

Deux exemples d'exploration d'opinions multiples

Céline POUDAT

UMR 7187 LDI, Université de Paris 13

cpoudat@gmail.com

Groupe de travail Nouv-com, IR corpus-écrits, séminaire en ligne du 31 mai 2012

Deux exemples d'analyse

- Analyse des controverses dans Wikipédia (Collab. Télécom ParisTech et Limsi, projet Autograph)
- Régulation des avis sur le web marchand : la critique littéraire sur Amazon.fr (collab. Legallois, projet Confidence)

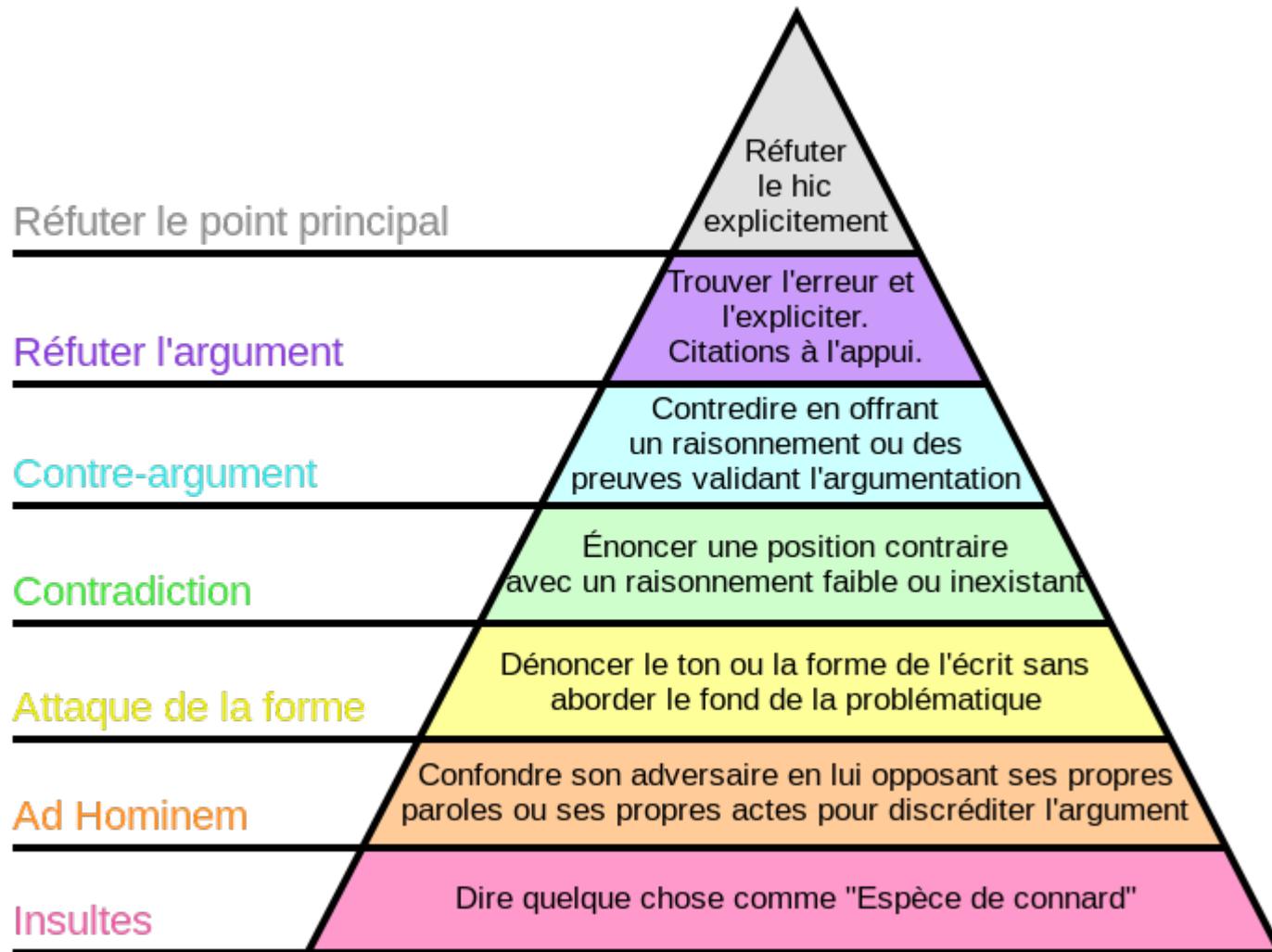
1. Analyse des controverses dans Wikipédia

*[Auray, Hurault-Plantet, Poudat & Jacquemin 2009]
[Jacquemin, Lauf, Poudat, Hurault-Plantet & Auray
2008a, 2008b] [Auray, Poudat & Pons 2007] [Poudat &
Loiseau 2007]*

Wikipédia

- Une ressource encyclopédique de référence
 - Quatre principes idéologiques forts
 - Projet encyclopédique couvrant tous les domaines du savoir
 - Logiciels et licences libres (*open source*)
 - Principe de la *neutralité de point de vue* (NPOV), garantie de l'objectivité
 - Gestion collective → mise en place progressive de divers instruments et procédures de régulation et de contrôle
 - Qualité discutée (Giles 2005, Gourdain et al. 2007...)
- ➔ **Qualité = non conflictualité?**

