

# Poésie et vérité d'un journalisme de données

**La chronique boursière  
au prisme du data mining (1815-1870)**



# #media19

**Pierre-Carl Langlais**

**GRIPIC/CELSA**

**Thèse** : [theses.fr/s96560](https://theses.fr/s96560)

**Twitter** : @Dorialexander

**Wikipédia** : User:Alexander Doria

**Github** : [github.com/Dorialexander](https://github.com/Dorialexander)

**Hypotheses** : <http://scoms.hypotheses.org/>

**Rue89** : [blogs.rue89.nouvelobs.com/les-coulisses-de-wikipedia](https://blogs.rue89.nouvelobs.com/les-coulisses-de-wikipedia)

# Le Projet Jourdain



# JOURDAIN

Journalisme & Data & Information

« ou comment on fait du datajournalisme sans le savoir. »



- le projet
- l'équipe
- analyses
- journée d'études
- séminaire
- travaux de recherche
- références

- JEUX DE DONNÉES
- DATAJOURNALISTES
- ORGANISATIONS
- PROJETS
- LÉGISLATION
- OUTILS ET LOGICIELS
- ÉVÉNEMENTS
- GLOSSAIRE

**DATA JOURNALISM HANDBOOK IN NUMBERS**

60 pages    2 days    6 chapters    20,000 words

55 contributeurs

**Cartographie d'un manuel communautaire**

*How journalists use our data to improve the news*

**The Data Journalism Handbook**

by Liliانا Bounegru and Lucy Chambers



# Une application : Pyllica

A tool to retrieve the text-version of newspapers in gallica (it works also for books and other single documents)

27 commits

1 branch

0 releases

1 contributor



branch: master -

Pyllica / +



Update pyllica.py ...



Dorialexander authored on 23 Apr 2014

latest commit 2748772f8d



.gitignore

Initial commit

a year ago



LICENSE

Initial commit

a year ago



README.md

Update README.md

a year ago



pyllica.py

Update pyllica.py

a year ago

README.md

## Pyllica

Pyllica is a small python tool to retrieve the text-version of newspapers in [gallica](#), the main french digital library. It works also for books and other single documents (and, soon perhaps, for periodicals).

It will not retrieve documents that are only accessible in image version (that is, that have not been [OCRized](#)).

**1.**

# Imprimé/numérique

*Les métamorphoses du document*

# Un premier grand projet d'indexation...

## e - INTERPRETES

### a) Acteurs

Alphonsine : 11. 9. F	Judith : 10. 5. 3C
Arnouly-Plessy : 23. 1. F	Lafont : 30. 1. F; 6. 11. F
Berton : 13. 2. F	Lafontaine V. : 16. 10. F
Bressant : 23. 1. F	Lagier S. : 7. 1. 3F
Brindeau : 13. 2. F	Lemaître F. : 27. 9. 2E; 2. 10. F
Carvalho : 25. 2. 2E <u>DIV</u>	Matthews C. : 11. 9. F
Champfleury : 15. 12. 2D	Mélingue : 13. 3. F
Chaumont : 20. 11. F	Michel : 9. 10. F
Déjazet : 30. 10. F	Miroy C. : 12. 6. F
Delabranche : 13. 12. F	Provost : 16. 10. F; 27. 12. 3A; 28. 12. 2B
Delaporte : 30. 1. F	Ristouri : 29. 5. F; 26. 6. 4F
Doche : 6. 2. F	Schneider H. : 11. 4. 1D
Duret : 27. 11. F	Talbot : 23. 1. F
Essler J. : 13. 11. F	Thuillier : 20. 11. F
Fargeuil : 13. 11. F	Vernet : 31. 7. F
Favart : 25. 12. F	Vertpré J. : 9. 11. 3A
Fébvre : 13. 11. F	Worms : 13. 2. F
Félix L. : 13. 2. F	
Got : 26. 7. 2F	

### b) Chanteurs

Battu : 28. 11. F	Maudit : 28. 11. F
Bloch : 15. 11. F	Monrose : 1. 4. 3D
Gontié C. : 5. 9. F	Roze : 5. 9. F
Gourdin : 2. 8. 3C	Theresa : 23. 2. 3D; 2. 4. F; 15. 5. 2D; 1. 3. 3B
Leroy : 5. 9. F	
Lichtmay : 25. 7. F	

**Les Tables du Temps** : un index du journal *Le Temps* entreprise au début des années 1960



“

*n'ont pas été inventoriées (...) les articles à  
périodicité régulière quotidienne ou  
hebdomadaire dont le contenu ne se prêtait  
pas à un inventaire systématique (tableaux  
quotidiens de la bourse ; Bulletin financier  
hebdomadaire...*

**Pierre Albert, *Les Tables du Temps* (2013)**



# FILTRE DE LECTURE

L'architecture documentaire est modelée par des conceptions culturelles du texte

ABONNEMENTS (PARIS) :

Trois mois : 13 fr. — Six mois : 23 fr. — Un an : 52 fr.

ANGLETERRE : M. DODD, 8, DILE-NEWPORT street, LONDON SQUARE, LONDON, W. G.

POUR la rédaction, s'adresser à M. A. NEFFZER, rédacteur en chef. (Affranchir.)

La rédaction ne répond pas des articles communiqués et ne se charge pas de les renvoyer.



# Le C

LUNDI 29 AVRIL.

EDITION DU SOIR.

### BOURSE DE PARIS

ET BULLETIN FINANCIER.

CLOTURE.	le 27	le 29	HAUSSE	BAISSE
<b>3 0/0</b>				
Comptant . . .	68 70	68 60	» »	» 10
Fin courant . .	68 70	68 65	» »	» 05
<b>4 1/2 0/0</b>				
Comptant . . .	95 50	95 25	» »	» 25
Fin courant . .	95 40	95 40	» »	» »

Les dispositions sont aujourd'hui moins bonnes. Dès le début, on s'aperçoit que la chaleur avec laquelle on poursuivait la lutte est tombée. Le 3 0/0 ouvre faible à 68 65; et, peu d'instants après, il se manifeste des offres assez importantes, et la Bourse tombe à 68 60.

Ces ventes, comme la plupart des autres qui seront faits ces jours derniers, n'ont d'autre cause, d'ailleurs, que des besoins de liquidation. C'est depuis le jour important, et chacun cherche à arranger sa position en vue de la réponse des primes.

On a cependant voulu expliquer par autre chose que des besoins de liquidation la faiblesse des cours au début de la Bourse; on a parlé de mauvaises nouvelles de Naples; on a surtout beaucoup parlé de l'emprunt du Brésil.

Au dire des nouvelles, cet emprunt ne serait plus de 200, mais de 340 millions. Nous ne pensons pas que ces appréciations soient exactes, et qu'il y ait rien de définitivement arrêté ni sur le chiffre de l'emprunt, ni même sur le mode d'émission. Quel qu'il en soit, on cherche maintenant à faire de l'excessif amené de cet emprunt un argument pour arrêter la hausse. La liquidation aigrit, on a réussi pour aujourd'hui. Le 3 0/0 s'est relevé lentement et n'a pas dépassé 68 70, il reste à 68 60 au comptant.

La spéculation, un peu limitée sur la Bourse, s'est portée avec un grand ensemble sur les chemins français. Il s'est fait ce matin des affaires énormes

Mais une part légitime est due et sera faite à toutes les autres marchandises; leurs mouvements, moins fréquents et plus caractéristiques, seront fidèlement notés avec leur signification la plus probable. Autrefois, la sagacité du commerçant s'exerçait dans les limites étroites d'une production de quelques lieues carrés, et d'une consommation purement locale. Aujourd'hui, les marchés d'Odessa ou de Melbourne pèsent avec une électricité promptitude sur ceux de Marseille ou de Strasbourg, d'où la nécessité évidente des plus précises informations.

Nous espérons avoir bientôt des correspondants sur chaque place importante, mais nous comptons aussi sur nos lecteurs, dont les communications seront accueillies avec un sympathique intérêt. Nous verrions avec bonheur s'établir une entremise entre ceux qui nous font part de nous lire et ceux qui nous donnent des renseignements et nous indiquent les prix de revient et les débouchés d'une marchandise naturelle ou artificielle.

