The Numapresse project

Tools & Methods

Impresso — July 5th 2018

Pierre-Carl Langlais
@Dorialexander
scoms.hypotheses.org
Alexander Doria (Wikipedia)

Our focus: the newspaper

The project builds up on distant reading of digitized news archives (chiefly by the French National Library) using the high computing infrastructure of Humanum.

The corpora comprises several hundred thousands documents in XML (about 1,5 to) from leading French dailies and weeklies (1814-1945).

Documents de presse numérisés en mode « OCR » du projet Europeana Newspapers

Mots-clefs: Presse, gallica, Europeana Newspapers

Ce jeu de données contient les documents numériques des collections de presse traitées durant le projet européen Europeana Newspapers avec une reconnaissance du texte (OCR, *optical character recognition*).

Sommaire

Contenu du jeu de données

Ce jeu contient la transcription réalisée par OCR d'environ 275 000 fascicules des collections de presse de Gallica traitées durant le projet Europeana Newspapers.

Tous les documents numérisés des titres suivants sont présents :

- Le Figaro
- L'Echo de Paris
- L'Univers
- La Presse
- L'Humanité
- Le Constitutionnel
- Le Petit Journal
- Le Siècle
- L'Action Française
- L'Intransigeant

Fiche technique

Date de création ou de mise à jour :
2015

Quantité :
1 287 500 pages, 275 000 fascicules

Langue :
Français

Formats :
METS, ALTO

Licence :
Gallica

Domaine :
Gallica , dans le même domaine

Télécharger

Métadonnées (23 Mo)
L'Action française (9 Go)
L'Echo de Paris (17 Go)
L'Humanité (9 Go)
L'Intransigeant (16 Go)
L'Univers (10 Go)
La Croix (12 Go)
La Presse (17 Go)
Le Constitutionnel (17 Go)
Le Figaro (21 Go)
Le Petit Journal (18 Go)
Le Siècle (21 Go)
Le Temps (25 Go)

Our focus: the newspaper

In Numapresse, newspaper archives are not studied as a set of historical facts but *as* newspapers and as a part of a wider media ecosystem.

In practice we are bound to raise the following issues, which may not always be solved by digital methods:

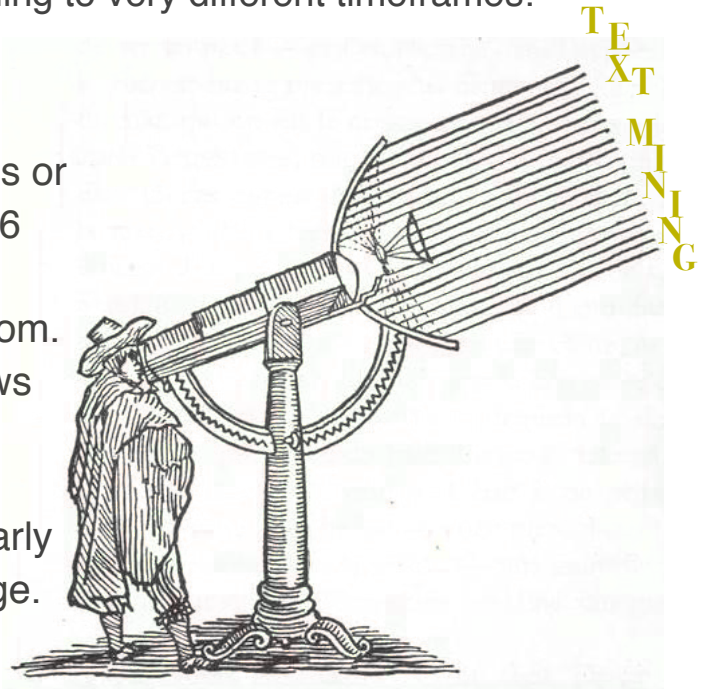
- What are the leading news genre? Where do they come from and how have they evolved?
- What are the main poetic features of newspaper genres and discourses? How are the heterogeneous components (texts, data, images…) of a newspaper page articulated?
- Who is writing the newspaper? How is the newsroom organized?
- How do newspapers connect to one another? By what means do news objects travel and are reprinted and adapted across the world?