Hiérarchie du désaccord de Graham



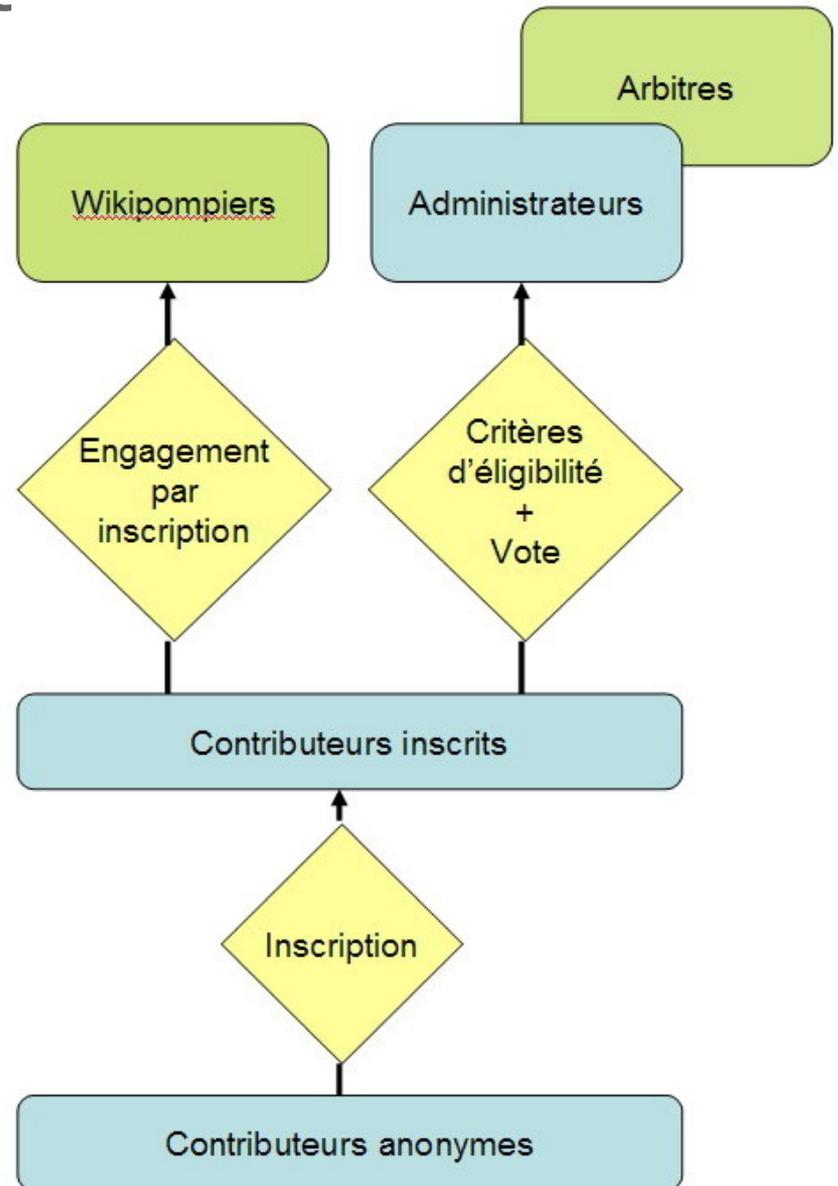
Les acteurs du conflit

Deux grandes catégories de participants

- les médiateurs, ou arbitres de la dispute (en vert)
- et ses impliqués potentiels (en bleu)

Nota Bene

- *wikipompiers* (éteindre les feux, conflits d'édition, supprimés en 2009)
- → **Salon de médiation**
http://fr.wikipedia.org/wiki/Wikip%C3%A9dia:Le_salon_de_m%C3%A9diation



Les lieux du conflit

- Les articles
- Les pages de discussion associées aux articles
- Les arbitrages

Les articles

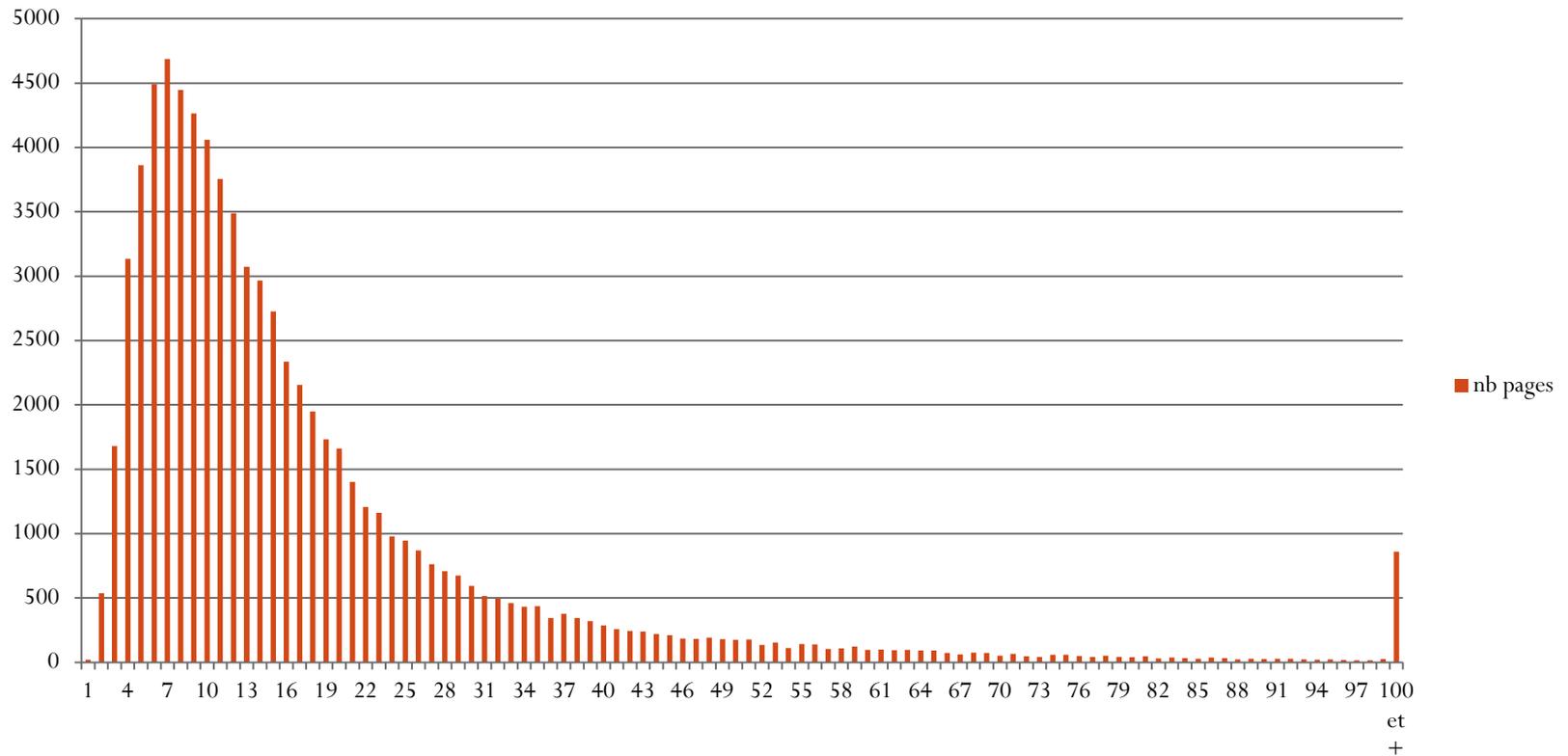
- Actes d'édition antagonistes
- Révocations (*reverts*)
 - 5% des éditions sont des révocations
 - 3% des articles ont fait l'objet d'au moins 3 reverts
 - l'article *homéopathie* a un taux de 10% de révocations effectuées par 18 contributeurs distincts
- Bandeaux
 - « articles soupçonnés de partialité » (ou « articles non neutres »)
 - 1000 articles non neutres en juin 2006
 - « articles protégés »

Pages de discussion

- Plus de vingt catégories de pages de discussion (*namespaces*)
 - *discuter :article*
 - *discussion :utilisateur*
- 93,38% des pages discussion
- 88,68% des contributions effectuées sur l'ensemble de ces pages

Un peu de quantitatif...

