente 27%	<i>Personnellement / Pour ma part, je ne pense pas, je ne crois pas que...</i> <i>ici je ne pense pas que ce soit judicieux</i>
	Doute Forme rare 5%	<i>je doute fort que les médecins étrangers de pays modernes (USA, UK, Allemagne, etc.) soient plus mauvais et moins bien formés que les médecins Français</i>

Les limites de l'approche Top-down

— — —

- Quelles catégories pour décrire les discussions ?
- A quel niveau ?
- Selon quel point de vue (acte de chaque contributeur, problème dans l'échange, etc.)?
- Beaucoup d'interactions peu "intéressantes"
 - 45% des messages qui ne correspondent pas à un acte de dialogue prédéfini
 - 88% des messages jugés conflictuels
 - difficile de juger si un désaccord est constructif ou problématique

⇒ Alternative : se laisser guider par les données en adoptant une approche bottom-up

4. Approche bottom-up

Faire émerger des profils depuis les données

Questions

— — —

Exploration plus large des discussions wikipédia

Approche inductive, sans viser un phénomène particulier

Concentration sur la nature des interactions et pas le contenu (thématique) des échanges

Méthodologie mixte :

approches quantitatives pour faire ressortir les tendances et les cas extrêmes

examen qualitatif d'individus atypiques ou d'échantillons ciblés



Corpus : <https://hdl.handle.net/11403/efg-wikicorpus>

	WP.en	WP.fr	WP.de	Total
Pages de discussion (+ Archives)	2 025 888	266 699	713 485	3 006 072

Profil commun : 1 ou 2 messages "échangé" dans la semaine

Messages par fil : moyenne (max)*	2,6 (651)	2,5 (149)	2,7 (305)	
Users par fil : moyenne (max)*	1,8 (97)	1,7 (43)	1,9 (120)	
Durée d'un fil : médiane moyenne (jours)**	5,3 260	2,1 184	6,1 349	
Durée moyenne entre deux messages (jours)**	4,1 233	1,5 172	4,6 317	

* **médiane** = 1, **min** = 1

** **min** = 1 min, **max** = 16 ans (2003 - 2019)

Profil commun : 1 ou 2 messages "échangé" dans la semaine, mais pas nécessairement

— — —

☰ Discussion:Wikipédia

🌐A ▾

[Article](#) [Discussion](#)

[Lire](#) [Voir le texte source](#) [Voir l'historique](#) [Outils](#) ▾

Une formulation à améliorer [[modifier le code](#)]

"Elle est devenue en quelques années l'encyclopédie la plus grande et la plus consultée au monde."

Je pense qu'il serait plus satisfaisant d'écrire : "En quelques années, elle est devenue la plus grande/importante/volumineuse [au choix] encyclopédie disponible dans le monde et la plus consultée."

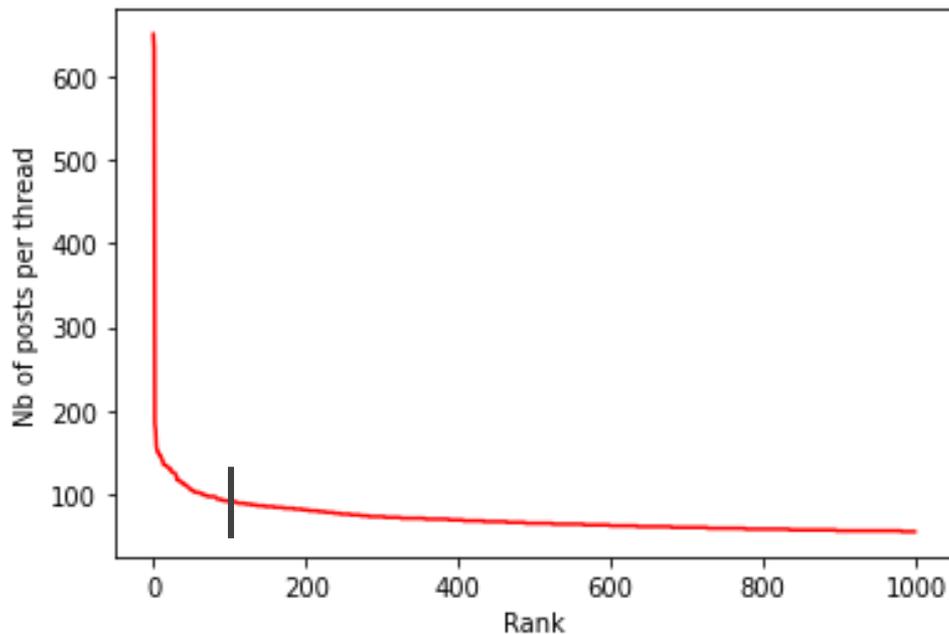
Cordialement.

--[92.91.74.18](#) ([discuter](#)) 27 juillet 2020 à 19:30 (CEST) [[répondre](#)]

Je vient de le faire. [0770he](#) ([discuter](#)) 12 février 2024 à 14:04 (CET) [[répondre](#)]

Profil atypique : Discussions (extrêmement) longues

Dans WP.en, les 100 discussions les plus longues ont plus de 90 messages



La 100 discussions les plus longues (WP.en)

- *Discussions standard (68/100)*
 - généralement conflictuelles
 - peuvent être structurées (ruptures, sections) pour une meilleure lisibilité
- *Sondages ou séries de sondages (26/100)*
- *Listes (6/100)*
Discussions avec un seul auteur, cf plus loin



☰ Talk:Gamergate (harassment campaign)/Archive 12

KotakuInAction moderators misogynist/anti-feminist/interested in female subjugation porn

A simple [investigation by BuzzFeed](#) determines that seven out of 10 KotakuInAction moderators are also moderators or former participants in sub-Reddits such as "struggleporn," "breakfeminazis," "ProlapseVille," "WhatFeministsLookLike" and other forums described by BuzzFeed as "devoted to either the physical and emotional deardadation and humiliation of women. or in subre

... [184 messages] ...

If you have issues with my behavior, this is not the forum to discuss it.—[Ryūlong \(琉竜\)](#) 20:56, 30 October 2014 (UTC)

ArbBreak

While [this edit](#) is appreciated to defuse / better back the statement (while we affirm BLP issues), the approach that statement again points to the issues in the past I've pointed to about laverina on subtle attacks / thumbing ones nose at the proGG side t

"ArbBreak" (coupures arbitraires) pour segmenter

When material is both verifiable and noteworthy, it will have appeared in more reliable sources.". This tabloid. [PseudoSomething \(talk\)](#) 02:40, 31 October 2014 (UTC)

ArbBreak 2

Let's get back to a key point: working on the assumption this is NOT a BLP issue, and we will take BF as a RS in conjunction with

piece from BuzzFeed). You really need to calm down.—[Ryūlong \(琉竜\)](#) 05:34, 31 October 2014 (UTC)

Stop talking about other editors here on the article talk page, it violates the expected standards of conduct, and if continued, users may be sanctioned per [Wikipedia:General sanctions/Gamergate](#). [Dreadstar](#) + 07:20 October 2014 (UTC)

[Gawker](#) also comments on the fact that a kotakuinaction mod also heads disreputable, misogynist ones such as breaker

La 100 discussions les plus longues (WP.en)

- *Discussions standard* (68/100)
 - généralement conflictuelles
 - peuvent être structurées (ruptures, sections) pour une meilleure lisibilité

- *Sondages ou séries de sondages* (26/100)

651 messages = 6 utilisateurs donnant leur avis au sujet de l'importance relative de chacun des 134 personnages de *One Piece*

- *Listes* (6/100)
Discussions avec un seul auteur, cf plus loin

Straw poll

Mr. 4

Resolved – To be mentioned in the Baroque Works section.