Text mining as a ajustable lenses

The leading dailies are published continuously for a very long time (as much as 145 years for the *Journal des débats*) and cultural transformations can occur according to very different timeframes:

- Several weeks for the dissemination of a news or short term contracts (such as serial novels).
- Several months or a year to uncover some editorial experiments or to analyze the *regular* flow of reprints in the media ecosystem (6 months according to Cordell)
- Several years for the inner sociological workings of the newsroom.
- Several decades for the emergence and metamorphosis of news genre.

Clearly to deal with all theses issues, we need to be able to regularly change our focus and shift between punctual and structural change.



Zoom reading

Not only are distant and close reading constantly articulated within the project, but we usually experiment with different *distances*

Text mining as a ajustable lenses

The XML files generated by the Library do not contain only the OCRed text but numerous contextual information that has been fundamental to integrate the multiple layers of the newspaper

- **Coordinates and persistent identifiers** of words, lignes, text blocks and images.
- **OCR confidence**
- **Size of the text and typographical layout** (bold, italic, font…).

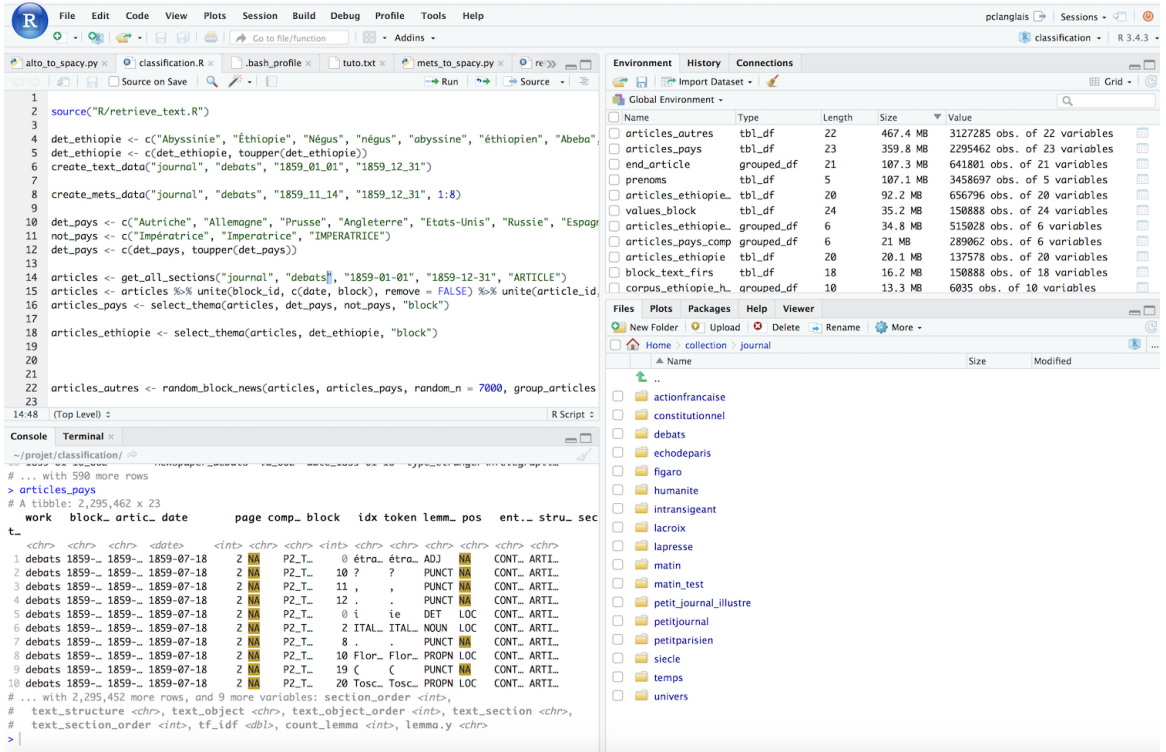

```
– <Layout>
  – <Page ACCURACY="99.68" HEIGHT="9782" ID="PAG_1" PHYSICAL_IMG_NR="1" QUALITY="OK" WIDTH="7166">
    – <PrintSpace HEIGHT="8513" HPOS="230" ID="PAG_1_PrintSpace" VPOS="440" WIDTH="6356">
      – <TextBlock HEIGHT="376" HPOS="2704" ID="PAG_1_TB000001" LANG="fr" STYLEREFS="TXT_1" TAGREFS="TAG_ST001" VPOS="509" WIDTH="23
        – <TextLine HEIGHT="376" HPOS="2704" ID="PAG_1_TL000001" STYLEREFS="TXT_1" VPOS="509" WIDTH="2354">
            <String CONTENT="LA" HEIGHT="372" HPOS="2704" ID="PAG_1_ST000001" STYLEREFS="TXT_1" VPOS="509" WC="1" WIDTH="574"/>
            <SP HPOS="3278" ID="PAG_1_SP000001" VPOS="543" WIDTH="1694"/>
            <String CONTENT="LIBERTÉ" HEIGHT="342" HPOS="4972" ID="PAG_1_ST000002" STYLEREFS="TXT_1" VPOS="543" WC="1" WIDTH="86"/>
          </TextLine>
        </TextBlock>
      – <TextBlock HEIGHT="60" HPOS="519" ID="PAG_1_TB000002" LANG="fr" STYLEREFS="TXT_2" VPOS="440" WIDTH="519">