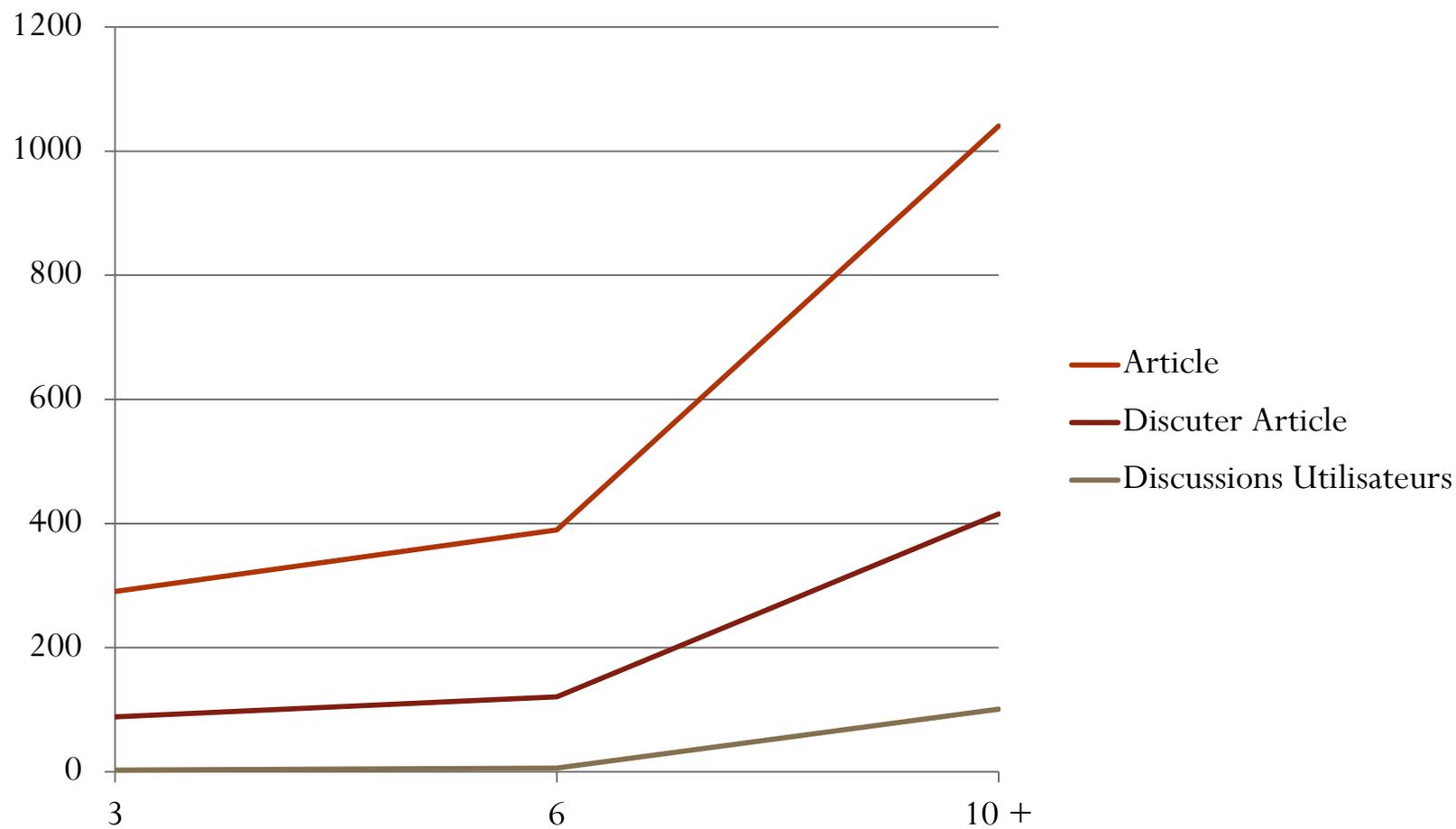
Distribution des articles hors redirections et > 1 an par nombre de contributeurs



Un peu de quantitatif...

Nombre de contributeurs (dans l'article)	3	6	10 et plus
Nombre moyen de mots par article	290,79	389,82	1040,47
Nombre moyen de caractères par article	1737,67	2295,52	6211,68
Nombre moyen de mots par page discuter:article	88,55	121,14	415,71
Nombre moyen de caractères par page discuter:article	501,63	690,63	2229,57
Nombre moyen de mots par page discussion:utilisateur	2,93	5,96	101,00
Nombre moyen de caractères par page discussion:utilisateur	15,88	33,15	655,85
Nombre moyen de mots toutes discussions	91,48	127,10	516,70

Un peu de quantitatif...



Un exemple d'arbitrage rendu

Description du conflit

Demande déposée par : [Guiggz](#) le 3 février 2012 à 11h25.

Parties concernées par le conflit : [Guiggz](#) (d · c · b) et [Floflo62](#) (d · c · b).

Nature du conflit : Reverts abusifs de la part de [Floflo62](#) (d · c · b).

Recevabilité : Déclaré recevable le 13 février 2012 à 03:25 (CET) par 4 voix pour, 1 voix contre et 1 récusation.

