

# Intégrité scientifique (2)

---

Karën Fort

[karen.fort@sorbonne-universite.fr](mailto:karen.fort@sorbonne-universite.fr) / <https://members.loria.fr/KFort/>

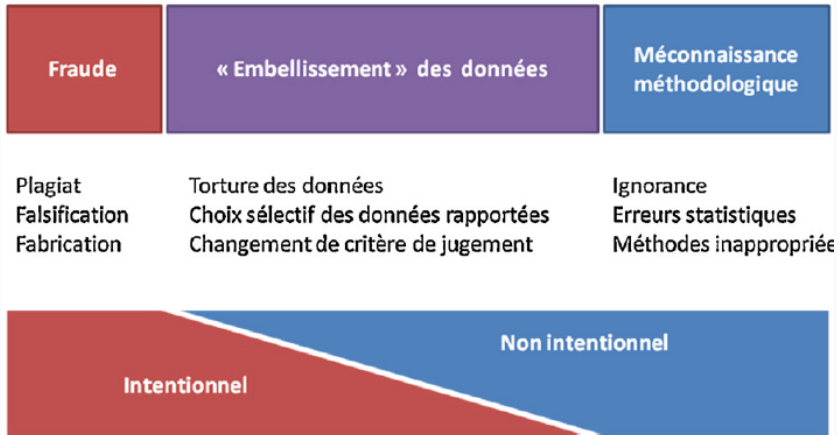
## Sources d'inspiration

- Atelier Intégrité scientifique. N. Heck (SU).
- Promouvoir et protéger une culture partagée de l'intégrité scientifique. P. Ouzoulis (sénateur) et P. Henriot (député), 2021.
- L'intégrité scientifique dans tous ses états : recherche, publications et enseignement. J. Haiech (Univ. Strasbourg), C. Forest (Univ. de Paris), 2021.
- Pour la promotion de l'Éthique de la Recherche et de l'Intégrité Scientifique. C. Strube (CNRS).
- Introduction à l'intégrité scientifique. A. Serres (Rennes 2), 2021.
- Bilan et propositions de mise en œuvre de la charte nationale d'intégrité scientifique. P. Corvol (collège de France), 2016.

# Les différents types de méconduites

---

# Un continuum



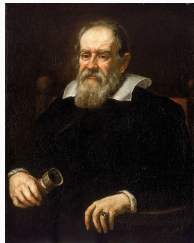
Seror, Ravaud. Presse Médicale, septembre 2012, via H. Maisonneuve  
<http://www.oncorea.com/Presentations/13e%20Rencontre/Maisonneuve.pdf>

*« il convient de différencier l'erreur (un manquement involontaire), de la fraude (un manquement intentionnel, dans l'espoir d'en tirer un éventuel bénéfice), de la faute (un manquement conscient). »  
[Henriet and Ouzoulias, 2021]*

## Fabrication

**création** de toutes pièces de données ou de résultats qui n'ont en vérité pas été mesurés ou observés. Par exemple, fausses données statistiques, ou manipulation d'images.

[Henriet and Ouzoulis, 2021]



Justus Sustermans -  
Portrait of Galileo  
Galilei

*« [...] aucun de ceux qui ont tenté de répliquer les résultats de ses expériences sur la chute des corps n'y est parvenu, du moins avec la précision proclamée par Galilée. Il semble donc que, [...] Galilée ait fait passer une brillante et féconde expérience de pensée pour un véritable travail empirique. Et pour cela, il aurait tout simplement fabriqué des données. » [Y. Pigenet, [journal du CNRS](#)]*

# Manquements graves (FFP) : la falsification

## Falsification

« **manipulation** de procédures de recherche, transformation ou omission de données » (Rapport [Alix](#)). Par exemple, laisser de côté un résultat qui n'irait pas dans le sens de l'article.