- **Mention** under a possible Baroque Works section. He has a one fight end. [Spindori \(talk\)](#) 19:55, 7 August 2010 (UTC)
- **Keep** -- He was an Officer Agent of Baroque Works. He deserves his own section. [Rico70](#)
- **Merge** into Baroque Works - an extremely minor character who has appeared in only one arc. [Sjones23 \(talk - contributions\)](#) 23:42, 7 August 2010 (UTC)
- **Mention** among the Baroque Works agents. His small role doesn't merit anything more. [Goodraise](#) 01:52, 10 August 2010 (UTC)
- **Keep** Same reason as Rico70. --[Dylandh \(talk\)](#) 23:20, 12 August 2010 (UTC)

Mr. 5

Resolved – To be mentioned in the Baroque Works section.

- **Mention** under a possible Baroque Works section. He has a one fight end. [Spindori \(talk\)](#) 19:55, 7 August 2010 (UTC)
- **Remove** from the list. I wasn't thinking when I made most of these (a mistake). Mr. 5 only appears during the Little Garden arc, and it by itself is too minor to accredit any casted character (the very reason the two giants aren't listed). [Spindori \(talk\)](#) 17:39, 16 August 2010 (UTC)
- **Keep** -- He was an Officer Agent of Baroque Works. He deserves his own section. [Rico70](#)
- **Merge** into Baroque Works - an extremely minor character who has appeared in only one arc. [Sjones23 \(talk - contributions\)](#) 23:42, 7 August 2010 (UTC)
- **Mention** among the Baroque Works agents. His small role doesn't merit anything more. [Goodraise](#) 02:02, 10 August 2010 (UTC)
- **Keep** Same reason as Rico70. --[Dylandh \(talk\)](#) 23:20, 12 August 2010 (UTC)
- **Merge**, Even though he was the main antagonist to the Whisky Peak arc and a supporting antagonist to the Little Garden arc, he is nothing more after that. Both Mr. 5 and Miss Valentine should be merged into the Baroque Works section. - [SuperTiencha \(talk\)](#) 01:22, 14 August 2011 (UTC)

... [651 messages] ...
une section par personnage

Yamakaji

Resolved – Yamakaji should not keep his section, and is to be deleted

- **Delete**, Minor character who is only involved in one story arc. [Sjones23 \(talk - contributions\)](#) 17:09, 29 July 2010 (UTC)
- **Remove** him from the list. As far as I know, his name is only mentioned in one of the data books and his role so far has been minor at best. [Goodraise](#) 23:55, 30 July 2010 (UTC)
- **Keep** -- He is one of the powerful Vice Admirals of the marines, and he and four other Vice Admirals led the Buster Call on Enies Lobby. He deserves his own section. [Rico70 \(talk\)](#) 04:49, 31 July 2010 (UTC)
- **Keep** Same reason as Rico70. --[Dylandh \(talk\)](#) 23:20, 12 August 2010 (UTC)

Yasopp

Resolved – Consensus is to establish mention of character under Shanks

- **Merge**, can be mentioned in Shanks' section. He is a minor background character. [Sjones23 \(talk - contributions\)](#) 17:09, 29 July 2010 (UTC)
- **Mention** him with Shanks and Usopp. That should be more than enough for a character with as little screen time as him. [Goodraise](#) 23:53, 30 July 2010 (UTC)

Autres cas extrêmes

— — —

Utilisateurs très productifs

- Parmi les 1000 plus gros contributeurs, pas de corrélation entre le nombre de messages et le nombre d'édicions
→ deux activités bien distinctes

Discussions dans le temps :

- Discussions filées (e.g. un ou deux commentaires tous les 3-4 ans sur 15 ans)
- Réponse à une question/demande très éloignée dans le temps (jusqu'à 16 ans pour répondre), souvent sans marque explicite (asynchronicité des CMC)
→ dimension patrimoniale des discussions Wikipedia (cf "Nostalgia Wikipedia")

Étude des schémas d'interaction

Choix méthodologiques : simplifications et orientations

— — —

Etant donné la difficulté à identifier la véritable structure d'une discussion (qui répond à qui) :

- Suppression des discussions avec des segments isolés (pas de signature ni de date)
- Pas de prise en compte du niveau d'indentation
- Mise à plat en gardant l'ordre textuel des messages
- Pseudonymisation des auteurs des messages suivant l'ordre (A, B, C...)

Concentration de nos études sur les métadonnées exploitables :
qui écrit où, quand, combien de fois, et après/avant qui ?

Vue d'ensemble : schémas

— — —

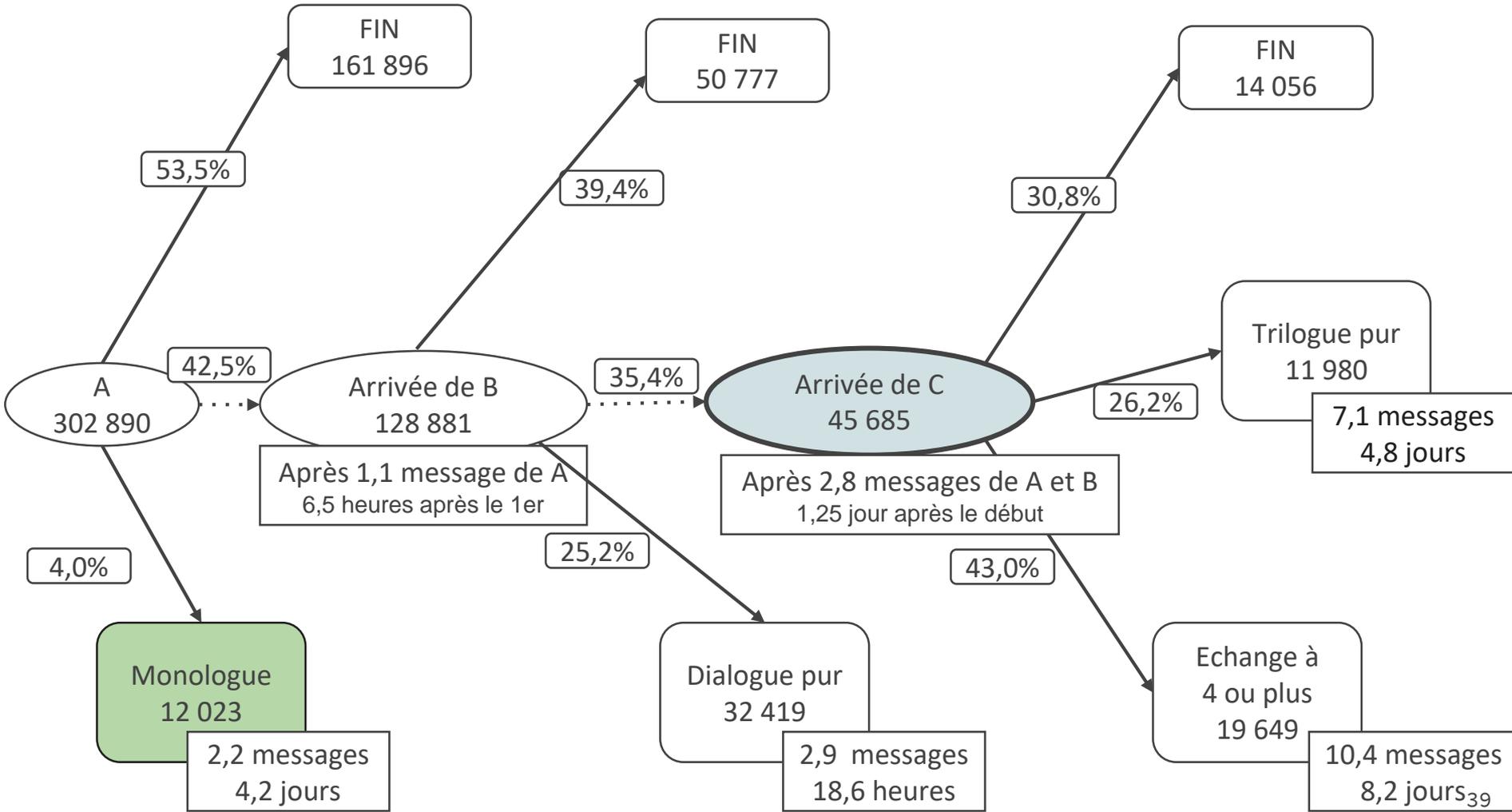
- Représenter simplement la structure d'une discussion
 - Considérer la suite des messages suivant leur auteur
 - Chaque participant différent est représenté par une lettre arbitraire : A pour celui qui initie la discussion, B pour le premier qui répond, etc.
 - Pas de prise en compte des structures complexes, on se base sur la linéarité observée sur la page
- Exemples :
 - ABAB : dialogue
 - AAA : monologue
 - ABCDEFGHIJKL : vote/consultation
 - ABCBCDEFA : c'est plus compliqué