Décision :

Considérant :

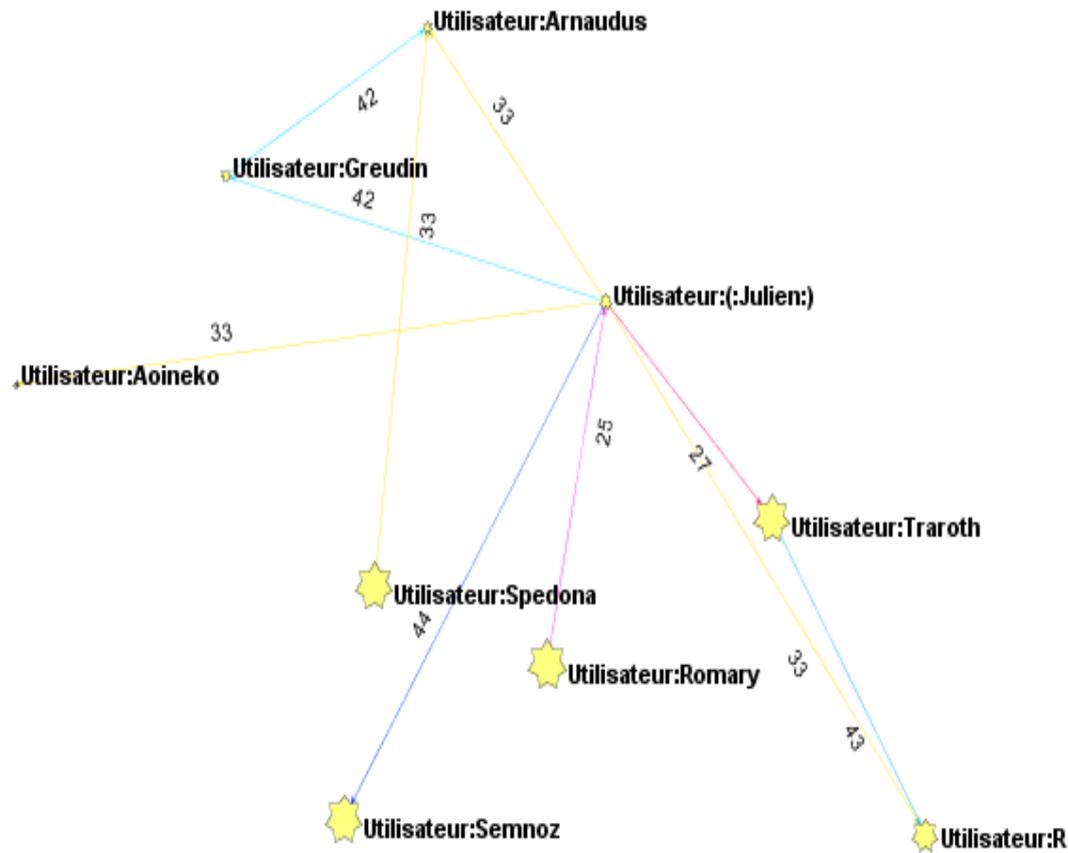
- que [Guiggz](#) (d · c · b) a tenté à plusieurs reprises de faire valider un type de contenu sur les clubs de football, qui n'a obtenu ni l'appui du [Projet:Football](#), ni un soutien après demande sur [Wikipédia:Le Bistro](#) ;
- que le [comité d'arbitrage](#) n'est pas compétent pour se prononcer sur les questions éditoriales et ne peut que vérifier si la discussion a eu lieu, ce qui est le cas, que, par suite, les révocations de [Floflo62](#) (d · c · b) sont justifiées ;
- que [Guiggz](#) (d · c · b) s'est livré, durant ses tentatives de modifier la ligne rédactionnelle, mais aussi durant le présent arbitrage, à des requêtes abusives, notamment auprès des administrateurs, ainsi qu'à des [attaques personnelles](#) à l'égard de [Floflo62](#) (d · c · b) ;
- que [Guiggz](#) (d · c · b) apporte un contenu intéressant à des articles sur le football, mais qu'il finit par lasser les autres membres du projet par son obstination ;
- que le comité d'arbitrage ne reproche rien à [Floflo62](#) (d · c · b), mais qu'il convient d'empêcher toute provocation entre [Guiggz](#) (d · c · b) et lui, du moins durant un temps ;
- que [Guiggz](#) (d · c · b) a finalement été bloqué pour deux mois le 19 mars,

Le Comité d'arbitrage :

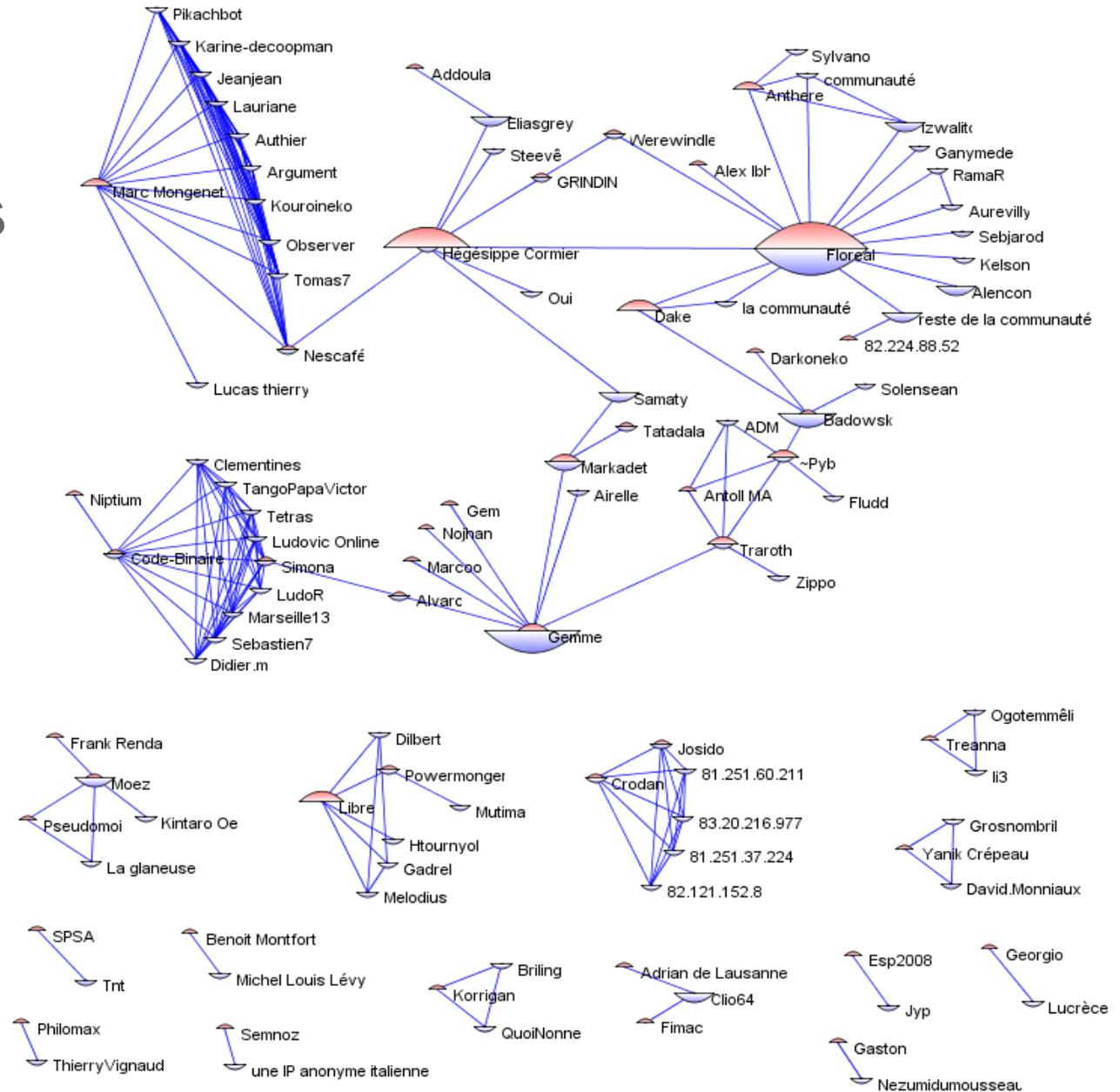
- défend à [Guiggz](#) (d · c · b) de contribuer aux sujets concernant le football, ainsi qu'aux pages de discussion de ces articles et au [café du foot](#), pour une durée de 3 mois prenant effet à compter de la fin de son blocage, et ce à peine d'un blocage d'une durée d'une semaine, pouvant être doublée en cas de récidive ;
- lui défend pareillement de relancer, sur tout autre espace de discussion, des propositions relatives aux résultats des clubs, pour une durée de 6 mois prenant effet à compter de la fin de son blocage, et ce à peine d'un blocage d'une durée d'une semaine, pouvant être doublée en cas de récidive ;
- demande à [Guiggz](#) (d · c · b) et à [Floflo62](#) (d · c · b) de ne pas se révoquer mutuellement, de ne pas écrire sur la page de discussion l'un de l'autre, et de ne pas dénigrer l'autre dans une autre discussion, pendant une durée de 7 mois à compter de la présente décision, et à peine d'un blocage d'une durée de 15 jours pour [Guiggz](#) (d · c · b), et d'un jour pour [Floflo62](#) (d · c · b).