[Henriet and Ouzoulis, 2021]





Gregor Mendel

*« Le moine tchèque Johann Gregor Mendel est considéré comme l'inventeur de la génétique moderne. [...] les résultats qu'il a publiés, à l'époque dans l'indifférence générale, sont statistiquement trop parfaits pour être vrais. Du moins avec la taille de ses échantillons. Il est donc très probable que le père de la génétique ait arrangé, donc falsifié, ses données. » [Y. Pigenet, [journal du CNRS](#)]*

### Plagiat

« **appropriation** des données d'autrui, de résultats de recherche ou de mots sans leur octroyer le crédit approprié » (Rapport [Alix](#)).

# Manquements graves (FFP) : le plagiat



Claude Ptolémée

*« Claude Ptolémée est réputé avoir été le plus grand astronome de l'Antiquité [...] Il appuyait son modèle sur des mesures astronomiques qu'il prétendait avoir réalisées lui-même sur la côte égyptienne au IIe siècle. Des mesures qui, en fait, avaient été effectuées 300 ans plus tôt par l'astronome grec Hipparque sur l'île de Rhodes. » [Y. Pigenet, [journal du CNRS](#)]*

# Les pratiques douteuses de recherche sur les données [Henriet and Ouzoulis, 2021]

Y compris :

- embellissement de données
- *salami slicing* ou saucissonnage
- mauvaises pratiques liées aux données (utilisation sélective, rétention ou non conservation de données)

# Les pratiques de recherche inappropriées [Henriet and Ouzoulis, 2021]

Violation des protocoles liés à l'expérimentation :

- chez l'homme,
- sur les animaux
- sur l'environnement

# Les pratiques douteuses liées aux publications

## [Henriet and Ouzoulis, 2021]

Y compris :

- hyperpublication
- citation d'un article sans vérifier la source
- indiquer un article en bibliographie sans l'avoir lu dans son entièreté
- conflits d'auteur

- la « trahison de la vérité » (fabrication et falsification)
- la « trahison de la confiance » (plagiat, non reproductibilité, auteurs fantômes, masquage de données, etc)
- le « risque pour la sécurité »

<https://retractionwatch.com/2018/12/17/is-it-time-for-a-new-classification-system-for-scientific-misconduct/>

## Exemple (de quoi ?) : l'affaire Andrew Wakefield

- 1998 : publication (avec 11 co-auteurs) dans *The Lancet* liant les vaccins ROR à l'autisme
- 2004 : B. Deer (*Sunday Times*) met au jour un conflit d'intérêts (financement anti-vax), retrait des co-auteurs
- 2010 : enquête du *British General Medical Council* : plus de 30 inculpations (dont 4 pour malhonnêteté et 12 pour abus sur enfants)



Andrew Wakefield with  
Justyna Socha, Warsaw, 2019  
(Bładość)



## L'affaire Andrew Wakefield : conséquences

Pour le chirurgien :

- rétractation de l'article du *Lancet*
- renvoyé de l'ordre des médecins (interdit d'exercer au Royaume Uni), s'exile aux États-Unis
- devient un activiste anti-vaccin aux États-Unis

Pour le monde :

- déclin de la vaccination au Royaume-Uni et augmentation des cas de rougeoles (et des décès liés)
- diminution de taux de vaccination contre la rougeole au Minnesota de 92 % en 2004 à 42 % en 2014 (et épidémie de rougeole aux États-Unis)

## Le plagiat

---

## Plagiat

« **appropriation** des données d'autrui, de résultats de recherche ou de mots sans leur octroyer le crédit approprié » (Rapport [Alix](#)).