Statistiques des schémas

— — —

	Anglais	Français	Allemand
Nombre de discussions	3 385 583	302 890	1 485 648
Nombre de schémas différents	116 893	13 103	54 834

Anglais (%)		Français (%)		Allemand (%)	
A	40.2	A	53.5	A	47.2
AB	35.1	AB	16.8	AB	21.1
ABC	4.2	ABA	3.8	ABC	5.1
AA	2.3	AA	3.4	ABA	3.3
ABA	2.2	ABC	3.2	AA	2.8
ABCD	1.2	ABAB	1.3	ABCD	1.4
ABAB	0.7	ABCD	0.9	ABAB	1.2
ABB	0.6	ABB	0.8	ABB	1.1
AAB	0.6	AAB	0.7	AAB	0.7
ABAC	0.5	ABABA	0.6	ABAC	0.7



4.1. Monologues

Méthode

— — —

8,5% des discussions avec au moins 2 messages n'impliquent qu'un seul auteur
Certains monologues peuvent contenir de nombreux messages (jusqu'à 28)

- Etude des monologues longs (cas extrêmes)
 - Permet d'identifier des phénomènes facilement
- Analyse des amorces (AA)
 - Typologie des raisons du deuxième message

Les cas extrêmes (FR)

— — —

136 fils avec un seul utilisateur contiennent plus de 5 messages

Trois usages principaux observés:

- Les **séries pures** : l'utilisateur énumère des informations, questions, remarques, etc. une par message **2011-04** [[modifier le code](#)]

Faut-il, à votre avis, signaler dans [Expansion terrestre](#) l'article de Robert Muir Wood, [Is the Earth getting bigger?](#), New Scientist, 8 février 1979, p. 387 ? [Visite fortuitement prolongée](#) (d) 2 avril 2011 à 22:16 (CEST) [[répondre](#)]

Faut-il signaler l'article de Robert Muir Wood, [Geological cul de sac](#), New Scientist, 30 juin 1988, p.85 ? [Visite fortuitement prolongée](#) (d) 2 avril 2011 à 22:16 (CEST) [[répondre](#)]

Faut-il signaler le livre *Expanding Earth Symposium*, 1983, compte rendu d'une conférence de 1981 ? [Visite fortuitement prolongée](#) (d) 2 avril 2011 à 22:16 (CEST) [[répondre](#)]

Faut-il signaler l'article de David Noel, [Fixed-Earth and Expanding-Earth Theories -- Time for a Paradigm Shift?](#), février 2005 ? [Visite fortuitement prolongée](#) (d) 2 avril 2011 à 22:16 (CEST) [[répondre](#)]

- Le **tableau de bord** : fil utilisé comme outil de suivi des opérations personnelles sur la page (pas d'interaction)
- Les **véritables monologues**

Les véritables monologues

Communes du Labourd [modifier le code]

Une petite visite aux Wikipédias espagnole et basque (je ne parle aucune de ces deux langues mais bon on voit à peu près ce que ça raconte...) me fait c. Media que la Wikipedia française semble la seule à fournir des frontières très précises pour les trois provinces basques du nord, en délimitant la "liste des communes" de chacune.

Pas du tout spécialiste d'histoire basque, mais de peu de sympathie instinctive pour les nationalismes de tout poil, je suis quand même un peu perplexe. Tout cela est-il bien sourcé ? (C'est surtout la présence du **Boucau** dans la liste qui me fait tiquer ; je vois que la wikipedia basque commence sur la page **eu:Bokale** par une phrase que je suis bien incapable de comprendre mais qui mêle Gascogne et Labourd. Quelles sont les sources qui justifient cette inclusion ?

De même (je poste le problème ici pour ne pas tout disperser) je vois que dans la page **Bidache**, cette commune est décrite comme «*aujourd'hui plus ou moins incluse dans le Pays Basque*» mais qu'elle figure dans la liste des communes de **Basse-Navarre**.

Bref, je suis spontanément un peu gêné de ces listes de communes du Pays Basque ; d'autant qu'elles ont vocation à dépasser la seule page qui les reprend mais se baladent sous forme de bandeau sur chacune des communes concernées. S'il ne tenait qu'à moi, je sabrerais impitoyablement, mais il vaut mieux consulter ceux qui connaissent quelque chose à la question que laisser l'initiative aux incompetents comme moi.

Bref, quelles sources fiables pour
suis beaucoup plus quand on e
(CEST) [répondre]

Ah oui intrigué aussi dans l'article
l'Adour s'est déplacé, mais que
(CEST) [répondre]

Je complète mon propos en fournissant
fournir des frontières précises pour le
sérieuses ? --Touriste 8 août 2006
transférera au ressort du Parlement d
règle pas la question de savoir où s'a

Bon je m'autoréponds un peu, de
et XVIIIème nettement moins. Il
Généralités sont disponibles sur
aux généralités, a été détaché
et la sénéchaussée des Lannes

Continuons à monologuer... La
Bayonne, la zone intra muros d
bailliage du Labourd. --Touriste

Et monologuons avec indentation
historique de Bayonne de D
(Discuter) 11 septembre 2006

Utilite, notamment pour la Basse-Navarre ou la liste des communes me posait davantage problème, la référence suivante : <http://www.cervantesvirtuar.com/servlet/SiveObras/hist/46848064104804728500080/p0000006.htm> --Touriste (Discuter) 15 août 2006 à 00:40 (CEST) [répondre]



Choisi ou subi?