        – <TextLine HEIGHT="60" HPOS="519" ID="PAG_1_TL000002" STYLEREFS="TXT_2" VPOS="440" WIDTH="519">
            <String CONTENT="Mardi" HEIGHT="52" HPOS="519" ID="PAG_1_ST000003" STYLEREFS="TXT_2" VPOS="448" WC="1" WIDTH="149"/>
            <SP HPOS="668" ID="PAG_1_SP000002" VPOS="447" WIDTH="29"/>
            <String CONTENT="1er" HEIGHT="47" HPOS="697" ID="PAG_1_ST000004" STYLEREFS="TXT_2" VPOS="447" WC="0.65" WIDTH="73"/>
            <SP HPOS="770" ID="PAG_1_SP000003" VPOS="444" WIDTH="24"/>
            <String CONTENT="mai" HEIGHT="50" HPOS="794" ID="PAG_1_ST000005" STYLEREFS="TXT_2" VPOS="444" WC="1" WIDTH="90"/>
            <SP HPOS="884" ID="PAG_1_SP000004" VPOS="440" WIDTH="24"/>
            <String CONTENT="1866" HEIGHT="51" HPOS="908" ID="PAG_1_ST000006" STYLEREFS="TXT_2" VPOS="440" WC="0.88" WIDTH="130"/>
          </TextLine>
        </TextBlock>
      – <TextBlock HEIGHT="52" HPOS="374" ID="PAG_1_TB000003" LANG="fr" STYLEREFS="TXT_3" VPOS="548" WIDTH="635">
        – <TextLine HEIGHT="52" HPOS="374" ID="PAG_1_TL000003" STYLEREFS="TXT_3" VPOS="548" WIDTH="635">
            <String CONTENT="papier" HEIGHT="39" HPOS="374" ID="PAG_1_ST000007" STYLEREFS="TXT_3" VPOS="561" WC="1" WIDTH="116"/>
            <SP HPOS="490" ID="PAG_1_SP000005" VPOS="560" WIDTH="19"/>
            <String CONTENT="et" HEIGHT="28" HPOS="509" ID="PAG_1_ST000008" STYLEREFS="TXT_3" VPOS="560" WC="1" WIDTH="33"/>
            <SP HPOS="542" ID="PAG_1_SP000006" VPOS="558" WIDTH="18"/>
            <String CONTENT="Tirage" HEIGHT="38" HPOS="560" ID="PAG_1_ST000009" STYLEREFS="TXT_3" VPOS="558" WC="1" WIDTH="118"/>
            <SP HPOS="678" ID="PAG_1_SP000007" VPOS="554" WIDTH="202"/>
            <String CONTENT="18" HEIGHT="33" HPOS="880" ID="PAG_1_ST000010" STYLEREFS="TXT_3" VPOS="554" WC="1" WIDTH="31"/>
            <SP HPOS="911" ID="PAG_1_SP000008" VPOS="550" WIDTH="23"/>
            <String CONTENT="fr." HEIGHT="31" HPOS="934" ID="PAG_1_ST000011" STYLEREFS="TXT_3" VPOS="550" WC="1" WIDTH="39"/>
            <SP HPOS="973" ID="PAG_1_SP000009" VPOS="548" WIDTH="27"/>
            <String CONTENT=")" HEIGHT="37" HPOS="1000" ID="PAG_1_ST000012" STYLEREFS="TXT_3" VPOS="548" WC="0.00" WIDTH="9"/>
          </TextLine>
        </TextBlock>
```