## Les différents types de plagats [Serres, 2021]

- copie de textes (y compris après traduction)
- vol de données
- vol de productions culturelles (par exemple, la musique)
- autoplaiat
- *salami-slicing* ou saucissonnage (pour multiplier les publications)

- Plagiat dans une thèse : motif d'invalidation de la thèse ⇒ retrait du titre de docteur
- Motif de rétractation d'articles
- Sanctions disciplinaires possibles (pour les enseignants-chercheurs)
- Sanctions judiciaires possibles (si plagiat qualifié en contrefaçon)

# La contrefaçon : quand le plagiat devient un délit

## Contrefaçon

« Est également un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une oeuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi. » (Code de la Propriété Intellectuelle, article L335-3)

« si la personne reprend la forme de votre œuvre, la même composition d'écriture, il y a contrefaçon. »

[https://www.avocats-picovschi.com/qu-est-ce-que-le-plagiat\\_article\\_836.html#intertitre2](https://www.avocats-picovschi.com/qu-est-ce-que-le-plagiat_article_836.html#intertitre2)

# Archéologie du "copier-coller"

[Accueil](#)[Tous les articles](#)[Plagiats et procès](#)[Wikiplagiat](#)[Les plagiés réagissent](#)[Presse](#)

Première instance, appel, cassation : Béatrice Durand (Univ. libre de Berlin) gagne définitivement contre Christophe Martin (Paris4 – Sorbonne), condamné pour contrefaçon et parasitisme.

Posté par **admin**

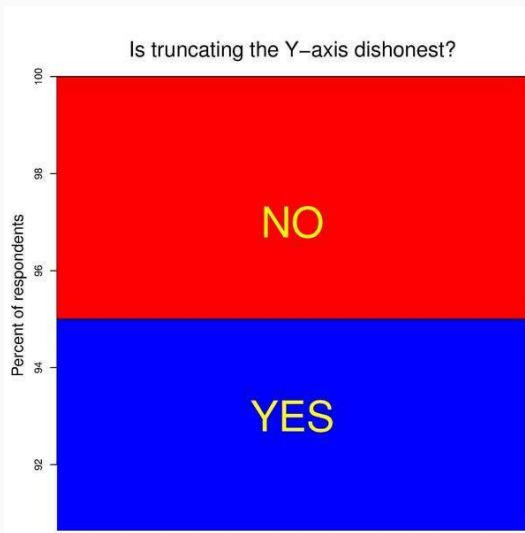
<http://archeologie-copier-coller.com/>

# La manipulation de données

---

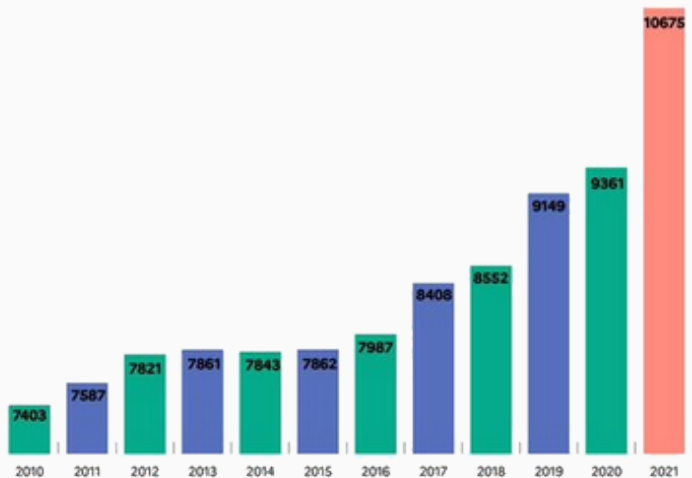


# Axe tronqué



<https://twitter.com/rshotton/status/865951825463312385/photo/1>

## Exercice : refaites le schéma en ajoutant un axe des ordonnées



# Exercice : solution

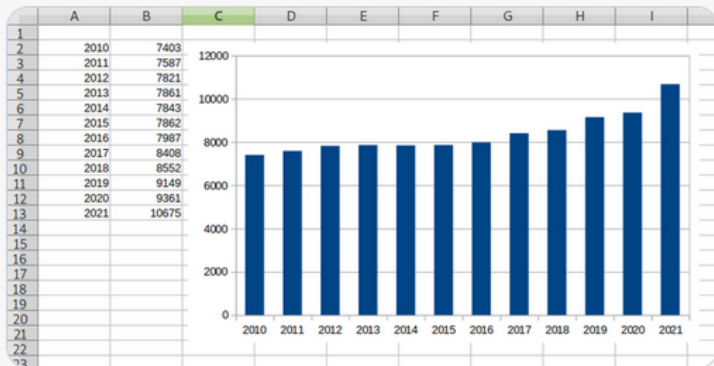


**Michel Billaud** 🍄 @MichelBillaud · 1h

...