En del nous avons porté une importance tripartite intérêts. La cause de la liberté commerciale est gagnée à cette heure; et une défense serait presque aussi tardive qu'un réquisitoire; mais on ne change pas impunément les conditions d'une industrie; et déjà plus d'une violente secousse signale ce déplacement. Le commerce intelligent doit

nous publions plus loin, et qui nous transmettent des impressions complètement différentes sur la réconciliation de M. de Cavour et du général Garibaldi. Notre correspondant de Turin pense que M. de Cavour a reculé devant les manifestations qui se préparaient en faveur de Garibaldi; notre correspondant de Gênes, au contraire, conteste la portée de ces manifestations, et soutient que la réconciliation a été toute spontanée.

Les dépêches de Turin annoncent la prochaine présentation d'un projet de loi concernant l'émission d'un emprunt de 500 millions de francs.

Le Journal allemand de Francfort parle de la prochaine arrivée d'une flotte anglaise dans l'Adriatique. C'est un bruit dont

**L'indexation systématique**  
 Sur Gallica, la requête "bourse"  
 porte sur tout le journal

Après les tentatives attentives des navies anglais croiseraient entre Antivari et admettent les lésions de la surveillance lais stationnaires devant Trieste, Venise, et les autres ports de la mer Adriatique. Les journaux à l'agitation qui commencent à paraître dans les journaux de Pétersbourg une note sur les Pologne, qui peut être considérée

Mouvement, qu'aux appréciations assez unanimes de la presse occidentale sur ces événements. Il est naturel que le Journal de Saint Pétersbourg cherche à justifier le gouvernement russe. Il donne d'ailleurs l'assurance que le gouvernement veillera à ce que les institutions accordées soient cons-

# Vers une saisie exhaustive de la presse ancienne..



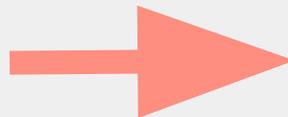
La présentation du projet Europeana Newspaper célèbre le quotidien dans son intégralité

## ...ou un déplacement des filtres de lectures ?



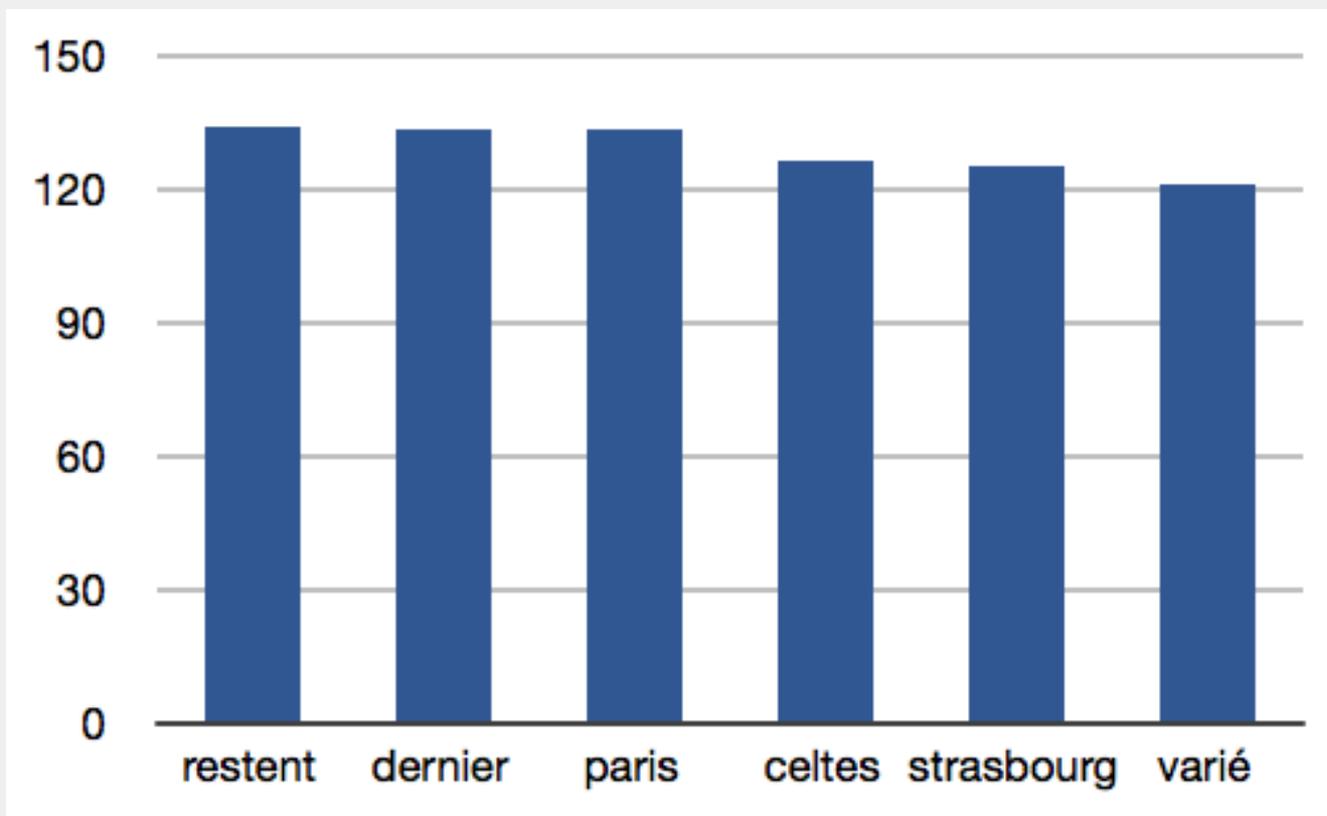
Les lettres sortent littéralement du journal pour s'intégrer dans le document numérique

## Une normalisation typographique...



Les algorithmes de reconnaissance de caractère (OCR) rapprochent chaque caractère d'un modèle idéal

## ...qui génère des erreurs structurelles



Le mot “celte” est l’un des plus fréquents dans la chronique boursière des *Débats* en 1852

# La sélection documentaire devient illisible



Le développement de grandes plate-formes écrase l' historicité des documents

## La bourse de Paris: mœurs, anecdotes, spéculations et ... - P...



<https://books.google.fr/books?id=qzINAAAAYAAJ>

Antoine Guitton - 1857 - Lire

mœurs, anecdotes, spéculations et conseils pour y faire valoir ses capitaux  
Antoine Guitton. baisses ou suivre les mouvements de hausse très-prononcés,  
comme il s'en présente presque toujours au retour des beaux jours, avec  
cette ...

## La Bourse de Paris: mœurs, anecdotes, spéculations et ... - ...



[https://books.google.fr/books?id=KDAjc2\\_Qfd4C](https://books.google.fr/books?id=KDAjc2_Qfd4C)

Q. G. de Mériclet - 1854 - Lire

mœurs, anecdotes, spéculations et conseils par... Q. G. de Mériclet. t. reur  
Napoléon, qui lui demandait le meilleur moyen de C... er le bonheur du  
peuple : Sire, doublez la gendarmerie. Je ne conteste pas que la sécurité du  
pays n'ait ...

Un soupçon de  
prédilection

Des duplicats  
sur Google Books

## La Bourse de Paris: mœurs, usages, spéculations et ... - Pag...



<https://books.google.fr/books?id=4i3vtADaHkoC>

Guitton de Mériclet - 1856 - Lire - Autres éditions

mœurs, usages, spéculations et conseils pour faire valoir des capitaux Guitton  
de Mériclet. XIII REVUE DES VALEURS DE LA BOURSE DE PARIS. <=&®> Le  
capitaliste et le spéculateur ont un véritable intérêt à acquérir les notions les  
plus ...

## La Bourse de Paris: Moeurs. Usages. Spéculations et ... - Pa...



<https://books.google.fr/books?id=0eVTAAAcAAJ>

A.G. de Mériclet - 1856 - Lire - Autres éditions

râurw' XIII REVUE DES VALEURS DE LA BOURSE DE PARIS. Le capitaliste  
et le spéculateur ont un véritable intérêt à acquérir les notions .les plus  
précises sur les valeurs de la Bourse. Nous avouons franchement que nous  
ne voulons ...