Our method: pragmatism

- “**Whatever work**”. Not everything fits one algorithm and, in fact, not everything can be modeled. Conversely some research can only be undertook through programming and automation, such as reprint detection or genre classification..
- **Build on the past work**. Previous projects have already initiated extensive databases and/or created precious tools (*ViraltText*). We are particularly keen on reusing library data and formats in a comprehensive way.
- **Ensure collaborations**. Numapresse has been greatly stimulated by the collaborative spirit of news studies in France. Our next steps would be to produce intermediary tools formatted for the research on newspaper, that do not rely on an extensive technical knowledge.
- **Systematical application**. While small and medium scale experiments are extremely useful, we are more and more focusing on applying programs on a very long time frame (>50 years) and on comprehensive corpora.

Our tool: Humanum super-computer

The programming interface hosted by Humanum allow us to run scripts on very big corpora and morph them into very big tables and to work collaboratively.



Our tool: Humanum super-computer

```
> matin_1926 %>% select(work, date, id_style, block_id, article_id, idx, token, lemma, pos, ent.type)
# A tibble: 13,689,656 x 10
   work  date       id_style       block_id                 article…   idx token   lemma   pos   ent.
   <chr> <date>     <chr>          <chr>                    <chr>     <int> <chr>   <chr>   <chr> <chr>
 1 matin 1926-01-01 TXT_4 PAR_LEFT  1926-01-01_1_1_P1_TB00004 1926-01…     0 IÏÏF    IÏÏF    PROPN ORG
 2 matin 1926-01-01 TXT_4 PAR_LEFT  1926-01-01_1_1_P1_TB00004 1926-01…     5 LATEMP… LATEMP… PROPN NA
 3 matin 1926-01-01 TXT_4 PAR_LEFT  1926-01-01_1_1_P1_TB00004 1926-01…     0 IÏÏF    IÏÏF    PROPN ORG
 4 matin 1926-01-01 TXT_4 PAR_LEFT  1926-01-01_1_1_P1_TB00004 1926-01…     5 LATEMP… LATEMP… PROPN NA
 5 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…     0 Os      Os      DET   NA
 6 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…     3 balles  balle   NOUN  NA
 7 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…    10 strenn… strenn… ADJ   NA
 8 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…    19 msteor… msteor… NOUN  NA
 9 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…    33 »       »       PUNCT NA
10 matin 1926-01-01 TXT_5 PAR_BLOCK 1926-01-01_1_1_P1_TB00005 1926-01…    35 i       ie      ADP   NA
# ... with 13,689,646 more rows
```

R text mining process preserve the complexities of the original text. Each word is a line in the table, that contain all the informations included in the digitized archives and add new ones (lemma, syntax analysis, named entities…) generated by Spacy.