Replying to @VidalFrederique @sup\_recherche and 2 others

Un vrai histogramme, qui part de 0 en bas



<https://twitter.com/MichelBillaud/status/1391409131798646785>

## Exercice : source (merci)



Frédérique Vidal

@VidalFrederique

...

L'effort sur la formation des futurs médecins, dans un contexte sanitaire difficile, n'a jamais été aussi important. Avec [#MaSanté2022](#), la barre des 10 500 étudiants admis en 2e année de médecine sera franchie à la rentrée 2021, soit +14% par rapport à l'an dernier.

[Translate Tweet](#)



<https://twitter.com/VidalFrederique/status/1390735956014256131>

## Grave ou pas grave ?

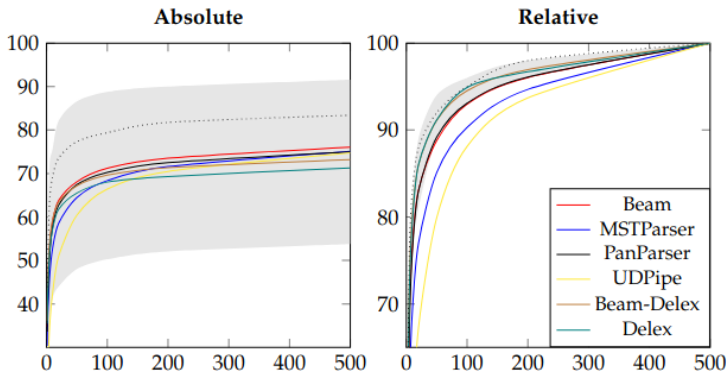
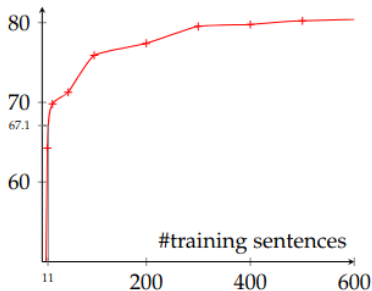


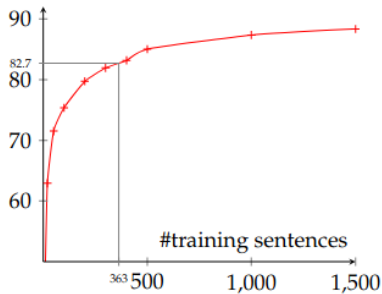
FIGURE 5.2: Learning curves (absolute and relative UAS, averaged across languages; see text for details) for increasing training sizes. The gray areas represent the span between the lowerbound (Ancient Greek) and upperbound (Japanese) curves for PANPARSER. The dotted curves represent French (PANPARSER).

<https://tel.archives-ouvertes.fr/tel-01838834/document>

# Grave ou pas grave ?



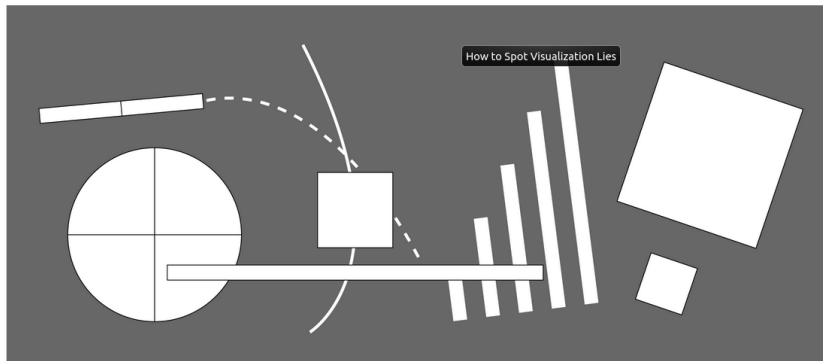
(a) Parser UAS.