# BIG DATA

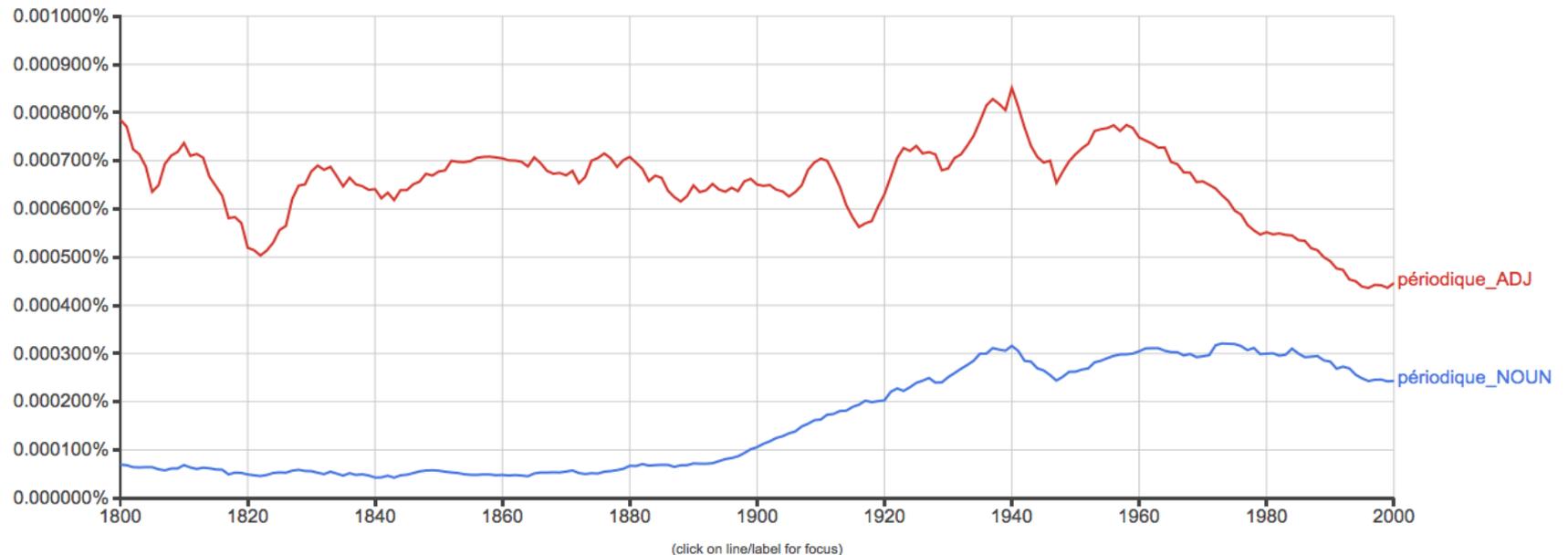
La promesse de transcender les biais par le nombre...

# Ngram Viewer

## Google books Ngram Viewer

Graph these comma-separated phrases:   case-insensitive

between  and  from the corpus  with smoothing of  [Search lots of books](#)



Mesurer les occurrences de mots et d'expressions dans un dixième des livres publiés

# Les biais de l'OCR



Quand le s long de l'ancien français devient un *f*.



“

*de nombreux périodiques ont des dates de publications attribuées par erreur à l'année du premier numéro. Ces journaux et publications sérielles représentaient également différents aspects culturels que les livres. Pour toutes ces raisons, nous avons décidé de filtrer les publications sérielles dans la mesure du possible*

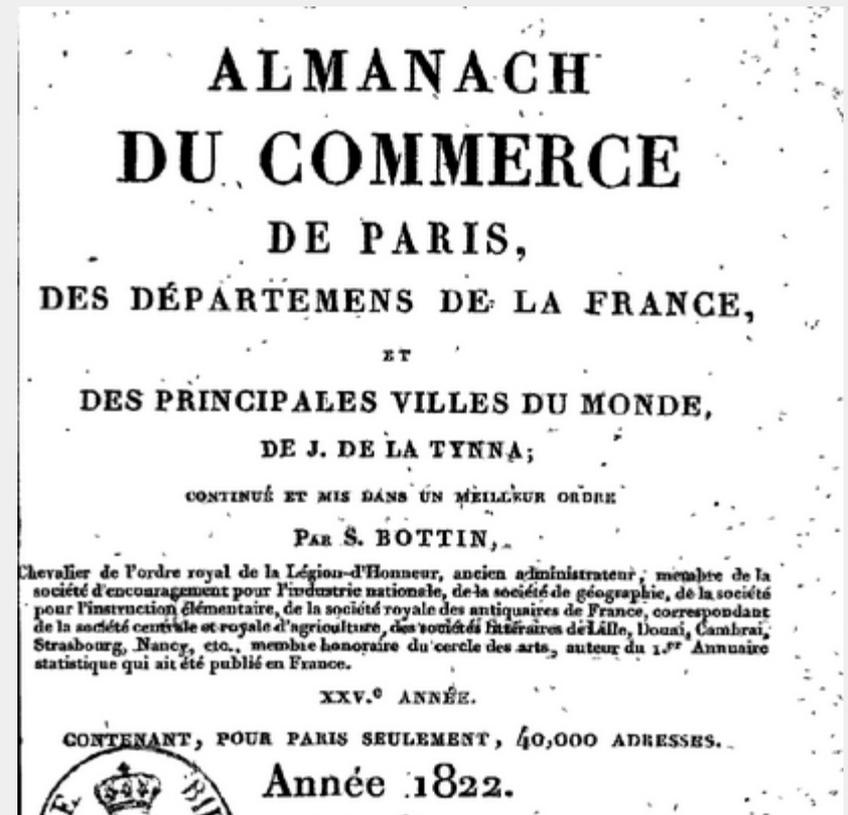
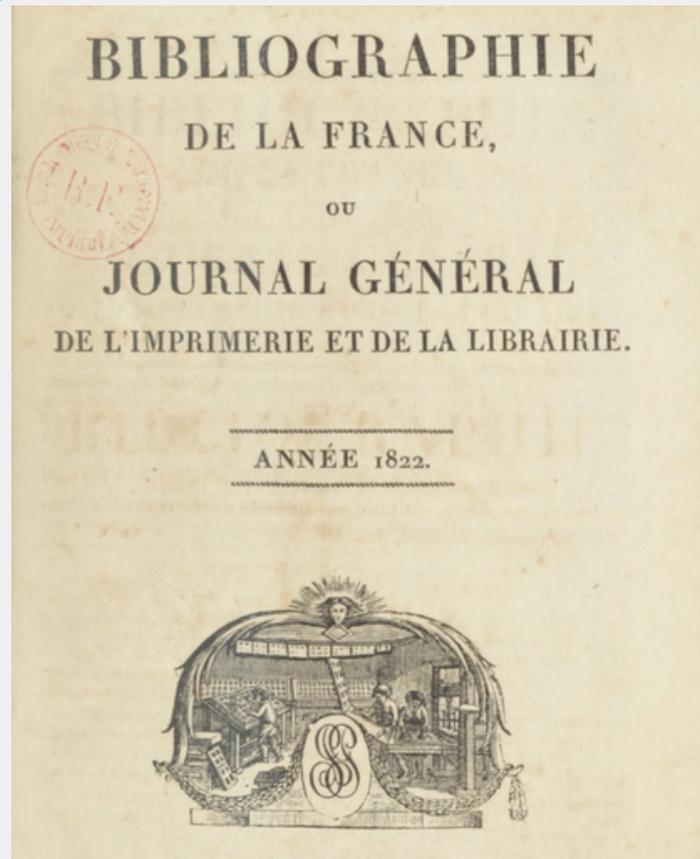
**Michel et al., « Supplement Online Material for Quantitative Analysis of Culture », *Science*, 2011**