Monologues (réflexion solitaire) vs monodialogues (plaintes, reproches)

Bon je m'autoréponds un peu, épaté de voir qu'un peu partout sur les pages sur des provinces l'histoire médiévale est assez détaillée, celle des XVIIème et XVIIIème nettement moins. Il existe en 1789 un "bailliage du Labourt" (orthographe d'époque) - dont notamment **les cahiers de vœux pour les États Généraux** sont disponibles sur le web. Ce bailliage appartient au **gouvernement** de Guyenne-Gascogne mais, au vu de la page Wikipedia consacrée aux **généralités**, a été détaché depuis très peu -en 1784- semble-t-il de la **généralité d'Auch**. Il existe donc bien une frontière entre le bailliage de Labourd et la sénéchaussée des Lannes. Où passe-t-elle, je n'en sais rien... --Touriste 8 août 2006 à 21:18 (CEST) [répondre]

Continuons à monologuer... La réponse semble être dans la page **Bayonne** - merci la Wikipédia ! La frontière passe dans la commune actuelle de Bayonne, la zone intra muros des remparts (et bien sûr la rive droite de l'Adour) se rattachant à la sénéchaussée des Lannes, et la partie sud de la ville au bailliage du Labourd. --Touriste 8 août 2006 à 21:31 (CEST) [répondre]

Et monologuons avec indentation, c'est tellement plus beau. Après consultation de diverses sources en bibliothèque, et notamment le Dictionnaire historique de Ducéré, il apparaît que ces infos de l'article Bayonne sont fantaisistes et je vais d'ailleurs m'atteler à les modifier. **Touriste** * (Discuter) 11 septembre 2006 à 23:51 (CEST) [répondre]

Comment s'installe un monologue?

Etude des amorces "AA"

— — —

Méthode

1. Sélection aléatoire de 100 discussions dont les deux premiers messages ont le même auteur (indépendamment de la suite)
2. Examen détaillé du contenu des deux messages
3. Etablissement d'une première typologie
4. Annotation (français et anglais)

Complément (27%)

— — —

A ajoute un ou plusieurs éléments (information, argument, détail, question) à son premier message

Renommage ? [\[modifier le code\]](#)

Hello,

Ahhh donc c'est Horner et pas Hörner, j'ai appris qqch :) La question du jour, c'est est-ce qu'on dit plus algorithme de H. ou méthode de H ? Selon google c'est kif-kif 730 contre 800. J'ai pas trop d'avis. En tout ca, une fois qu'un choix aura été fait je ferais un redirect. (:Julien:) ➡ 26 juillet 2007 à 13:16 (CEST)

[\[répondre\]](#)

Y'a aussi schéma de Horner avec 474 hits. Pas plus d'avis. (:Julien:) ➡ 27 juillet 2007 à 13:20 (CEST) [\[répondre\]](#)

Correction (12%) et auto-réponse (2%)

— — —

A a identifié une erreur dans son précédent message (et vient la corriger), ou a trouvé la réponse à sa question

Ses mémoires [\[modifier le code \]](#)

Je lis : *Elle commence peu avant l'exécution de son père (le 21 janvier 1793) à écrire ses mémoires*. A 14 ans, cela me paraît douteux d'autant que j'ai lu dans la thèse d'Hélène Becquet qu'elle avait commencé à écrire son journal en 1799.
--Priper (d) 7 février 2013 à 14:36 (CET) [\[répondre \]](#)

Au temps pour moi. La biographie de Castelot confirme la rédaction à cet âge des fameux mémoire. Je supprime la demande de réf.--Priper (discuter) 1 septembre 2013 à 12:12 (CEST) [\[répondre \]](#)

Union City, une ville du comté, est la plus peuplée des États-Unis. [\[modifier le code \]](#)

J'ai du mal à comprendre ou à croire cette phrase...

Sxilderik 2 novembre 2006 à 19:11 (CET) [\[répondre \]](#)

Ok, d'après en:Union City, New Jersey, il faut comprendre: la plus densément peuplée... je corrige
Sxilderik 2 novembre 2006 à 19:13 (CET) [\[répondre \]](#)

Relance (2%)

— — —

A n'ayant pas eu de réponse à son premier message, il relance un appel (ciblé ou non) pour obtenir une réaction

Sur la réécriture de la section "apologie par l'extrême droite" [[modifier le code](#)]

Comparatif de la version en cours et la version proposée par Parmatus [[modifier le code](#)]

Afin de pouvoir mieux comprendre l'ampleur des modifications proposées, voilà un petit tableau synthétisant selon moi les différences entre la version en cours et la version proposée par Parmatus. Certaines sont des modifications de style (ex : faire deux phrases au lieu d'une), d'autre de fond et légères (ex : ajout de "en 1995"), d'autres de fond plus conséquentes.

#	Version en cours	Version proposée par Parmatus	commentaire de Parmatus	commentaire de GiL GooL	Commentaire de Raconish	Commentaire de GastelEtwane
---	------------------	-------------------------------	-------------------------	-------------------------	-------------------------	-----------------------------

...

Le but est de débattre par la suite sur les modifications auxquelles tient Parmatus et qui sont contestées par au moins une autre personne, et d'éviter d'ergoter dans des débats chronophages ou énergivores sur des détails à propos desquels personne ne tient réellement ou personne n'est contre. En particulier, pour GiL GooL qui a proposé une autre version, la question pourrait être : quelles modifications proposées par Parmatus serais-je prêt à intégrer à ma proposition de texte? Ou encore : si je devais partir de la proposition de texte de Parmatus, et que je voulais le modifier au minimum pour le rendre acceptable par moi, qu'est-ce que je laisserais, autrement dit quelles sont les modifications faites par Parmatus qui me paraissent acceptables?

Je ferai demain de même avec le texte proposé par Gil GooL (un deuxième tableau, avec liste des différences entre la version en cours et la version proposée par lui).[Thémistocle](#) ([discuter](#)) 2 avril 2014 à 21:59 (CEST) [[répondre](#)]

Parmatus pourrait-il corriger ce tableau s'il y a lieu, ou confirmer qu'il n'y a rien à corriger, et éventuellement compléter la colonne "Commentaire de Parmatus" s'il veut apporter des précisions? Merci.[Thémistocle](#) ([discuter](#)) 6 avril 2014 à 22:13 (CEST) [[répondre](#)]

Tableau de bord / Série (11%)

— — —

A énumère des actions, questions, réflexions ou éléments d'information (cf monologues longs)

Voix Française [\[modifier le code \]](#)

André Valmy double dans Pendez vous les Hauts et courts, La cité des dangers. Yannick455 (discuter) 9 juillet 2019 à 14:25 (CEST) [répondre]

Jacques Sarthout double Monsieur Joe. Yannick455 (discuter) 9 juillet 2019 à 14:30 (CEST) [répondre]

Roland Ménard double dans La charge Héroïque. Yannick455 (discuter) 9 juillet 2019 à 14:34 (CEST) [répondre]

Pierre Traud double dans le convoi des braves. Yannick455 (discuter) 9 juillet 2019 à 14:34 (CEST) [répondre]

Yves Furet double dans Rio Grande. Yannick455 (discuter) 9 juillet 2019 à 14:35 (CEST) [répondre]

Richard Francoeur double dans l'homme de la vallée perdue. Yannick455 (discuter) 9 juillet 2019 à 14:37 (CEST) [répondre]

Raymond Loyer double dans La vengeance aux deux vengeance. Yannick455 (discuter) 9 juillet 2019 à 14:40 (CEST) [répondre]

Albert Augier double dans Les Cheyennes. Yannick455 (discuter) 9 juillet 2019 à 14:41 (CEST) [répondre]