(b) Tagger accuracy.

FIGURE 5.1: Monolingual learning curves, compared to cross-lingual best scores.

<https://tel.archives-ouvertes.fr/tel-01838834/document>



## How to Spot Visualization Lies

Keep your eyes open.

By Nathan Yau

<https://flowingdata.com/2017/02/09/how-to-spot-visualization-lies/>

# La (non-)reproductibilité de la recherche

---



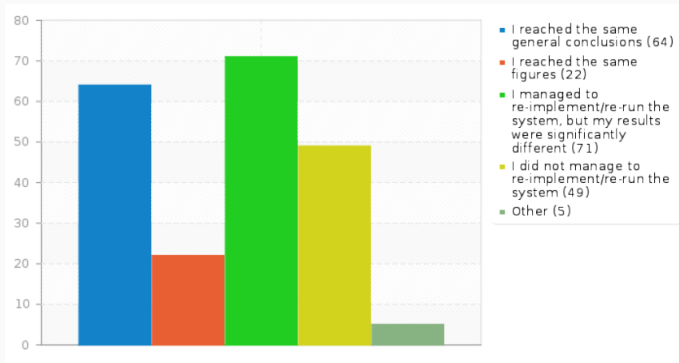
## Répliquabilité

La répliquabilité est une propriété d'une **expérience** : la capacité de répéter – ou non – l'expérience décrite dans une étude.

## Reproductibilité

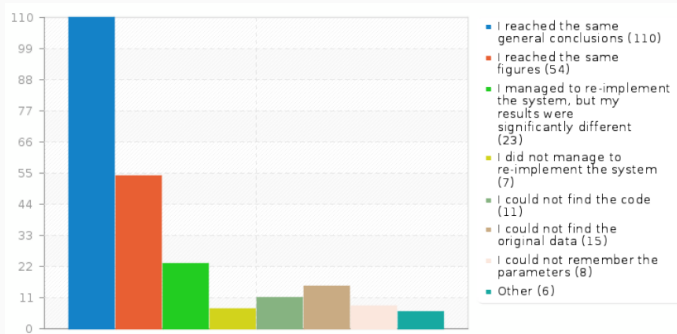
La reproductibilité est une propriété des **résultats** d'une expérience : parvenir – ou non – aux mêmes conclusions, trouvailles ou valeurs.

Lorsque vous avez essayé de reproduire la recherche de quelqu'un d'autre, quels résultats avez-vous obtenus ?



Verticalement : nombre de répondants (plusieurs réponses possibles)

Lorsque vous avez essayé de reproduire vos **propres** recherches, quels résultats avez-vous obtenus ?

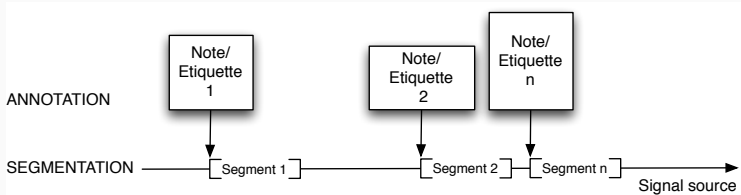


Verticalement : nombre de répondants (plusieurs réponses possibles)

# Et vous ?

Dans quelle mesure votre recherche est **reproductible** ?

# Répliquabilité et biais : l'annotation manuelle de corpus



Ajout d'informations **interprétatives** [Leech, 1997, Habert, 2005]

Annoter la phrase suivante avec les étiquettes :  
NOM, ADJ, ADV, VERBE, PREP, PRONOM

Je suis fatiguée.