# Corriger les biais

```
def thegreatregexsynt(file, newfile):
    textfile = codecs.open(file, 'r', "utf-8")
    fileread = textfile.read()
    correctlist = getstockname(fileread, "Saint-Germain")
    for item in correctlist:
        fileread = re.sub(item, "Saint-Germain", fileread, re.UNICODE)
    ofileread = re.sub(u' ello ', u' elle ', fileread, re.UNICODE)
    ofileread = re.sub(u'l tl', u'M', fileread, re.UNICODE)
    ofileread = re.sub(u' quo(,?) ', u' que\2 ', ofileread, re.UNICODE)
    ofileread = re.sub(u'\'', u'\'', ofileread, re.UNICODE)
    ofileread = re.sub(u' do ', u' de ', ofileread, re.UNICODE)
    ofileread = re.sub(u'fi', u'fi', ofileread, re.UNICODE)
    ofileread = re.sub(u'fl', u'fl', ofileread, re.UNICODE)
    ofileread = re.sub(u' [Itli]r.', u' fr.', ofileread, re.UNICODE)
    ofileread = re.sub(ur'\b(\d) 0.0\b', ur'\1%', ofileread, re.UNICODE)
    ofileread = re.sub(ur'V[ec]rsaill[ec]s', ur'Versailles', ofileread, re.UNICODE)
    ofileread = re.sub(ur'(\b ?)Pannée( ?\b)', ur"\1l'année\2", ofileread, re.UNICODE)
    ofileread = re.sub(ur'(\b ?)Pavait( ?\b)', ur"\1l'avait\2", ofileread, re.UNICODE)
    ofileread = re.sub(ur' ellet ', ur' effet ', ofileread, re.UNICODE)
    ofileread = re.sub(ur' le marche\b', ur" le marché ", ofileread, re.UNICODE)
    ofileread = re.sub(ur' a[é]té ', ur" a été ", ofileread, re.UNICODE)
    ofileread = re.sub(ur' pou ', ur" peu ", ofileread, re.UNICODE)
    ofileread = re.sub(ur'chemins de(\n| ) [t]er', ur"chemins de\1fer", ofileread, re.UNICODE)
    ofileread = re.sub(ur' ta ', ur" la ", ofileread, re.UNICODE)
    ofileread = re.sub(ur'(\d) n ', ur"\1 » ", ofileread, re.UNICODE)
    ofileread = re.sub(ur' défalcation ', ur" défalcation ", ofileread, re.UNICODE)
    ofileread = re.sub(ur'i(\d+)( |\n)', ur"i\1\2", ofileread, re.UNICODE)
    ofileread = re.sub(ur'(\d+ [1-4]).(4)', ur"\1/\2", ofileread, re.UNICODE)
    ofileread = re.sub(ur'(\d+ [1357]).(8)', ur"\1/\2", ofileread, re.UNICODE)
    ofileread = re.sub(ur' (l)(ond)(s?) ', ur" f\2\3 ", ofileread, re.UNICODE)
```

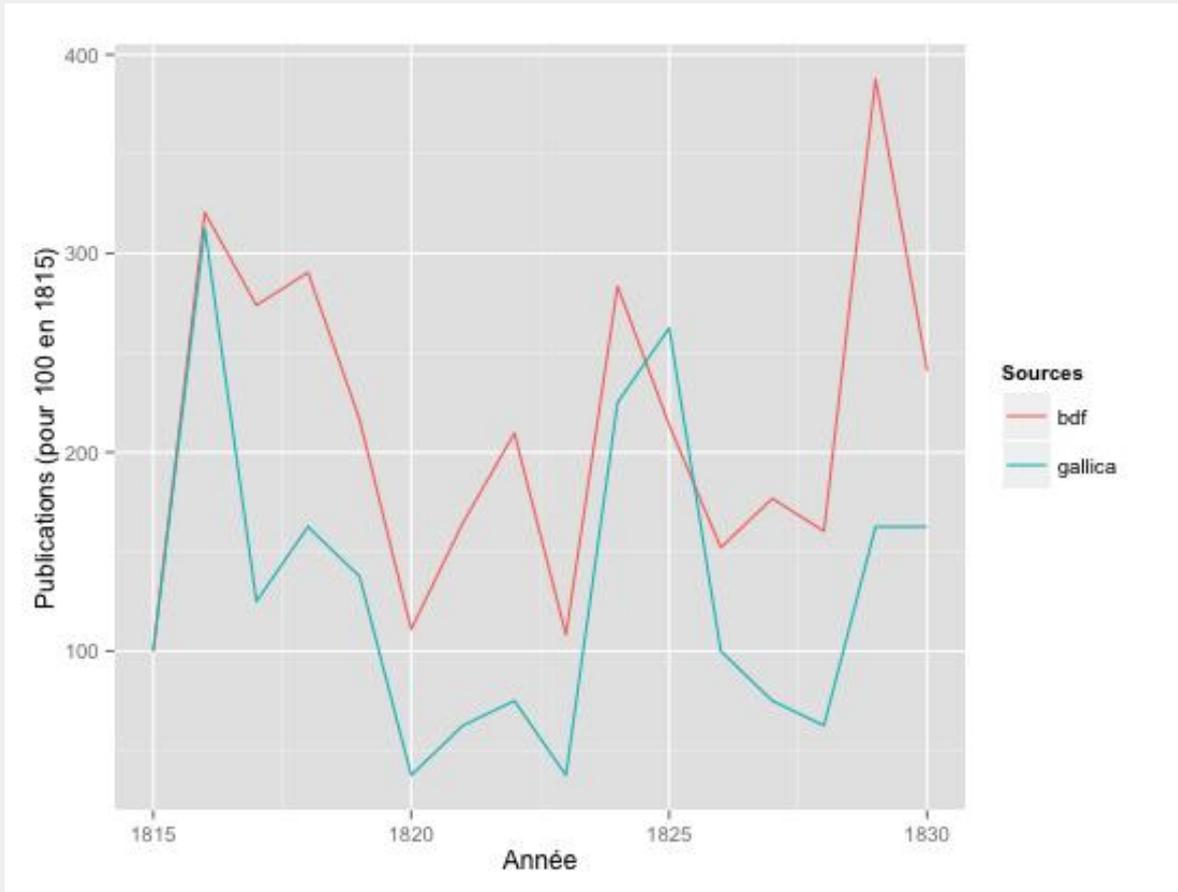
Pour amortir les erreurs de l'OCR, le chercheur doit se faire éditeur...

# Corriger les biais



Les compilations bibliographiques anciennes permettent d'évaluer l'ampleur des déformations documentaires

# Corriger les biais



Les ouvrages d'économie en France (1815-1830)  
Bibliographie de France vs. Gallica

**2.**

# Une convergence médiatique

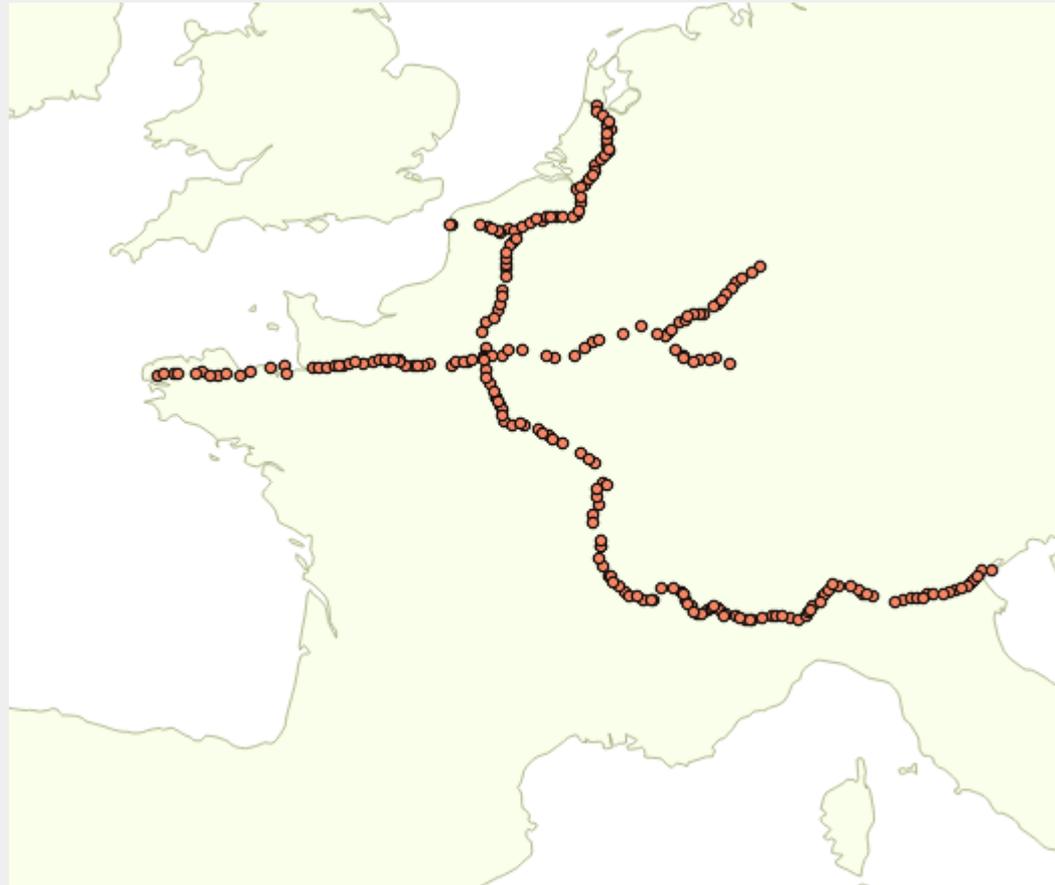
*La chronique boursière au  
point de tension*

“

Avec des télégraphes convenablement disposés, Paris pourrait être instruite (...) de tous les mouvements des ports, et de toutes les opérations des principales bourses (...) et d'après les avis que les négociants recevraient, ils pourraient faire passer dans le même jour, des ordres de vente et d'achats pour l'Allemagne, la Hollande, Hambourg, la Suisse, l'Italie, le Levant, l'Espagne, etc.