Pour le CA, [O. Morand](#) (d) 20 avril 2012 à 23:39 (CEST)

Arbitres: sous-graphe des désaccords (seuil < 50%)



Graphe des protagonistes du CA



Ex. Floreal

Floreal contre Alençon

- (actu) (dern) 13 jun 2005 à 23:50 Floreal m (élimination d'un parasite.)
- (actu) (dern) 13 jun 2005 à 23:19 Alençon (Revert d'une troisième tentative dans la même journée de truquage de discussion ; cette fois-ci par destruction d'historique)
- (actu) (dern) 13 jun 2005 à 23:03 Floreal m (Suppression d'un parasite)
- (actu) (dern) 13 jun 2005 à 23:01 Alençon m (Tu prétends ce matin que c'est moi le vandale (sic)... En attendant, c'est quand même toi la multi-récidiviste pour blanchir les discussions qui ne te plaisent pas)
- (actu) (dern) 13 jun 2005 à 22:35 Floreal m
- (actu) (dern) 13 jun 2005 à 17:46 Floreal m (Réponse à un parasite.)
- (actu) (dern) 13 jun 2005 à 17:46 Floreal m (Réponse à un parasite.)

Page *Féminisme*

- ::::Il est triste de voir un être humain tomber dans de tels tréfonds. J'espère pour toi que tu ne te rends pas compte combien bas tu es tombée. Tu m'as fait perdre assez de temps, et comme déjà dit plus tôt, je retourne à ce qui devrait t'occuper: la rédaction d'articles pour l'encyclopédie. Et je veux dire articles encyclopédiques, pas des pamphlets grotesques. [[Utilisateur:Alex lbh | Alex lbh]] 6 août 2005 à 00:50 (CEST)
- ::::Il est surtout triste de voir comment un puits d'ignorance tel que toi mais doté d'un arrivisme forcené doublé d'une arrogance sans limite et d'une bassesse d'âme prête à toutes les hypocrisies et aux compromis les plus odieux réussisse à imposer sa loi: les parfaits ingrédients de l'inquisiteur et du dictateur en herbe. Pense à la carrière qui s'ouvrirait à toi dans certains pays! Tu serais un chef de bande révééré!-- [[Utilisateur:Floreal | fl0]] 6 août 2005 à 08:45 (CEST)fl0

2. La critique littéraire sur Amazon.fr

*Régulation des avis sur le web marchand : la critique
littéraire sur Amazon [Legallois&Poudat 2008]*

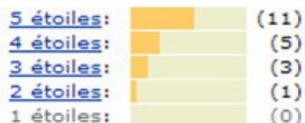
Amazon

- Création en juillet 1995 par Jeff Bezos
- 2000: ouverture de la filiale française, Amazon.fr
- Système d'évaluation collaborative (*collaborative peer review*) articulé autour des produits mis en vente
 - Chaque livre possède ainsi son évaluation moyenne propre (étoilage)
 - Chaque critique, ou *commentaire*, est pourvue des métadonnées suivantes : *auteur, étoilage, titre, date de publication et 'utilité', obtenue par les votes des clients inscrits sur le site.*

Commentaires en ligne

La chambre des curiosités

20 évaluations



Moyenne des commentaires client

★★★★☆ (20 commentaires client)

Partagez votre opinion avec les autres clients

Créer votre propre commentaire

Evaluation moyenne du livre

Rechercher des commentaires

Rechercher uniquement parmi les commentaires portant sur ce produit

GO

Le commentaire favorable le plus utile

26 internautes sur 26 ont trouvé ce commentaire utile :

★★★★★ un must

j'avoue, j'ai eu du mal à l'ouvrir... mais une fois fait impossible de le refermer!

tout y est : le suspense, les fausses vraies pistes, une psychologie des personnages qui laisse sans voix, un voyage historique dans manhattan documenté sur le bout des doigts, une écriture qui coule, bref... un des meilleurs policiers que j'ai jamais lu avec une intrigue palpitante...

[Lire toute l'évaluation >](#)

Publié il y a 14 mois par hayabuza90

> Voir plus de commentaires

Le commentaire critique le plus utile

12 internautes sur 18 ont trouvé ce commentaire utile :

★★★★☆ Faible curiosité

Il est difficile d'entretenir le suspens sur 700 pages, il faut beaucoup de génie pour y parvenir.

La Chambre des curiosités est remarquablement écrite et les personnages attachants, cela ne suffit pas, ça se traîne, ça manque de punch. Il faut être très indulgent pour donner du crédit à cette histoire d'assassin en série qui traverse le temps. Une fois que le...

[Lire toute l'évaluation >](#)

Publié le 24 août 2006 par Hano Nyme

Vs.