Plan

An introduction in five parts

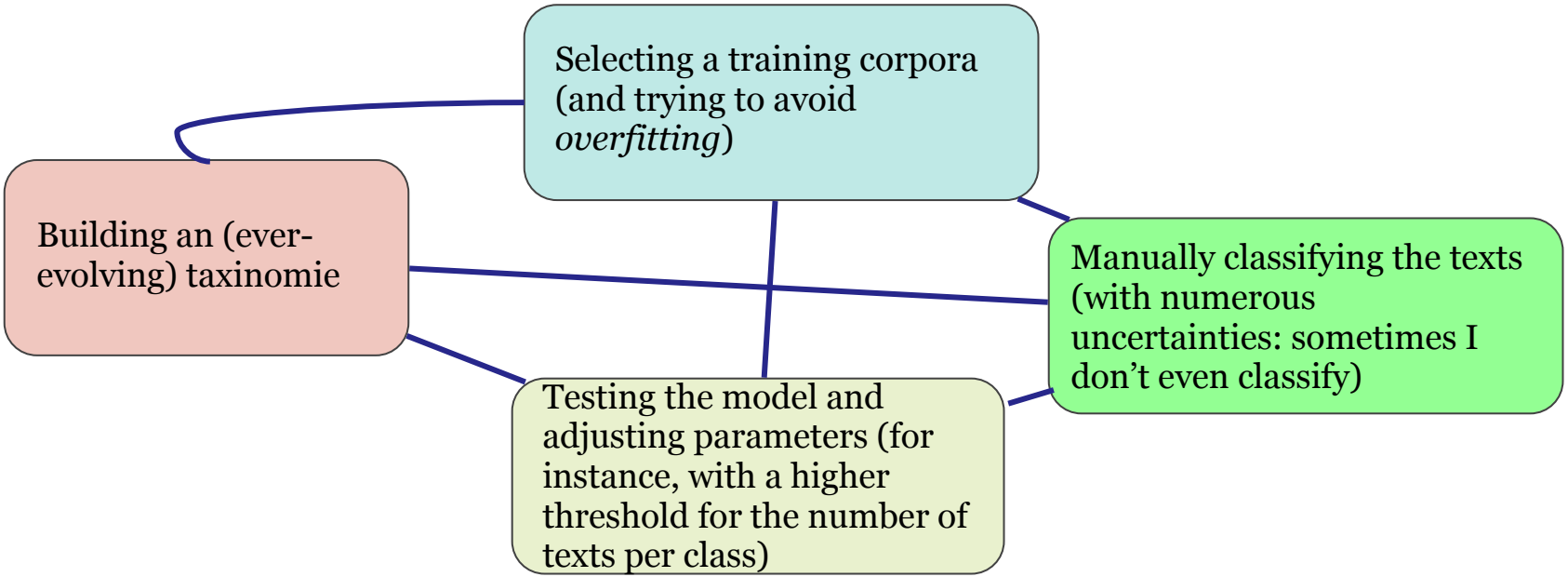
- Classify the newspaper : making automated recognition of newspaper genre and forms.
- Toward a poetic analysis of news images : tracking aesthetic mutations through *deep learning*.
- Reconstruction of editorial structures.
- Mapping news circulation.
- The Numapresse database

1. Classify the newspaper

Rules and grammar of automated recognition of newspaper genre and forms.

The principles of classification

A four-step process (with a lot of recursions)



The principles of classification

Un événement imprévu vient de mettre en émoi toute la ville d' Ussel , un des chefs-lieux d' arrondisse- ment de la Corrèze . M. Mignon , banquier de cette vn)e , a disparu subitement le samedi 23 de ce mois , à quatre heures et demie du matin ; il est softi de son domicile armé d' un fusil , et n' est pas rentré de tonte ta journée . Vers le soir , on s' est enquisde sa on a été vainement le chercher à Saint - Exu- péry , à Saint - Ange ! et ailleurs . La justice a fait pen- dant la nuit une descente à son domicile et a posé les N scellés . Dimanche dernier , 24 janvier , cette disparition faisait le sujet de toutes les conversations les uns , et c est le plus grand nombre , pensent que M. Mignonc s' est suicidé ; les autres , qu' il a passé à l' étranger . Ott . l' a cherché dans les champs et te long des rivières , mais on n' est parvenu jusqu' à présent à découvrir aucune trace . On a remarqué que son bureau était /"/ très en ordre teut y était étiqueté . On a.trouvé dans des sacs une quantité d' enets bien arrangés et envirom une dtzaine de mille francs en or et en argent . Som départ a été , à ce qu' on assure , sans préparatifs point de sacoches , pas de cheval ni de voiture . Voi) & ` tout ce que l' on sait actuellement . Encore qnehmes . jours , et ce mystère sera peut-être expliqué . On écrit de Luçon (Vendée) , le 27 janvier Un ouragan épouvantable , qui a eu lieu dans 1~ mut du 25 au 26 de ce mois , a démoli une partie d & la voûte de notre belle et antique cathédrale , qui avait depuis bien des siècles résisté aux terribles coups de vent de sud-ouest . Le magnifique jeu s' est trouvé abtmé dans cette chute , ainsi les belles sculptures en bois qui le décoraient . Le vent a été si violent , que le coq qui surmontait ta ° Hèche , et qui devait peser au moins de 25 à 30 a été enlevé et jeté à plus de 300 mètre : . Le drapeau tricolore en tôle , qui se trouvait aux basses galeries de la nèche depuis la révolution de Juillet , et qui était maintenu par trois barres de fer dont chacune avait au moins 12 .ou 15 centimètres de tour , a été arraché et est allé tomber à plus de 300 ' mètres dans la cour d' un habitant , qui a. été fort étonné de trouver chez lui cet étendard national . Heureusement il n' y avait alors personne dans car autrement , que de malheurs de plus à On estime à 40,000 ou 50,000 fr . au moins causé par cet ouragan