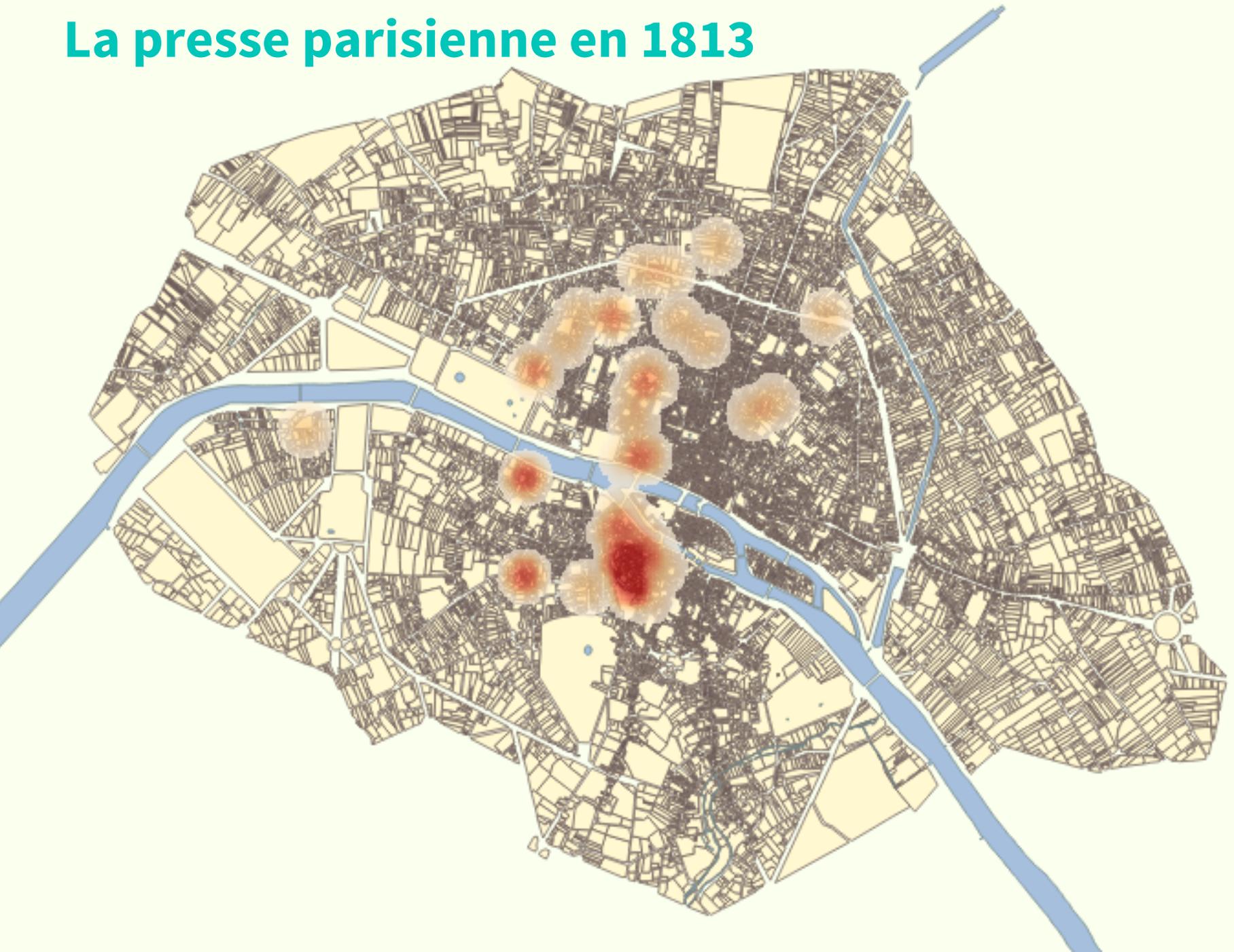
**Claude Chappe**, *Journal des Débats*, 3 mai 1802, p. 5

## Le rêve de Chappe...

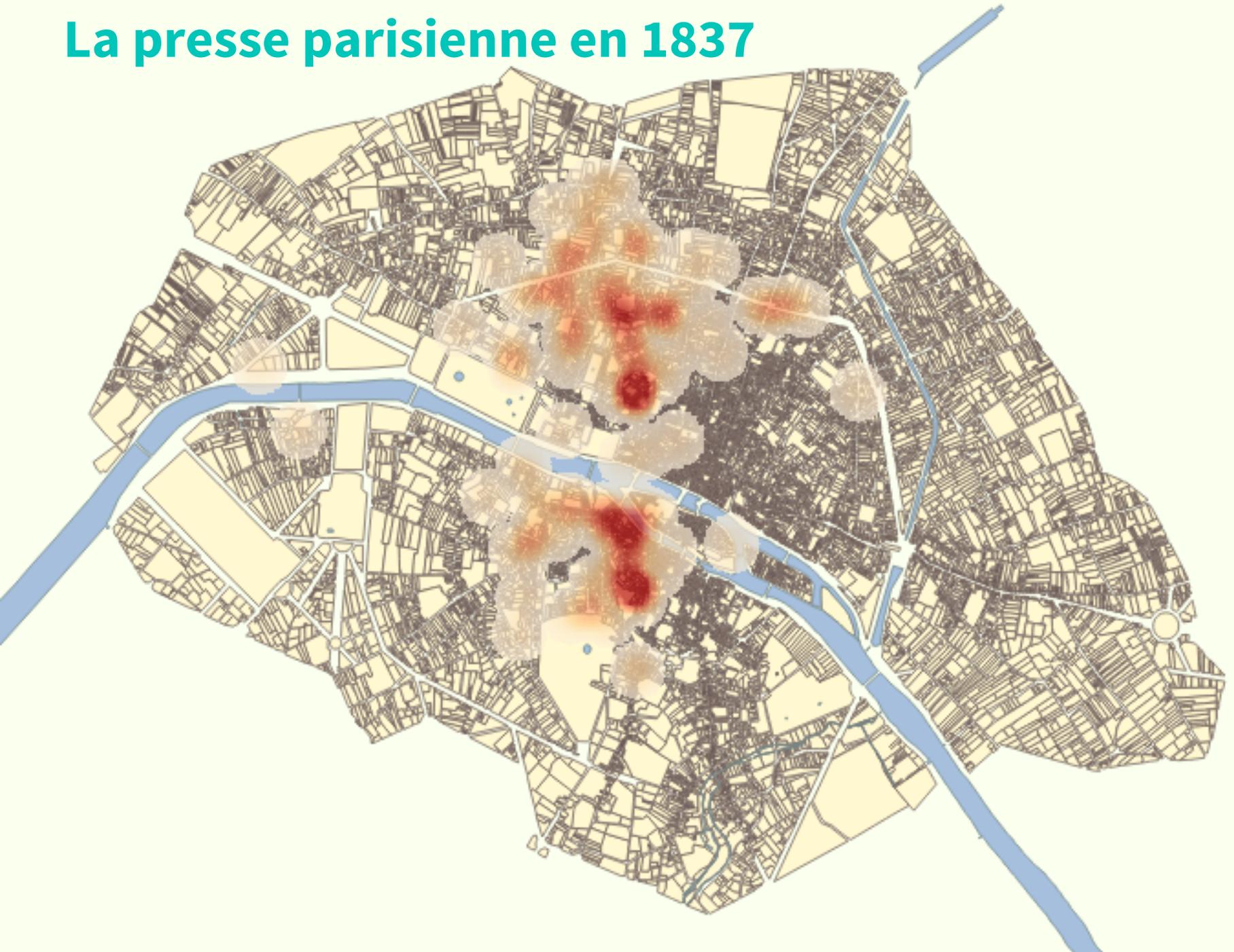


À la fin de l'Empire, le réseau télégraphique prend temporairement une dimension européenne...