1 étoile, 1 étoile

[plus récent au plus ancien](#)

26 internautes sur 26 ont trouvé ce commentaire utile :

★★★★★ un must, 20 juin 2007

Par [hayabuza90](#) (Belfort, France) - [Voir tous mes commentaires](#)

j'avoue, j'ai eu du mal à l'ouvrir... mais une fois fait impossible de le refermer!

tout y est : le suspense, les fausses vraies pistes, une psychologie des personnages qui laisse sans voix, un voyage historique dans manhattan documenté sur le bout des doigts, une écriture qui coule, bref... un des meilleurs policiers que j'ai jamais lu avec une intrigue palpitante qui nous pousse à se prendre au jeu de l'enquête!

Aidez d'autres clients à trouver les commentaires les plus utiles

Avez-vous trouvé ce commentaire utile ?

Oui Non

Après un démarrage presque poussif, même si c'est bien écrit, la suite est vraiment excellente. En effet, c'est au moment où l'on se demande pourquoi on l'a acheté (honnêtement, pour moi c'est parce que les commentaires des autres lecteurs étaient très

Problématique

- Est-ce que l'annotateur d'avis fait confiance à la masse ou met en place des mécanismes de confiance dans des sites d'annotation collaborative ?
- Amazon: commentaires ouverts
 - Problème: lexique non contraint qui entrave, voire invalide l'application d'un traitement automatique
 - ➔ Annotation semi-automatique des **valeurs** fondant l'évaluation littéraire sur Amazon – intérêt d'XML et d'Oxygen pour l'annotation.
- Relation entre *commentaires, contributeurs et étoilage*

Corpus de critiques

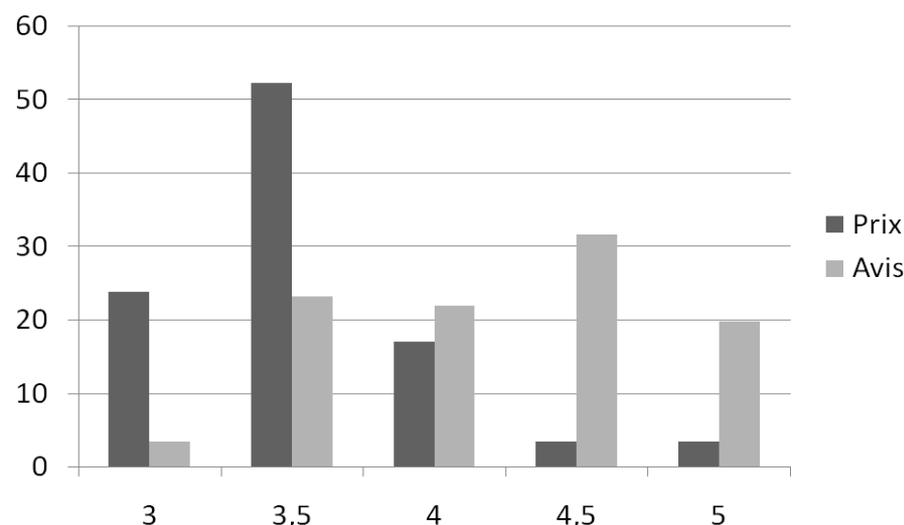
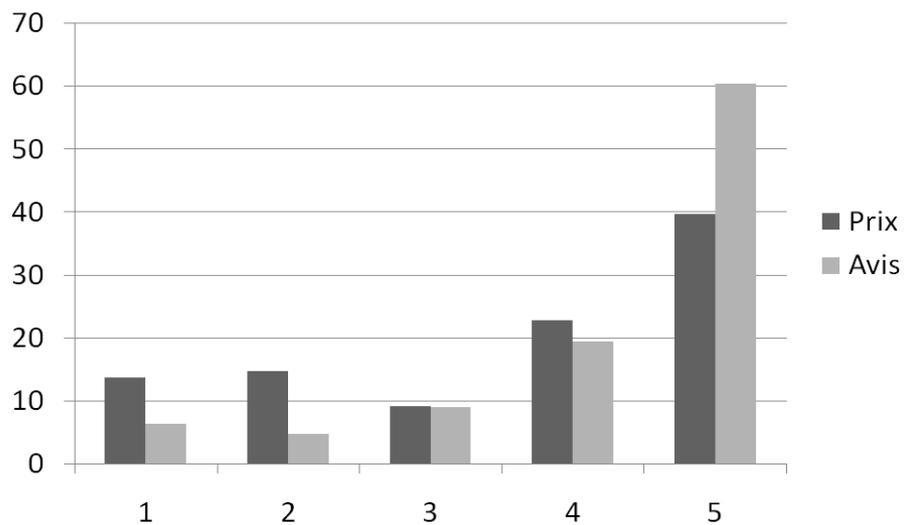
Corpus AVIS

- 21 œuvres classiques et contemporaines
 - *L'Homme qui voulait vivre sa vie* (29 critiques), *Les Bienveillantes* (24), *Les particules élémentaires* (23), *Si c'est un homme* (21), *Le Pendule de Foucault* (20), *Plateforme* (20), *Voyage au bout de la nuit* (18), *Cent ans de Solitude* (17), *Les Trois mousquetaires* (11), *Ravel* (11), *Au Bonheur des dames* (10), *Eugénie Grandet* (10),...
- 319 critiques publiées entre août 2000 et juillet 2008
- 283 contributeurs distincts (dont 34 anonymes)

Corpus PRIX

- 7 œuvres primées en 2007 par la critique littéraire professionnelle
 - *Alabama Song* de Gilles Leroy, Goncourt (21 critiques)
 - *Ap. J.-C.* de Vassilis Alexakis, Grand Prix du Roman de l'Académie Française (3 critiques)
 - *Baisers de cinéma* d'Eric Fottorino, Femina (5)
 - *Birmane* de Christophe Ono-Dit-Biot, Interallié (7)
 - *Chagrin d'école* de Daniel Pennac, Renaudot (39)
 - *Ouest* de François Vallejo, Prix du Livre Inter (10)
 - *La stratégie des antilopes* de Jean Hatzfeld, Médicis (3)

Évaluation associée aux critiques et aux œuvres du corpus – distribution comparée exprimée en pourcentages – corpus Prix et Avis