A genre is marked by the use of specific words (here, a *fait divers* of 1844)

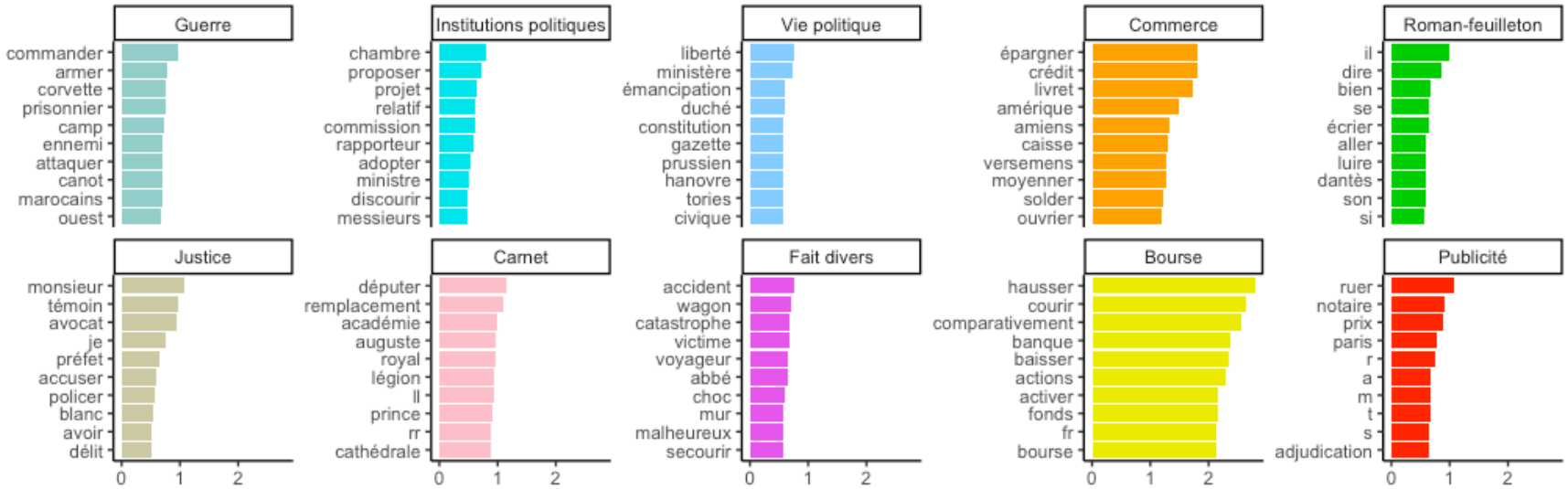
The principles of classification

Currently, two models have been trained using supervised SVM.

- A « 1835-1855 » model, from 10 annotated issues of the *Journal des débats*. It recognizes 10 different genres..
- A « 1920-1940 » model from 20 annotated issues of *Le Petit Journal*, *L'Intransigeant*, *Le Matin* and *Le Petit Parisien*. It recognizes 20 different genres.