# La presse parisienne en 1813



# La presse parisienne en 1837





*Nouvelles Annales des Voyages et des sciences géographiques*

*Revue et Gazette musicale de Paris*

*Journal de la Société d'encouragement*

*A. Boïot*

*Journal des Connaissances médicales pratiques*

*Le Polonais*

*La Loi*

*L'Echo de la jeune France*

*Abeille musicale*

*Revue de Paris*

*The Paris Literary gazette*

*La Sentinelle*

*L'Artiste*

*Mouvement des Ports*

*Le Monde dramatique*

*Le Constitutionnel*

*Grandes affiches de France*

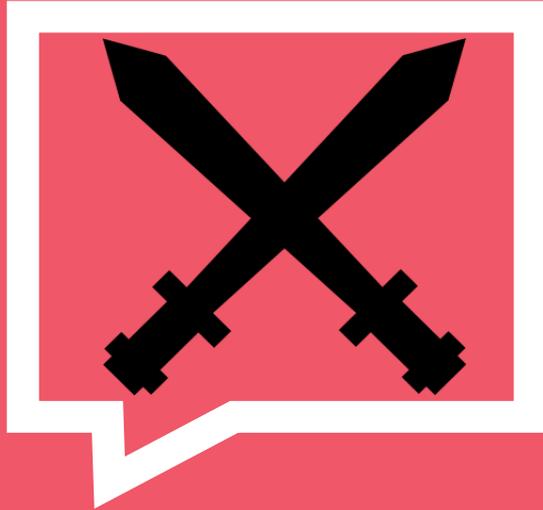
*La France*

*Prix courant général et légal des*

*Galignani's Messenger*

*The London and Paris Observer*

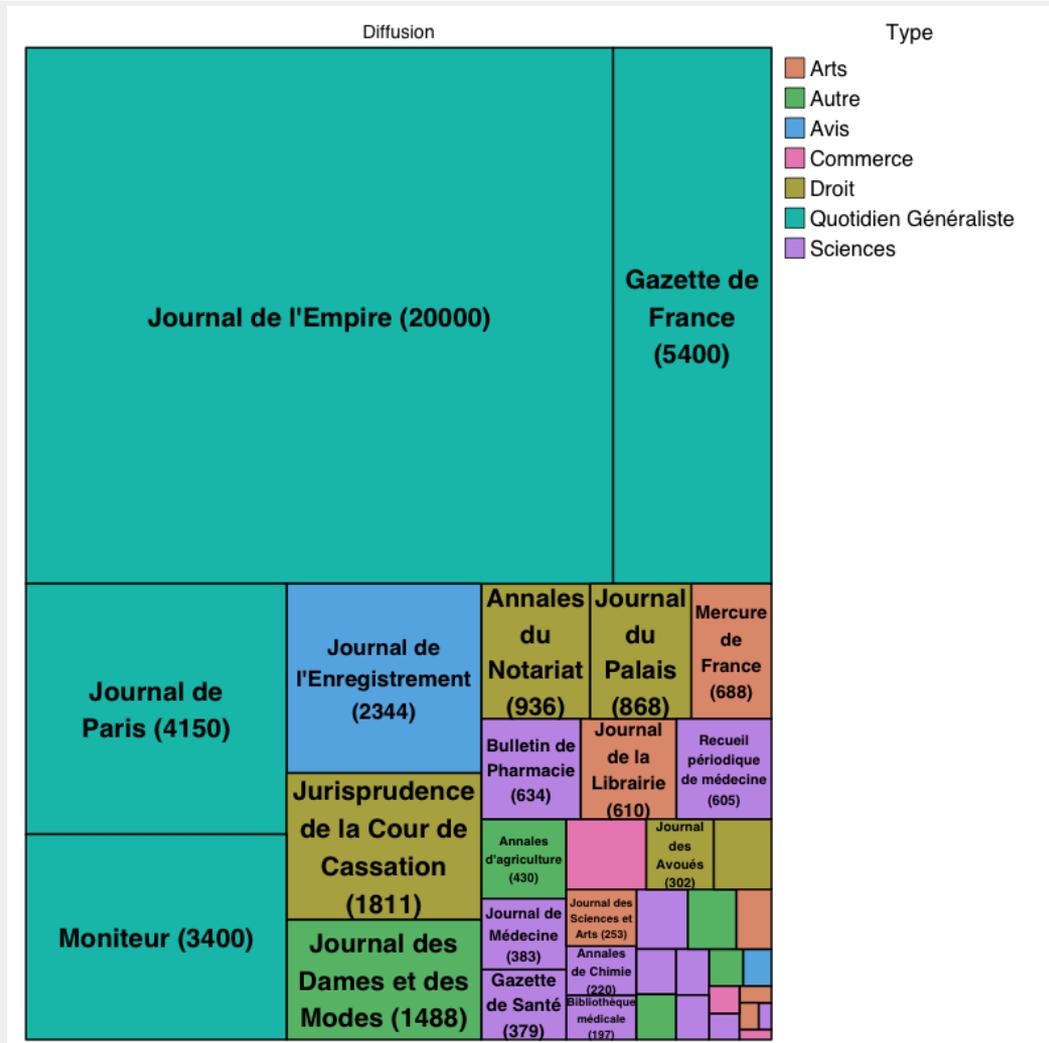
OFFICE DE CORRESPONDANCE



# GUERRE TEXTUELLE

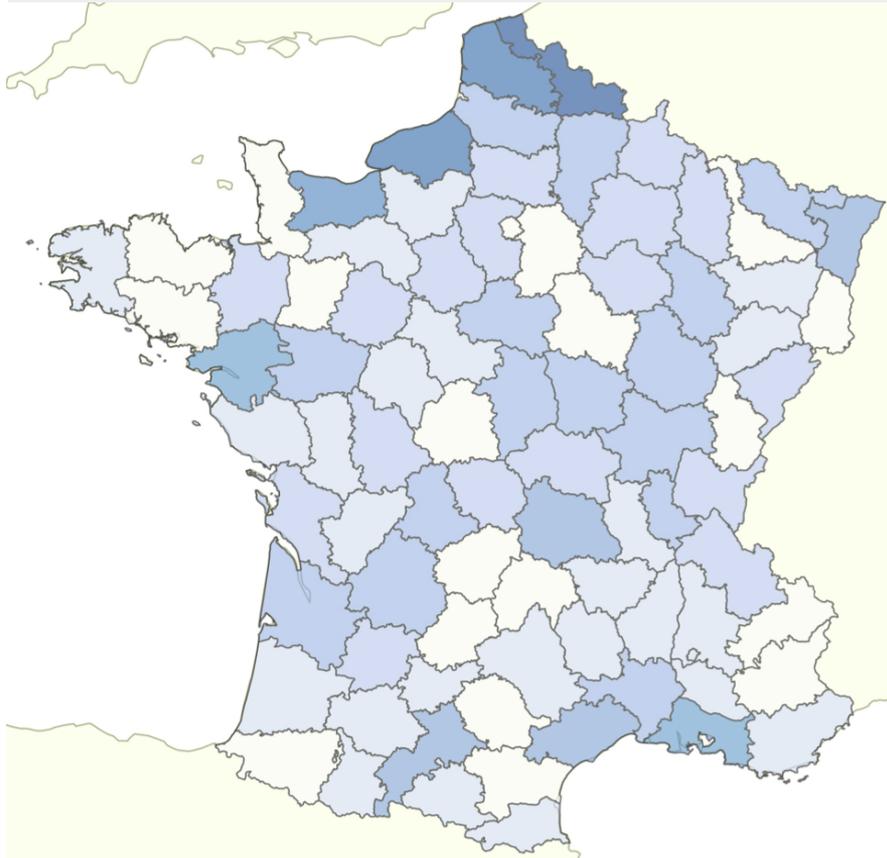
Plusieurs acteurs tentent de contrôler la circulation de l'information boursière