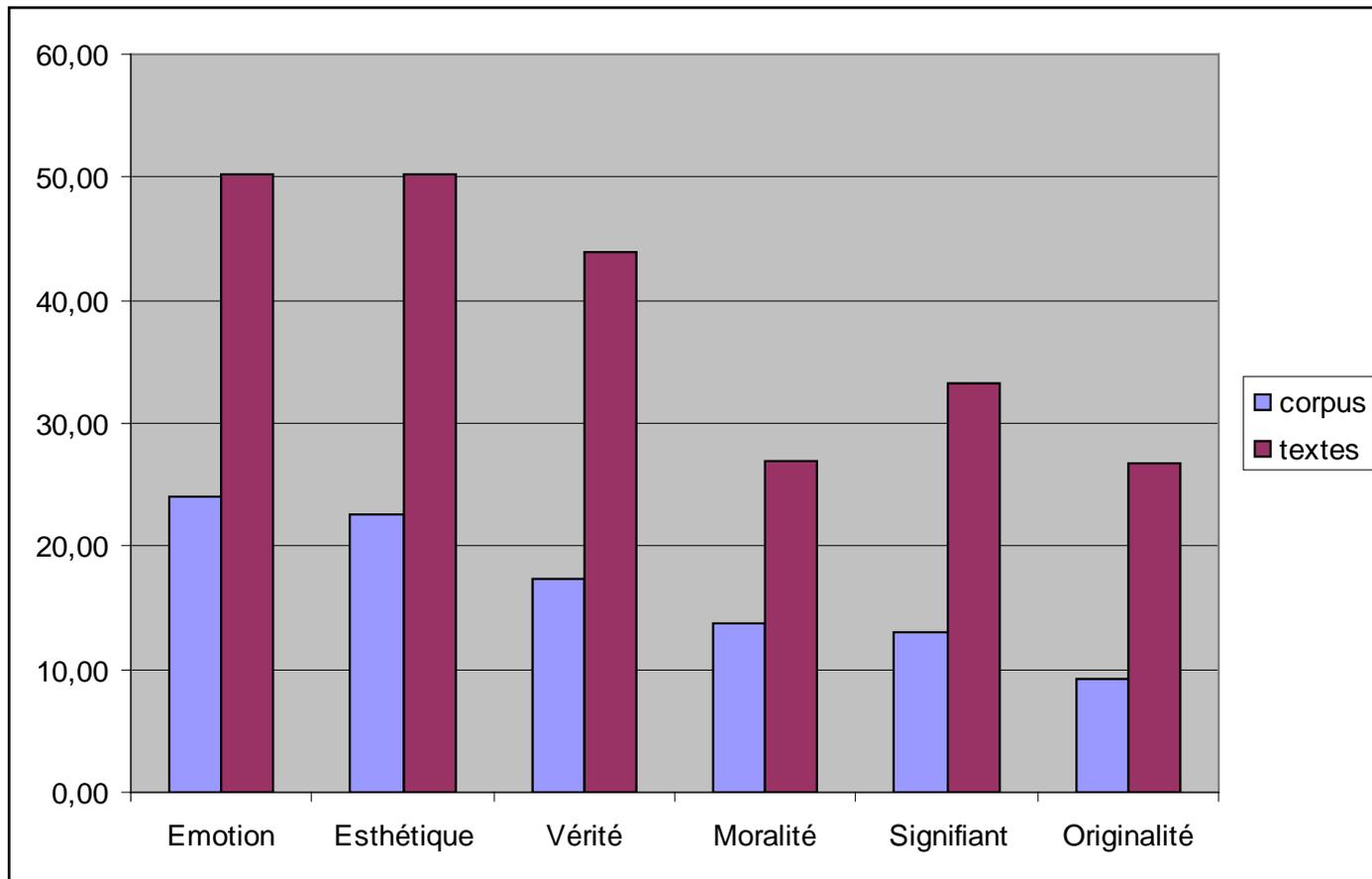


Catégorie	Sous-catégorie	Descriptif
Émotion	horreur	Effet psychologique
	rire	Effet psychologique
	tristesse	Effet psychologique
	attachement	Effet psychologique
	identification	Réaction transréférentielle
	happage	Effet psychologique
Esthétique	intrigue	Construction de l'intrigue
Originalité	genre	Fidélité du texte aux canons du genre
Signifiant	accessibilité	Facilité de lecture
	édition	Clarté de l'édition
	brièveté	Effort de lecture
Vérité	historique	Conformité au fait historique
	personnage	Réalité des personnages
Citation		Valeur démonstrative / intertexte
Inhibition		Difficulté (rhétorique) à s'exprimer
Intertexte		Référence à d'autres œuvres, d'autres auteurs
Lecture		Lecture d'œuvres du même auteur
Prescription		Recommandation du livre
Relecture		Relecture du livre

Annotation XML

```
- <text>
- <header>
  <post>22</post>
  <auteur>VIAN</auteur>
  <titre>J'irai cracher sur vos tombes</titre>
  <auteur_post>anonyman</auteur_post>
  <eval_book>4</eval_book>
  <nb_com>12</nb_com>
  <rang>37350</rang>
  <com>1</com>
  <nb_votes_utiles>3</nb_votes_utiles>
  <eval>2</eval>
  <eval_com>03/10</eval_com>
  <titre_com>direction Trashland</titre_com>
  <date>15 septembre 2003</date>
  <num_oeuvre>2</num_oeuvre>
</header>
- <body>
- <p>
  "J'irai cracher sur vos tombes" est
  <valeur type="originalité">un livre surprenant</valeur>
  :
  <valeur type="signifiant" subtype="accessibilité">le style est simple donc facile à lire</valeur>
  mais
  - <valeur type="émotion" subtype="horreur">
    le dénouement est insoutenable, voire vraiment immonde
  </valeur>
  .
  - <valeur type="moralité">
    Le combat de ce Noir "blanc" est louable et son envie de vengeance est aussi la notre mais il y a d'autres façons de se venger ou plutôt de décrire une vengeance...
  </valeur>
  - <valeur type="prescription" pol="neg">
    Un livre à ne pas mettre entre toutes les mains donc.
  </valeur>
</p>
</body>
</text>
```