André Valmy double dans Will Penny le cavalier Solitaire. Yannick455 (discuter) 9 juillet 2019 à 14:44 (CEST) [répondre]

Henry Djanick double dans La Horde Sauvage. Yannick455 (discuter) 9 juillet 2019 à 14:46 (CEST) [répondre]

Henry Djanick double dans La Horde Sauvage. Yannick455 (discuter) 9 juillet 2019 à 14:46 (CEST) [répondre]

Pierre Leproux double dans les Géants de l'Ouest. Yannick455 (discuter) 9 juillet 2019 à 14:47 (CEST) [répondre]

Jean Amadou double dans Major Dundee, Chisum. Yannick455 (discuter) 9 juillet 2019 à 14:48 (CEST) [répondre]

Nicolas Vogel double dans Rio Verde. Yannick455 (discuter) 9 juillet 2019 à 14:51 (CEST) [répondre]

Philippe Dumat double dans junior Bonner, Le dernier Bagarreur. Yannick455 (discuter) 9 juillet 2019 à 14:53 (CEST) [répondre]

Nicolas vogel double dans le voleur des trains. Yannick455 (discuter) 9 juillet 2019 à 14:55 (CEST) [répondre]

Henry Djanick double dans Dillinger. Yannick455 (discuter) 9 juillet 2019 à 14:58 (CEST) [répondre]

Claude Bertrand double dans Guet apens . Yannick455 (discuter) 9 juillet 2019 à 14:58 (CEST) [répondre]

Jean Martinelli double dans Suthgarland Express. Yannick455 (discuter) 9 juillet 2019 à 15:00 (CEST) [répondre]

Pierre Garin double dans la chevauchée sauvage. Yannick455 (discuter) 9 juillet 2019 à 15:02 (CEST) [répondre]

Jean Paul Coquelin double dans le solitaire de Fort Humboldt. Yannick455 (discuter) 9 juillet 2019 à 15:08 (CEST) [répondre]

Michel Barbey double dans L'inévitable catastrophe. Yannick455 (discuter) 9 juillet 2019 à 15:13 (CEST) [répondre]

Georges Brousseau double le monstre de train. Yannick455 (discuter) 9 juillet 2019 à 15:41 (CEST) [répondre]

Georges Atlas double dans Le Chasseur. Yannick455 (discuter) 9 juillet 2019 à 15:42 (CEST) [répondre]

Jean Violette double dans L'aube Rouge. Yannick455 (discuter) 9 juillet 2019 à 15:46 (CEST) [répondre]

Jean Violette double dans Cherry 2000. Yannick455 (discuter) 9 juillet 2019 à 15:48 (CEST) [répondre]

Jean Brousseau double dans Le rêve de Bobby. Yannick455 (discuter) 9 juillet 2019 à 15:49 (CEST) [répondre]

Pascal Renwick puis Claude Préfontaine double dans une équipes aux anges . Yannick455 (discuter) 9 juillet 2019 à 15:52 (CEST) [répondre]

[Ajouter un sujet](#)

Action (28%)

— — —

- 1 : A suggère une action d'édition
- 2 : A annonce avoir effectué la dite action

Nombre de locuteurs par pays [\[modifier le code \]](#)

La source est au départ unique, indiqué en référence en haut du tableau (Dieter W. Halwachs, *Speakers and Numbers*, University of Graz, 2004). Des modifications ont été faites et défaites pour la Roumanie pour passer de 1030000 à 535250 à partir de [Démographie de la Roumanie#La population par ethnie \(recensement 2002\)](#), qui précise pourtant « certaines populations semblent sous évaluées. C'est particulièrement le cas des roms. » Il vaudrait mieux avoir des chiffres cohérents tirés tous de la même source, ou sinon faire une deuxième colonne pour les recensements (par ailleurs la colonne % n'est pas claire, apparemment l'explication là-dessus est dans la version en anglais de l'article). — Oliv [Éppen hozzám?](#) 19 septembre 2013 à 16:47 (CEST) [\[répondre \]](#)

✓ Fait. [↗](#) — Oliv [Éppen hozzám?](#) 19 mars 2014 à 08:19 (CET) [\[répondre \]](#)

Réaction à un événement (16%)

— — —

A réagit à un événement (action d'édition typiquement) qui a eu lieu à la suite de son premier message

Courage pour ta traduction !

Nous comptons sur toi :-)

[Fabos](#) ✉ 28 sep 2005 à 09:55 (CEST)

Super ça avance ! [Fabos](#) ✉ 6 octobre 2005 à 23:23 (CEST) [[répondre](#)]

Incohérences majeures par rapport au livre [[modifier le code](#)]

Et un paragraphe inutile de plus :(Résumé un livre de plus de 700 pages en un film d'une longueur acceptable est impossible. Et puis, un film **n'est pas** un livre. Il y aura donc inévitablement des différences entre l'œuvre originale (le livre) et son adaptation (le film). Mais ma crainte principale est que certains (beaucoup ?) vont confondre **incohérences majeures** et **différences** (les dernières modifications vont déjà dans ce sens). Cette analyse n'a pas, selon moi, sa place dans Wikipédia. Laissons cela aux sites spécialisés et autres blogs de fans. --[H2o](#) [[parlons-en](#)] 6 août 2009 à 17:40 (CEST) [[répondre](#)]

[Suprememangaka](#) a transformé l'intitulé en **Différences majeures par rapport au livre**. Soit. Mais qui déterminera que la différence est **majeure**. Autant se limiter à ce qui constitue effectivement des incohérences, qui amènent des contradictions entre les films et/ou livres précédents et suivants. --[H2o](#) [[parlons-en](#)] 9 août 2009 à 20:15 (CEST) [[répondre](#)]

Une forme de dialogue multicanal

— — —

Dans certaines situations, une forme de dialogue se met en place, articulant messages sur différentes pages, éditions, commentaires

Ah, a propos de musique "savante" [[modifier le code](#)]

On appelle bien musique savante toute musique dans laquelle l'exécutant et le compositeur sont dissociés. Donc en occident: musique savante égale plus ou moins musique dite classique. Ca peut paraître péjoratif et méprisant pour les autres formes de musique, mais ça ne l'est pas. Du moins c'est comme ça qu'on dit en musicologie. [BiffTheUnderstudy](#) 15 février 2007, 20hquelquechose

je t'ai expliqué ça sur ta page de discussion: merci de la lire. ~~cet article n'est pas à sa place, il remplacera l'actuel musique de la période classique. En attendant je suis en train d'écrire ailleurs un musique classique au sens large qui remplacera celui là. Donc merci de ne pas toucher d'ici là.~~ --[BiffTheUnderstudy](#) 24 février 2007 à 13:27 (CET)

[[répondre](#)]

Distribution

Distribution similaire sur les deux jeux
Types spécifiques à la Wikipédia :
40-50%
Faux positifs dus aux problèmes
de parsing (usages non-standards
du wikicode)

Categories	English	French
Addendum / Complément	30	24
Action report / Action	24	26
Reaction to event / Réaction	8	14
List / Liste	8	10
Self-correction / Autocorrection	4	11
Self-answer / Autoréponse	6	2
Chasing up / Relance	4	2
False positive / Faux positif	16	11
Total	100	100

Annotation humaine et automatique

Expérimentations

— — —

1/ Deux étudiant.e.s (licence SDL) ont appliqué la typologie à nos jeux de données

- Objectif : évaluer la stabilité de la typologie et notre guide d'annotation
- Accès aux données originales (navigateur Web), en ne considérant que les 2 premiers messages