# Les grands quotidiens



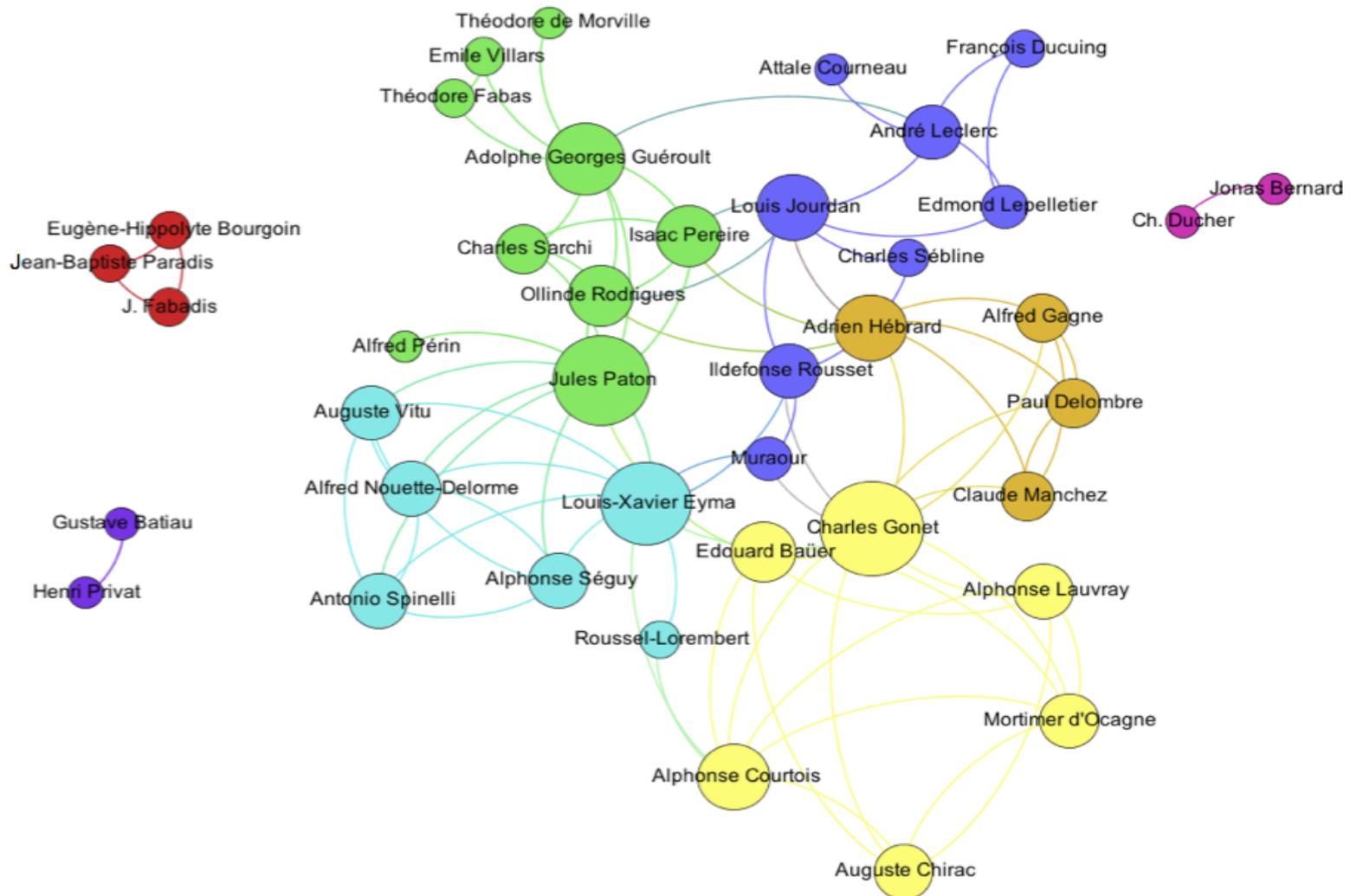
Sous l'effet de la censure économique, la presse française est très concentrée

# Les correspondances de journaux



En 1830, l'Office-Correspondance fédère 100 titres régionaux et 100 partenaires internationaux

# Les Bulletiniers

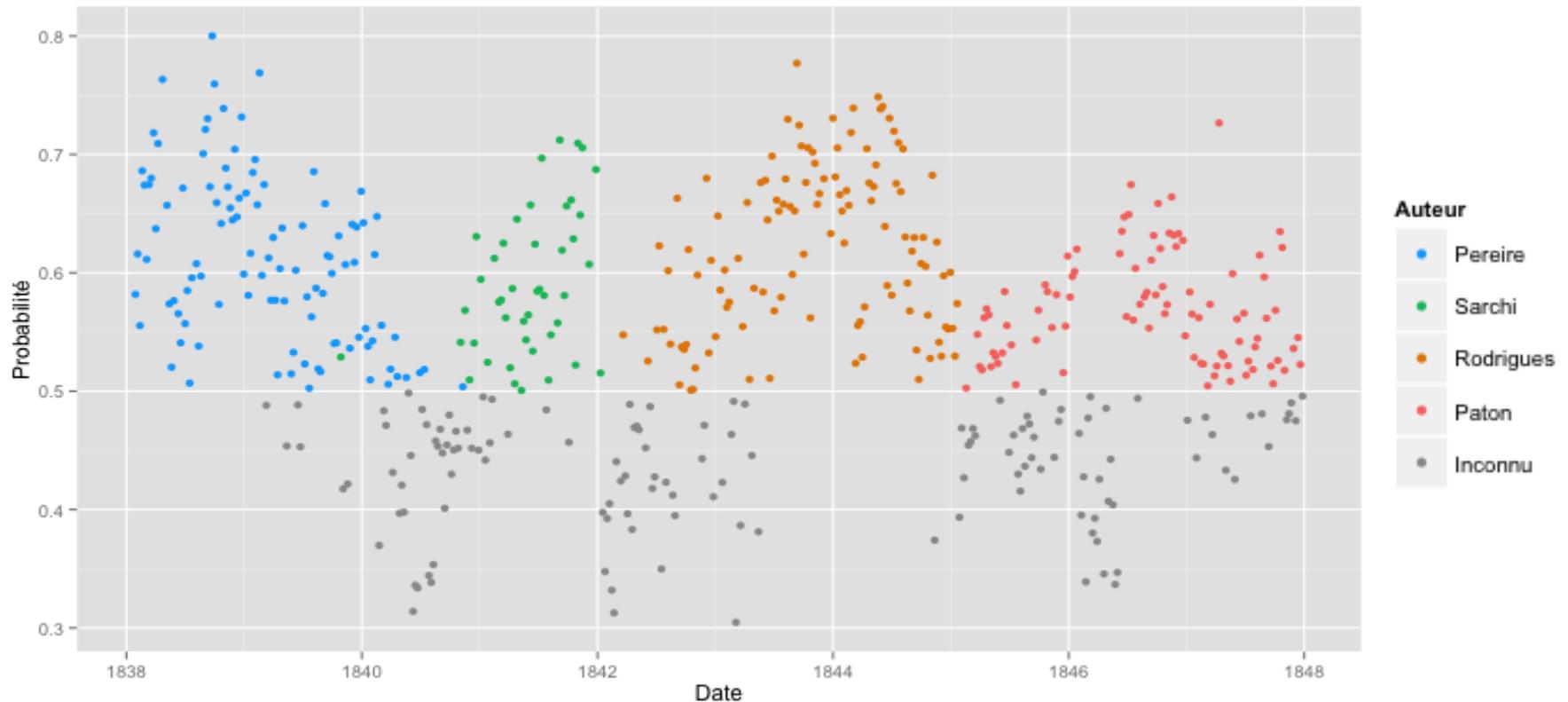


**3.**

# La codification d'un discours

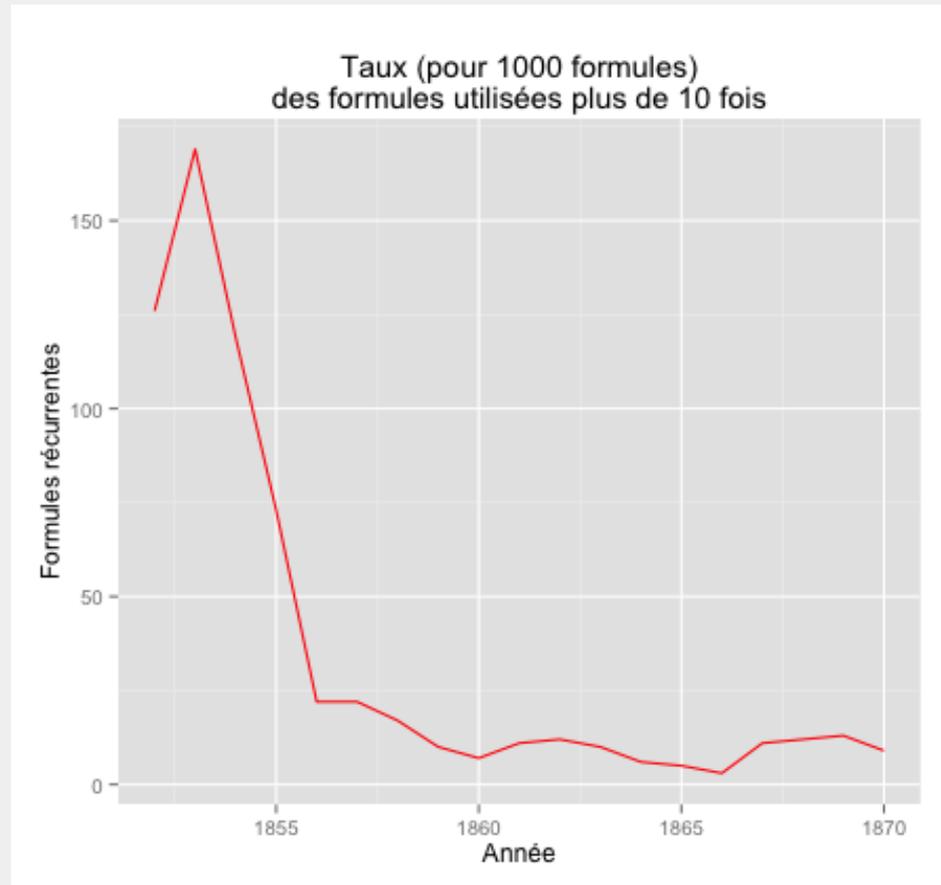
*De la forme hybride au genre  
médiatique*