L'annotation est de l'ordre de la **quantification** [Desrosières, 2008] :

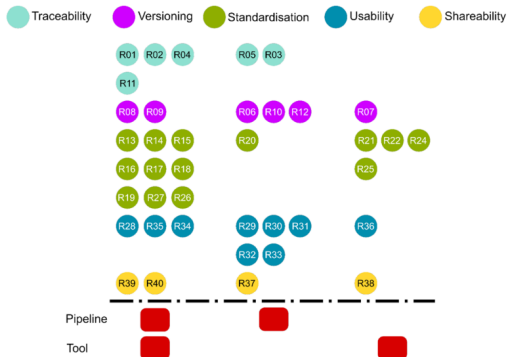
- **mesurer** : implique une forme mesurable (par ex. la hauteur du Mont Blanc)
- **quantifier** : suppose des conventions d'équivalences préalables

Outiller le consensus :

- guide d'annotation (12 p. pour le football)
- **évaluer** le consensus (la cohérence)  $\Rightarrow$  calculer l'accord inter-annotateurs

# Comment rendre une recherche répliquable ?

**Figure 2.** Classification of recommendation at tool and pipeline level. 21 recommendations are applicable both to tools ...





Nécessité de mettre les données à disposition pour assurer la  
reproductibilité de la recherche

VS

Respecter la vie privée des participants (leurs données)

[http://www.ethique-et-tal.org/2020/11/26/  
replicabilite-vs-protection-des-donnees-limpossible-equilibre/](http://www.ethique-et-tal.org/2020/11/26/replicabilite-vs-protection-des-donnees-limpossible-equilibre/)

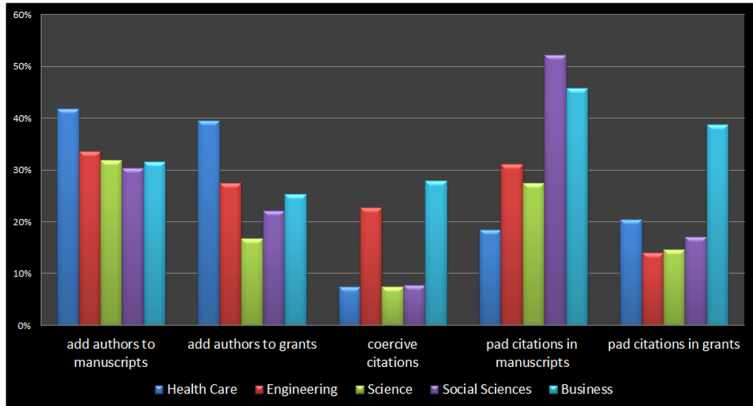
- Éviter les revues « parasites » (difficile quand on est doctorant)
- Mettre ses publications sur HAL (formation possible)
- Attention aux revues/éditeurs prédateurs (pas de bonne qualité scientifique)

## Signature d'articles

---

# Qui a déjà ajouté un auteur de complaisance à un article ?

Étude portant sur plus de 12 000 chercheurs de plusieurs champs disciplinaires [Fong and Wilhite, 2017] :



Manipulation of authorship and citation across academia.

# Qu'est-ce qu'un auteur ?

Selon l'ICMJE 4 (International Committee of Medical Journal Editors), un auteur doit :

- apporter une « contribution substantielle à la conception ou aux méthodes de la recherche » ET
- participer à la « rédaction préliminaire de l'article ou sa révision critique » ET
- donner son « approbation finale de la version à publier » ET
- prendre « l'engagement à assumer l'imputabilité pour tous les aspects de la recherche »

<http://icmje.org/recommendations/translations/french2017.pdf>

# Comment éviter les conflits d'auteurs ? (recommandations)

- se mettre d'accord dès le **début** (de la thèse, de la recherche, du projet) sur un fonctionnement
- utiliser une plateforme d'écriture collaborative qui permette de **tracer** les auteurs et leur participation effective
- garder des **traces écrites** des échanges sur ces sujets si la situation est tendue