2/ Premier essai avec une annotation par des LLMs

- Objectif : tester la possibilité d'une classification automatique pour une étude quantitative large
- "Zero-shot prompting" avec les détails de la tâche et une courte description de chaque catégorie (en anglais). Chaque discussion traitée indépendamment, sur la base du contenu en texte brut des deux premiers messages

Mesure du kappa de Cohen et de métriques de classification par classe

Details (F1)

- *Complément et auto-correction* faciles pour les humains et les LLMs
- *Action* très facile pour les humains mais plus difficiles pour les LLM, idem pour *réaction*
- *auto-reponse* et *liste* inaccessibles aux LLMs

Category	Anotator1 French k=0.67	Annotator2 English k=0.69	Mistral Openorca English k=0.17
Addendum	0.65	0.71	0.55
Self-correction	0.59	0.54	0.57
Self-answer	0.50	0.67	0.00
Chasing up	0.33	0.80	0.15
Action report	0.88	0.84	0.39
Reaction to event	0.67	0.55	0.17
List	0.60	0.57	0.00
Macro-average F1	0.60	0.67	0.28

Principales difficultés identifiées

— — —

Complément trop large (rappel élevé, précision faible)

Auto-réponse ne correspond pas toujours à une question explicite (actes de langage indirects du type “Je me souviens plus”, “je comprends pas”)

Listes ambiguës et nécessitent plus de 2 items pour être identifiées

Réactions très difficiles dans le cas de communications multicanal (besoin d’inférer un dialogue implicite)

Premiers résultats

— — —
Le monologue : un des phénomènes apparus dans une étude plus globale des interactions

Une première typologie des séquences monologiques dans les discussions Wikipédia des situations classiques et fréquentes (complétion, correction / réponse, relance)
applicable à d'autres situations de communication en ligne

De nouvelles pratiques identifiées, spécifiques à la Wikipédia
liste / tableau de bord / dialogue multicanal
conséquences dans l'usage des discussions WP comme données de CMR

Trouver une meilleure stratégie pour la classification automatique

4.2. Irruption d'un tiers dans un dialogue en ligne

Questions

— — —

Le dialogue est une forme privilégiée d'interaction (cf. Garrod and Pickering 2004)

Qu'en est-il des trilogues, formes d'interaction plus élaborées?

Comment s'engage une conversation en ligne à plusieurs (multilogue ou multiparty conversation)?

Autrement dit : qu'est-ce qui motive l'intervention d'un tiers dans une discussion déjà entamée ?

Premier bilan quantitatif

- Les échanges faisant intervenir 3 contributeurs sont très courants, et même majoritaires lorsqu'une discussion s'étend sur plus de deux messages (57% des cas)
- 63% des échanges impliquant un 3e contributeur suivent un schéma *ABC* i.e. où les 3 premiers messages sont postés par 3 contributeurs différents (A suivi de B suivi de C)
- L'intervention de C se fait avec un temps de réaction plus long (délai médian de réponse de 7h) que de façon générale (2h de délai entre deux messages)

Ex.

Discussion:Troubles au Tibet en mars 2008/archives 2009

🔍

Article Discussion

Lire Modifier le code Ajouter un sujet Voir l'historique Outils

< Discussion:Troubles au Tibet en mars 2008

Youtube [modifier le code]

- 1 A J'ai supprimé les liens Youtube, après avoir lu la discussion suivante :[1]🔗. Cela fait toujours autorité. --Rédacteur Tibet (d) 21 février 2010 à 19:26 (CET) [répondre]
- 2 B Tu retire ces vidéos parce qu'elles montrent des témoignages accablants qui contredisent tes convictions. Le contenu de ces vidéos est une compilations d'extrait de témoignages diffusés sur de nombreuses télé publiques. La suppression est totalement arbitraire et abusive. Tu profite de la déstabilisation apportée par deux contributeurs pour relancer une guerre d'édition. --Aacitoyen (d) 21 février 2010 à 19:47 (CET) [répondre]
- 3 A J'ai donné l'explication ci-dessus, merci de ne pas faire de procès d'intention.--Rédacteur Tibet (d) 21 février 2010 à 19:49 (CET) [répondre]
- 4 C Ces videos n'ont pas été importer par les propriétaires des droits d'auteurs. Elles y sont donc illégalement. Les vidéos sont donc à retirer. Par contre, la vidéo peut être citée comme source ... Cdt, Kyro^{Tok Wiz Mi} le 21 février 2010 à 20:28 (CET) [répondre]
- 5 B Si ce n'est pas tes convictions qui ont fait retirer ces deux vidéos, pourquoi n'as tu pas retiré les vidéos qui vont dans le sens de tes convictions dans le même chapitre, et qui posent la même question de copyvio ? --Aacitoyen (d) 21 février 2010 à 20:37 (CET) [répondre]
- 6 D Question à Kyro : Je ne connais pas très bien la question, mais en quoi un simple lien actif (donc du code et non la vidéo elle-même) vers une vidéo située sur un site tiers est-il susceptible de constituer une violation du droit d'auteur ? --Elnon (d) 21 février 2010 à 20:40 (CET) [répondre]
- ... *10 messages et deux jours plus tard*
- 17 D J'ai remis les intitulés des vidéos retirées, mais sans lien actif, de façon à rétablir l'équilibre des points de vue et la neutralité dans cette section.--Elnon (d) 23 février 2010 à 23:25 (CET) [répondre]

Approche qualitative



Exploration :

Extraction des 45685 discussions avec 3 participants ou plus

Calcul des spécificités des contenus du premier message de chacun (A, B, C)

Typologie :

Proposition des principales raisons pour l'intervention de C

Annotation :

Etiquetage manuel de deux échantillons

(trois co-auteurs, calcul de l'accord inter-annotateur)

Spécificités des mots-formes des premiers messages de A, B et C

	A	Indice	B	Indice	C	Indice
Spécificités positives	depuis	1000,0	tu	1000,0	✓	118,7
	révisions	248,8	vous	154,1	c'	77,5
	propose	243,8	pas	121,5	contre	63,4
	-il	231,0	toi	95,6	Pour	47,7
	Wikipédia	212,7	te	94,5	pas	45,2
	pensez	144,1	accord	91,6	effectivement	43,5
	Bonjour	116,5	Oui	87,8	aussi	42,5
	Quelqu'un	90,9	Tu	85,7	est	39,5
	Qu'	73,5	c'	85,4	pour	38,5
	Est	71,4	as	81,2	rejoins	37,8
	-on	64,2	est	81,0	n'	28,8
Spécificités négatives	tu	-1000,0	propose	-84,1	-il	-82,9
	accord	-1000,0	-il	-66,0	Bonjour	-77,9
	c'	-294,6	signé	-51,2	propose	-77,7
	vous	-281,5	pensez	-50,2	pensez	-48,2
	pas	-275,5	révisions	-42,7	article	-42,1
	Oui	-207,1	la	-41,7	Quelqu'un	-33,9
	est	-202,0	l'	-36,2	suggérée	-32,0
	Pour	-182,6	Quelqu'un	-28,0	wikipédia	-31,9
	effectivement	-178,1	Est	-26,4	Qu'	-31,1
	contre	-163,5	-on	-24,7	transfert	-28,9

Typologie : 10 raisons pour l'intervention de C

— — —

Appui : “+1”, “Oui bien sûr”, “C’est clair !”