# Une énonciation polyphonique



L'attribution automatisée des chroniques du *Journal des Débats* de 1838 à 1848 suggère l'existence de nombreux auteurs inconnus

# Un processus d'individualisation



Taux de formules récurrentes  
dans la chronique du *Journal des Débats* de 1852 à 1870

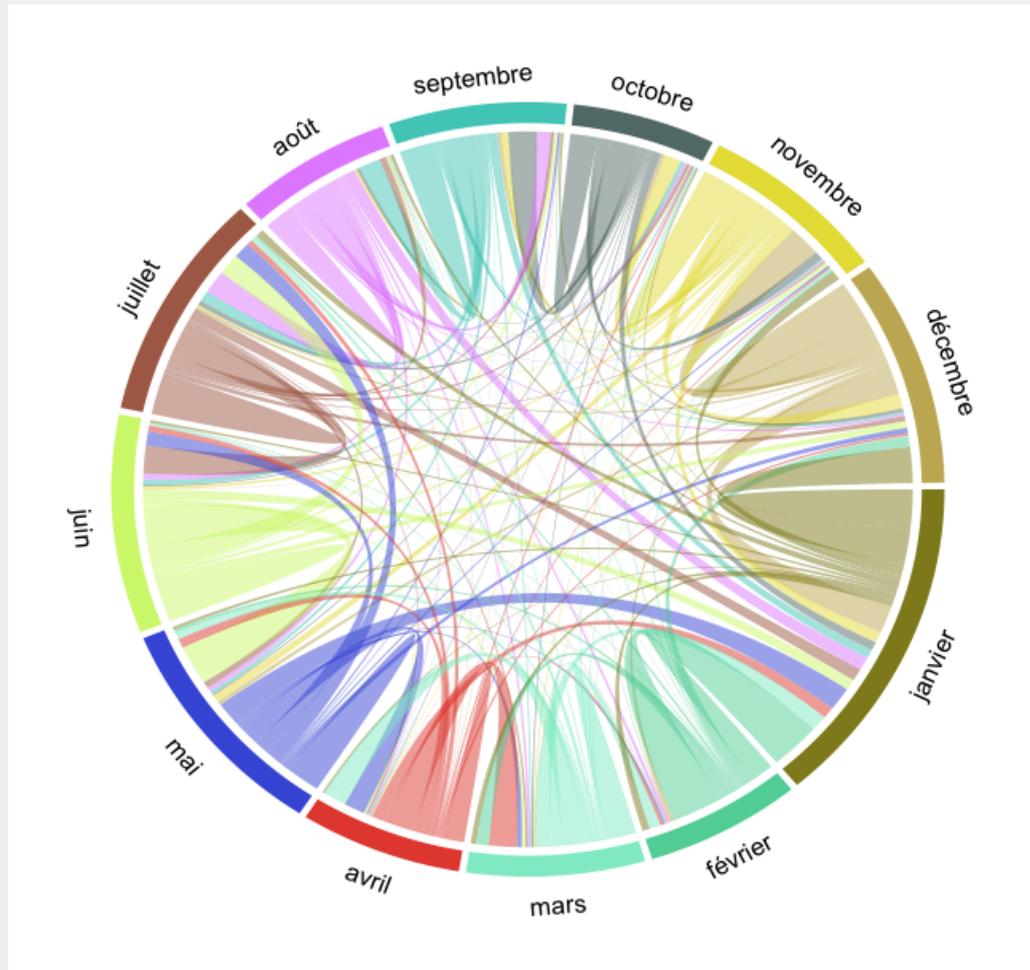
# Une poétique journalistique de la donnée

tes: On jugera de cette amélioration et des avantages pronostiqués pour un certain nombre de ces lignes par l'examen du tableau suivant, qui indique le dernier cours du samedi 13, le cours le plus haut atteint depuis lors, le dernier cours du samedi 20, et enfin la hausse d'un samedi à l'autre.

	Cours du 13.	Cours plus haut.	Cours du 20.	Hausse.
Saint-Germain.....	590 »»	625 »»	625 »»	35 »»
Versailles (droite)....	335 »»	340 »»	340 »»	5 »»
Versailles (gauche)....	247 50	255 »»	253 75	6 25
Orléans.....	1170 »»	1205 »»	1190 »»	20 »»
Rouen.....	687 50	715 »»	700 »»	12 50
Havre.....	275 »»	280 »»	278 75	3 75
Marseille.....	255 »»	275 »»	262 50	7 50
Bâle.....	230 »»	235 »»	231 25	1 25
Centre.....	550 »»	577 50	568 75	18 75
Bordeaux.....	500 »»	580 »»	567 50	67 50
Nord.....	628 75	658 75	648 75	20 »»
Strasbourg.....	528 75	567 50	547 50	18 75
Lyon.....	595 »»	617 50	605 »»	10 »»
Nantes.....	322 50	357 50	346 25	23 75
Boulogne.....	320 »»	335 »»	335 »»	15 »»
Montereau.....	162 50	170 »»	165 »»	2 50
Dieppe.....	232 50	240 »»	235 »»	2 50
La Teste.....	106 25	110 »»	110 »»	3 75

La chronique du *Siècle* (22 mars 1852) utilise le tableau à des fins de démonstration rhétorique

# Du périodique au cyclique



Les références aux mois dans les *Débats* de 1852 à 1870



“

Voir et surtout prévoir au jour le jour les actes de la Bourse n'est pas chose facile mais une étude faite de haut et embrassant de larges périodes permet, en exposant les causes, de déduire certains effets

**Mortimer d'Ocagne, *La Presse*, 6 octobre 1862**

# Les urgences de l'écriture

## 211<sup>e</sup> jour.—Bourse de Paris et Bulletin financier du vendredi 30 juillet 1858.

2 heures.—Le début du parquet présageait une nouvelle répétition des journées précédentes. Les prix étaient lourds sur la rente, et il y avait absence absolue d'affaires. Cependant, la Bourse présente depuis une demi-heure une amélioration sensible. Deux ou trois achats de 3 0/0 ont suffi pour relever le cours de 68 10 à 68 20. C'est aujourd'hui que la Conférence de Paris tient sa 13<sup>e</sup> séance, qui avait été ajournée par suite d'une indisposition de l'ambassadeur de Prusse, et l'on croit que les protocoles seront signés dans cette séance.

Les Consolidés anglais sont venus deux fois en hausse, par suite des dernières nouvelles favorables de l'Inde. La première cote était de 95 7/

fléchi aujourd'hui. On les négociait à 70 fr. On a repris ensuite à 72 50.

Les Petites-Voitures étaient plus fermes à 33 75. On signe en ce moment au siège de cette Compagnie une pétition à l'empereur afin d'obtenir le dégrèvement des droits payés par la Société à la ville de Paris. Les actions des Docks étaient faibles de 171 75 à 172.

Les cours étaient nominaux sur presque toutes les autres valeurs. Les Omnibus de Paris étaient calmes à 852 50; le Gaz parisien à 717 50; les Immeubles de Rivoli à 95.

Les actions du Crédit mobilier ont été tenues pendant quelque temps à 610; elles viennent de remonter à 613 75.

À partir de mars 1852, les bulletiniers des journaux du soir délivrent leurs bulletins juste avant l'impression.



**Conclusion**

# Crédits

Un grand merci à ces concepteurs de ressources libres.

- Modèle de présentation par SlidesCarnival
- Cadenas ouvert et fermé : Joel Bryant pour The Noun Project, CC-by-SA 3.0 ; Icône de journal : RRZEicons, CC-by-SA 3.0 ; Icône de loupe : David Vignoni