**Un contexte de recherche  
favorisant les méconduites ?**

---

ScienceAdvances

Contents ▾ News ▾ Careers ▾ Journals ▾

[Read our COVID-19 research and news.](#)

**SHARE** RESEARCH ARTICLE | SCIENTIFIC COMMUNITY

## Nonreplicable publications are cited more than replicable ones

 Marta Serra-Garcia<sup>\*,†</sup> and  Uri Gneezy<sup>†</sup>

[+ See all authors and affiliations](#)

Science Advances 21 May 2021:  
Vol. 7, no. 21, eabd1705  
DOI: 10.1126/sciadv.abd1705

<https://advances.sciencemag.org/content/7/21/eabd1705>




Mais les articles avec données associées sont plus cités...

# PLOS ONE

 OPEN ACCESS  PEER-REVIEWED

RESEARCH ARTICLE

## Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar  Roger S. Day, Douglas B. Fridsma

Published: March 21, 2007 • <https://doi.org/10.1371/journal.pone.0000308>

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308>

Les recherches négatives finissent dans le tiroir :

*« Modern science's professional culture prizes positive results far above negative results, and also far above attempts to reproduce earlier research. Scientists therefore steer away from replication studies, and their negative results go into the file drawer. » [Randall et al., 2018]*

<https://www.nas.org/reports/the-irreproducibility-crisis-of-modern-science/full-report>



## Une lauréate d'un Nobel rétracte un de ses articles

[\[Accueil\]](#) / [\[Société\]](#) / [\[Science\]](#)



Photo: Damian Dovarganes Associated Press La lauréate du prix Nobel de chimie en 2018, l'Américaine Frances Arnold, professeur au California Institute of Technology

<https://www.ledevoir.com/societe/science/570410/>

[la-laureate-du-prix-nobel-de-chimie-en-2018-a-du-retracter-un-article](https://www.ledevoir.com/societe/science/570410/)

## Discussion

---

# Questions d'intégrité scientifique dans votre domaine ?



**Merci de votre participation !**

- Reproductibilité en linguistique (EN)
- Pratiques bibliographiques (un peu long)
- Etude qualitative
- L'art de la citation (hyperliens)
- Plagiat en philosophie
- Plagiat
- Contexte des méconduites scientifiques
- Réplicabilité et citation (EN)
- COMETS
- fake data

Rétractations : post-publication, scientifique triste (France Culture)



# Bibliographie

---



Cohen, K. B., Xia, J., Zweigenbaum, P., Callahan, T., Hargraves, O., Goss, F., Ide, N., Névéol, A., Grouin, C., and Hunter, L. E. (2018).

**Three Dimensions of Reproducibility in Natural Language Processing.**

In Proceedings of LREC.



Desrosières, A. (2008).

**Pour une sociologie historique de la quantification : L'Argument statistique.**

Presses de l'école des Mines de Paris.



Fong, E. A. and Wilhite, A. W. (2017).

**Authorship and citation manipulation in academic research.**

PLOS ONE, 12(12) :1–34.



Habert, B. (2005).

**Portrait de linguiste(s) à l'instrument.**

Texto !, vol. X(4).



Henriet, P. and Ouzoulis, P. (2021).

**Rapport au nom de l'office parlementaire d'évaluation  
des choix scientifiques et technologiques sur promouvoir  
et protéger une culture partagée de l'intégrité  
scientifique.**

Technical report.



Leech, G. (1997).

**Corpus annotation : Linguistic information from  
computer text corpora, chapter Introducing corpus  
annotation, pages 1–18.**

Longman, Londres, Angleterre.



Mieskes, M., Fort, K., Névéol, A., Grouin, C., and Cohen, K. B. (2019).

### **NLP Community Perspectives on Replicability.**

In Recent Advances in Natural Language Processing, Varna, Bulgaria.



Serres, A. (2021).

### **Introduction à l'intégrité scientifique.**

webinaire.