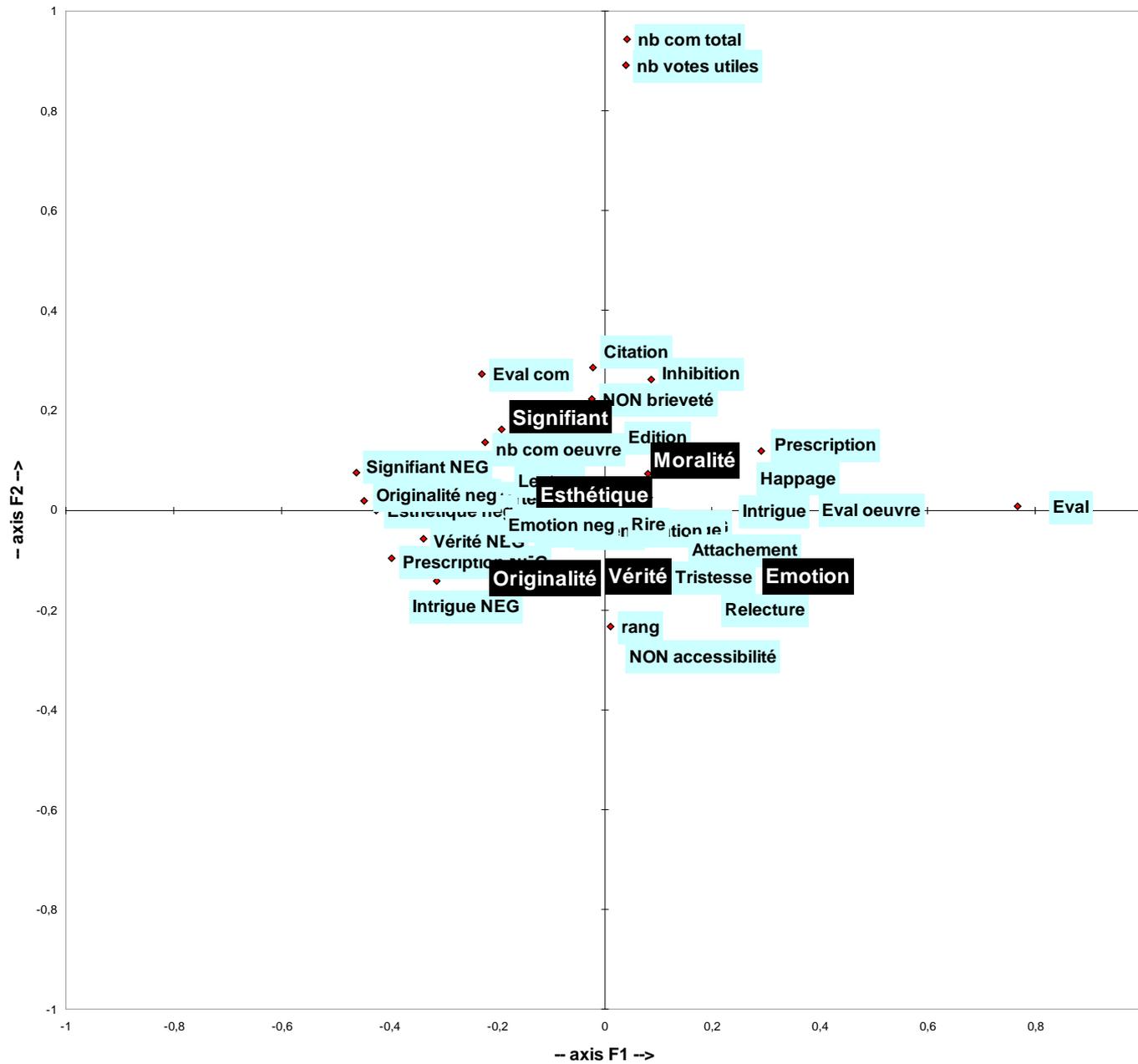
Les valeurs privilégiées



Les critères privilégiés

- L'intrigue
 - une intrigue évaluée comme pauvre entraîne ainsi une recommandation négative de l'œuvre (corrélation +0.30)
 - Forte corrélation avec le happage
 - statut ambigu de l'intrigue, construction formelle plébiscitée ici non pour sa forme, mais pour l'illusion référentielle qu'elle permet

Variables (axes F1 and F2)



Commentaires, commentateurs et étoilages

- Évaluation de l'œuvre et commentaires
 - Plus l'œuvre est positivement évaluée (étoilage élevé), moins elle est commentée
- Autorité du commentateur
 - Plus l'auteur a un rang élevé, plus ses commentaires sont bien évalués
 - rang = confiance
- Une œuvre très commentée génère plus de commentaires à la critique postée
 - stratégie pour avoir de l'autorité sur Amazon : ne commenter que les œuvres les plus commentées

Perspectives

Perspectives de recherche nouv-com

- Wikipédia
 - Conflits
 - Articles de recherche / articles encyclopédiques (académique / vulgarisé)
 - Sciences du langage
- Typologies narratives et narrations de soi
 - Exploration quantitative (analyse de données textuelles)
 - Réseaux sociaux
 - Blogs, sites de rencontre

Références

- AURAY, Nicolas, HURAUULT-PLANTET, Martine, POUDAT, Céline et JACQUEMIN, Bernard (2009). « La négociation des points de vue. Une cartographie sociale des conflits et des querelles dans le Wikipédia francophone » in Cardon, D. (coord.), « Web 2.0. », *Réseaux* n°154, 2009-2, pp. 15-50.
- JACQUEMIN, Bernard, LAUF, Aurélien, POUDAT, Céline, HURAUULT-PLANTET, Martine et AURAY, Nicolas (2008). “Managing conflicts between users in Wikipedia” in Flejter, Dominik, *et al.* (éds.), BIS 2008 Workshops Proceedings: Social Aspects of the Web (SAW 2008), Advances in Accessing Deep Web (ADW 2008), E-Learning for Business Needs. Innsbruck, Autriche, CEUR-WS.org, 81-93.
- JACQUEMIN, Bernard, LAUF, Aurélien, POUDAT, Céline, HURAUULT-PLANTET, Martine et AURAY, Nicolas (2008). « La fiabilité des informations sur le Web : le cas Wikipédia » in Actes de CORIA 2008, Trégastel, 12-14 mars 2008, pp. 449-456.
- LEGALLOIS, Dominique et POUDAT, Céline (2008). « Comment parler des livres que l’on a lus ? Discours et axiologie des avis des internautes » in Bertelli, D. et Chauvin-Vileno, A. (coord.), « De la médiacritique culturelle comme métadiscours. Objets, genres, dispositifs », *Semen*, n°26, pp.49-80.
- AURAY, Nicolas, POUDAT, Céline et PONS, Pascal (2007). « Democratizing scientific vulgarization. The balance between cooperation and conflict in French Wikipedia » in *Observatorio*, Vol. 1, n°3.
- POUDAT, Céline et LOISEAU, Sylvain (2007). « Représentation et caractérisation lexicale des sciences dans Wikipédia » in Tutin, A. (coord.), « Lexique de la langue scientifique », *Revue Française de Linguistique Appliquée*, XII, 2007-2, pp. 29-44.