Appui et complément : “Bonjour, je suis d'accord avec la structure proposée ici [...] Sur la partie I, je pense qu'on aura du mal [...] donc je suggère qu'on se limite [...]”

Opposition : “Bonjour, je ne suis pas d'accord avec toi. Quels sont les critères de notoriété des plats (ou mets) ? Je ne parle pas de recettes [...]”

Vote : “Pour”, “Contre. Rien à ajouter.”

Rapport d'activité : “Fait !”, “J’ai remis les intitulés des vidéos retirées [...]”

Complément : “Bonjour, je viens mettre mon grain de sel [...] il faut prendre en compte le fait que les 3 premiers ont un nom unique à travers le monde alors que Le Tombeau des lucioles est le titre français. Cherche avec le titre anglais Grave of the Fireflies et tu vas déjà obtenir beaucoup plus de résultats”

Question : “Open source : pourquoi ne pas le dire en français ?”

Réponse : A : “ [...] il se pourrait que les Kurdes aient quitté pendant quelque temps la région puis y soient revenus après la P.G.M. ? une opinion qq ?” B : “Je connais très peu l'histoire kurde [...] Peut-être Bouarf pourra-t-il en dire plus ?” C : “Les kurdes sont présents dans la région depuis longtemps, mais étant nomades, la population a varié [...]”

Ouverture : “En tout état de cause il faut distinguer ce qui est sourcé de ce qui ne l'est pas, ce qui est régulier de ce qui est anecdotique. “

Pacification : “Quoiqu'il en soit, il faut arriver à refléter les sources du mieux que l'on peut. Et dire que l'oeuvre a longtemps était attribuée à Goya mais qu'aujourd'hui le Prado préfère l'attribuer à son élève me semble une présentation neutre”

Echantillons

Deux échantillons sélectionnés (42 discussions dans chaque) :

- 1 : échantillon aléatoire parmi toutes les discussions à 3+ participants
- 2 : échantillon spécifique d'arrivée tardive de C (10e message ou plus tard)
Permet de focaliser sur des fonctions plus spécifiques

Ciblage du message de C

Test : le message de C conserve-t-il son sens si on supprime le ou les message(s) de A ou de B ?

Scission de l'article [\[modifier le code \]](#)

Ne serait-il pas judicieux de scinder l'article en deux, un article pour l'émission anglaise et un autre pour l'émission américaine ? Dans la WP anglophone, il y a également deux articles. Qu'en pensez-vous ? --[Berdea](#) (d) 10 mai 2013 à 01:23 (CEST) [\[répondre \]](#)

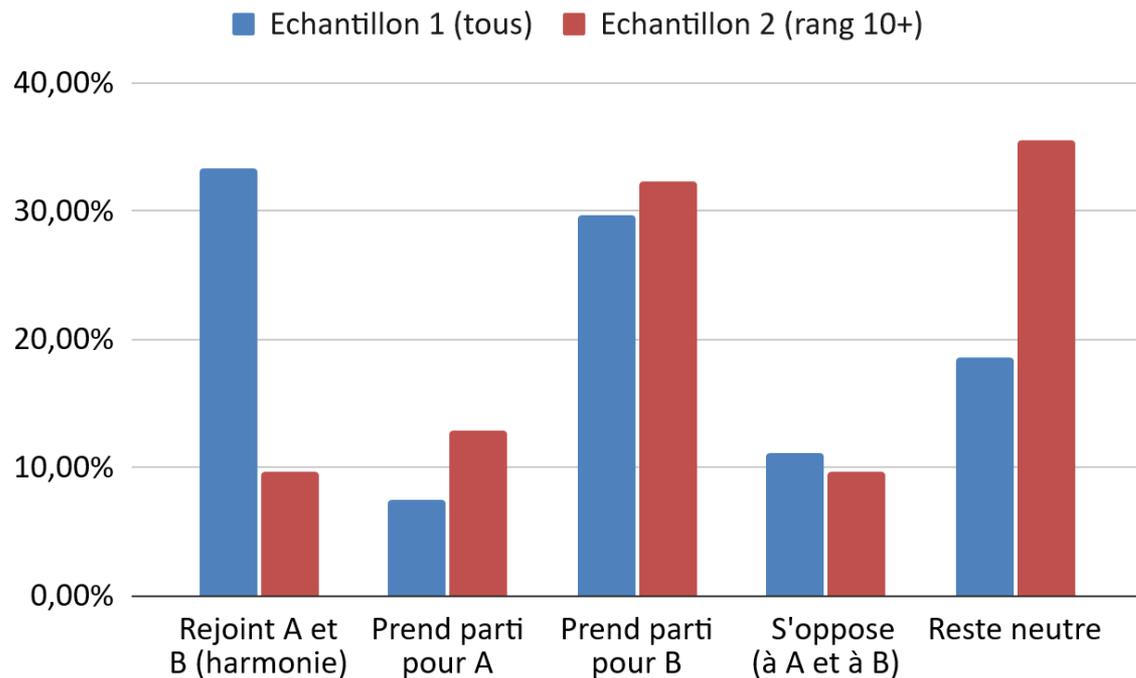
Il est vrai qu'il serait mieux de diviser les anglais et les américains, comme pour les anglophones. [TiboF](#)® 19 juillet 2014 à 16:59 (CEST) [\[répondre \]](#)

Je ne pense pas, le concept, le déroulé sont identiques. Les chaînes rediffusant en France ne font même pas de différences dans leurs annonces de programme. [Askwhyale](#) (discuter) 8 février 2015 à 16:44 (CET) [\[répondre \]](#)

Seule une minorité des messages de C ne s'adresse qu'à un des deux (33% / 27%)

(Kappa = 0,64)

Alignement de C



L'alignement partiel ou total est majoritaire dans les deux échantillons.

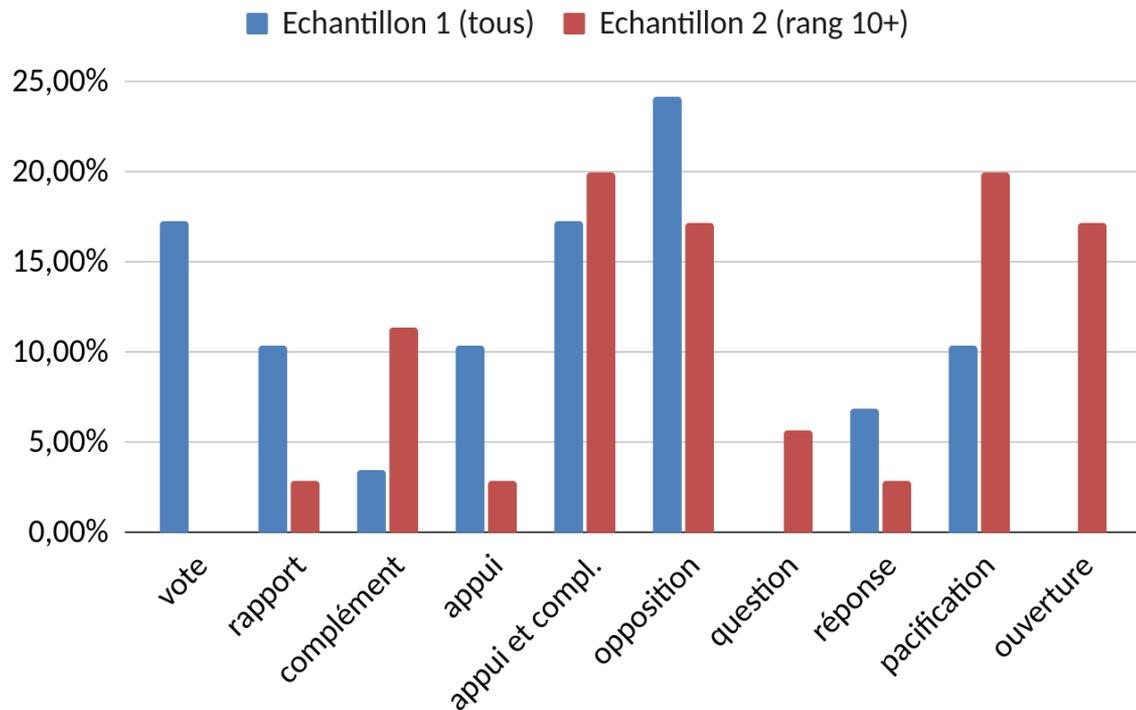
Préférence pour B vs A.

Différences marquées entre les échantillons sur l'harmonie et la neutralité.

kappa = 0,59



Objectif de l'intervention



La prise de position est bien le cas le plus fréquent (49%)
Différences logiques
1 : votes et rapports
2 : complément, ouverture et pacification
 $\kappa = 0,51$

Conclusion

Les discussions Wikipédia: une face cachée mais cruciale

— — —

- Des volumes et des activités très importantes
 - Des contributeurs parfois très productifs et des fils potentiellement très volumineux
 - Des fils qui peuvent s'étendre sur une temporalité très longue
- Des données variées permettant d'examiner différents phénomènes
 - Conflits, monologues, polylogues, argumentation...
 - Mais avec des spécificités de comportement
 - liées au dispositif Wiki, à l'objectif partagé et à la sociologie des utilisateurs

Les discussions Wikipédia: un objet riche et complexe avec ses propres normes

— — —

- Des formes de discussion spécifiques
 - Sous-tendues par une activité argumentative constante
- Des formes conflictuelles particulièrement indirectes
 - Très peu d'insultes directes, des désaccords et des reproches exprimés indirectement
- Mais un espace d'une grande complexité
 - Référence aux pages et aux éditions
 - Des discussions dynamiques qui peuvent se poursuivre d'un espace à l'autre

Méthodologie : une approche mixte

- Analyse quantitative : cas extrêmes et phénomènes inattendus
- Analyse qualitative : examen manuel d'un échantillon
- Reste à faire le voyage retour par annotation automatique
 - quantification des phénomènes identifiés
 - croisement avec d'autres caractéristiques (thèmes, auteurs, périodes, etc.)
 - phénomènes complexes à traiter

Pistes à explorer

- Analyse plus fine des interactions
 - Au-delà des débuts de discussion
 - Lien avec l'édition des articles
 - Activité argumentative des patterns à explorer - Même dans un monologue, une to-do liste ou dans un sondage, le contributeur peut se justifier et argumenter
- Différences entre langues
 - Faibles en terme de grandes tendances et de structures
 - Peuvent apparaître sur des phénomènes plus spécifiques
- Extension à d'autres contextes et données
 - Gros problème : accessibilité de plus en plus réduite des données des réseaux sociaux

Et si on en demandait plus aux LLM ?

Dans le contexte des pages de discussion de Wikipédia, voici quelques raisons spécifiques qui pourraient amener un utilisateur à répondre à son propre message :

1. **Ajout de références ou de sources** : L'utilisateur peut avoir trouvé des sources supplémentaires ou des références qui renforcent ou clarifient son point de vue.
2. **Clarification de propos** : L'utilisateur peut vouloir clarifier une ambiguïté dans son message initial ou reformuler une idée pour qu'elle soit mieux comprise.
3. **Proposition de questions supplémentaires** : Pour encourager le dialogue, l'utilisateur pourrait poser des questions supplémentaires liées au sujet, même sans avoir reçu de réponses.
4. **Mise à jour de la situation** : Si des changements ont eu lieu concernant l'article en question, l'utilisateur peut vouloir les signaler pour informer les autres contributeurs.
5. **Contexte historique** : L'utilisateur peut ajouter un contexte historique ou un précédent pertinent à la discussion pour enrichir le débat.
6. **Inviter à l'action** : Il peut aussi inciter les autres contributeurs à agir sur une question ou un problème soulevé dans le premier message.
7. **Réflexion personnelle** : L'utilisateur peut avoir eu une nouvelle réflexion ou un changement d'opinion sur le sujet après avoir écrit son premier message.
8. **Démontrer l'intérêt pour le sujet** : En ajoutant un second message, l'utilisateur montre qu'il est engagé et souhaite que d'autres s'impliquent également.

Et encore plus ?

Here is a list of possible reasons for self-reply in online discussions based on the extracts you provided:

1. **Adding New Information:** The user may realize they missed important details or found new information after posting the first message, prompting a follow-up reply (e.g., correcting a date or sharing additional sources).
2. **Clarifying or Correcting the Initial Post:** The user may notice an error or ambiguity in their original post and provide a correction or clarification in a second message (e.g., correcting coordinates or dates).
3. **Addressing Another User or Directing a Question:** The user may remember to direct a question or suggestion to another participant after submitting their first post, as seen in some cases where the user asks for input or directs a comment to a specific individual (e.g., "Maybe this is something you can help us out with").
4. **Updating Action or Progress:** Sometimes, the user provides an update on actions taken or changes made, such as reverting edits, modifying content, or implementing decisions based on feedback (e.g., "I will change back to the MoS-compliant version").
5. **Reassessing a Position:** The user might reconsider their stance or action and either withdraw or modify it based on new insights, as in the case of correcting conversion errors or withdrawing a claim (e.g., "withdrawn, it's correct, made a conversion error").
6. **Soliciting Feedback or Confirming No Objections:** Users may follow up to ensure there are no objections to a proposed action, giving others a chance to intervene or comment before they proceed (e.g., "If there will be NO objections I will remove the size tomorrow").
7. **Engaging in Self-Reflection or Humor:** Occasionally, users may engage in self-reflection, pointing out their own mistakes or adding a humorous or self-deprecating remark about an earlier post (e.g., "Oh, this is amusing. Apparently the unsourced 1963 establishment date was added by me").
8. **Reiterating or Strengthening an Argument:** In some cases, the user adds further reasoning, examples, or sources to bolster their original point, showing they are committed to their argument and want to provide additional evidence (e.g., "To further my argument...").
9. **Attempting to Prompt Discussion:** The user may post again to stimulate further discussion or to encourage others to participate in the conversation, especially if there hasn't been an immediate response (e.g., "I wanted to hear his thoughts").
10. **Switching Topics:** The user may reply to themselves to introduce a new, sometimes unrelated, topic that came to mind after their first post, often marked by an off-topic notation (e.g., "Off Topic").

Pour plus de détails

— — —

*Investigating Wikipedia:
linguistic corpus building,
exploration and analyses*

Poudat, Lungen, Herzberg
(eds)

Studies in Corpus Linguistics



Studies in Corpus Linguistics

General Editor  **Ute Römer** | Georgia State University

Founding Editor **Elena Tognini-Bonelli** | The Tuscan Word Centre/University of Siena

SCL focuses on the use of corpora throughout language study, the development of a quantitative approach to linguistics, the design and use of new tools for processing language texts, and the theoretical implications of a data-rich discipline.

A paraître fin 2024!