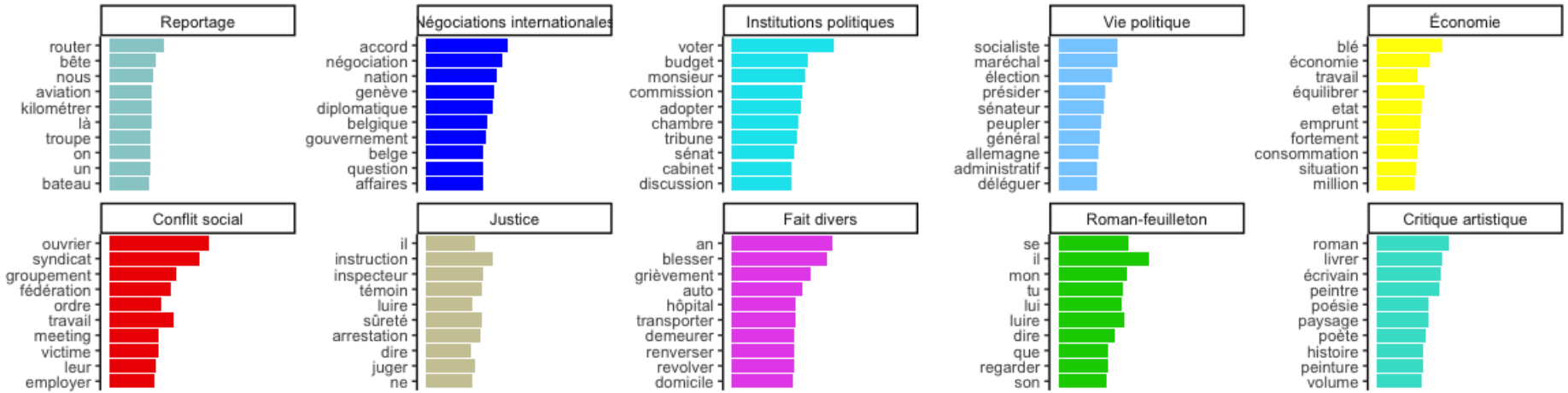
Later on, models will be implemented for every 20-30 years period of the history of the French daily press, as well as specialized models for some corpora (for instance, news magazine).

1835-1855 model



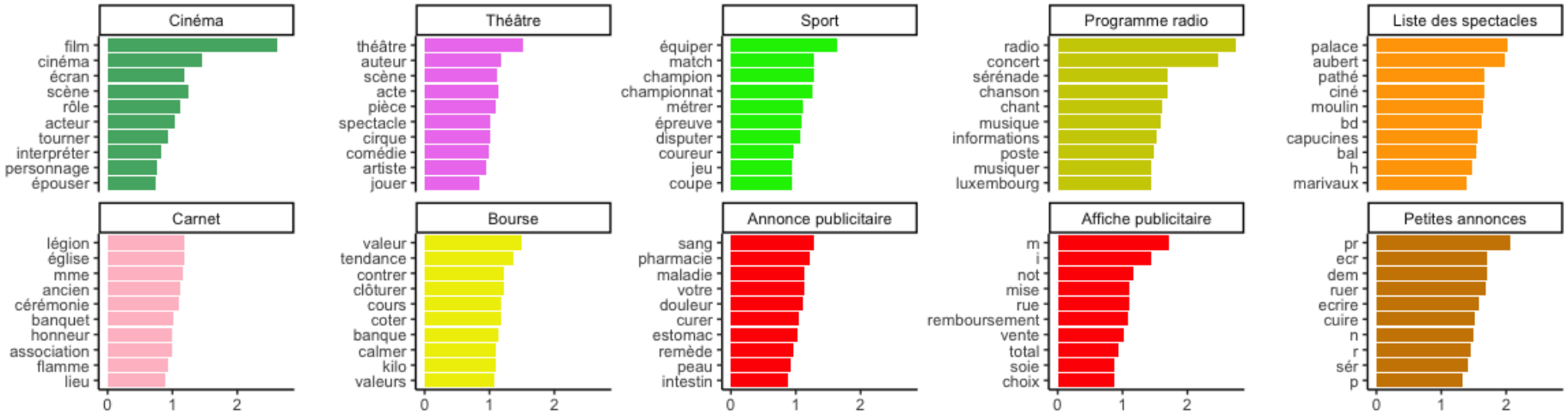
Models contain a “lexical sketch” of every genre, from the words that characterize them the most.

1835-1855 model



Models contain a “lexical sketch” of every genre, from the words that characterize them the most.

1835-1855 model



Models contain a “lexical sketch” of every genre, from the words that characterize them the most.


27 JUILLET 1857.

JOURNAL DES DÉBATS

POLITIQUES ET LITTÉRAIRES.

JEUDI.

Grèce.

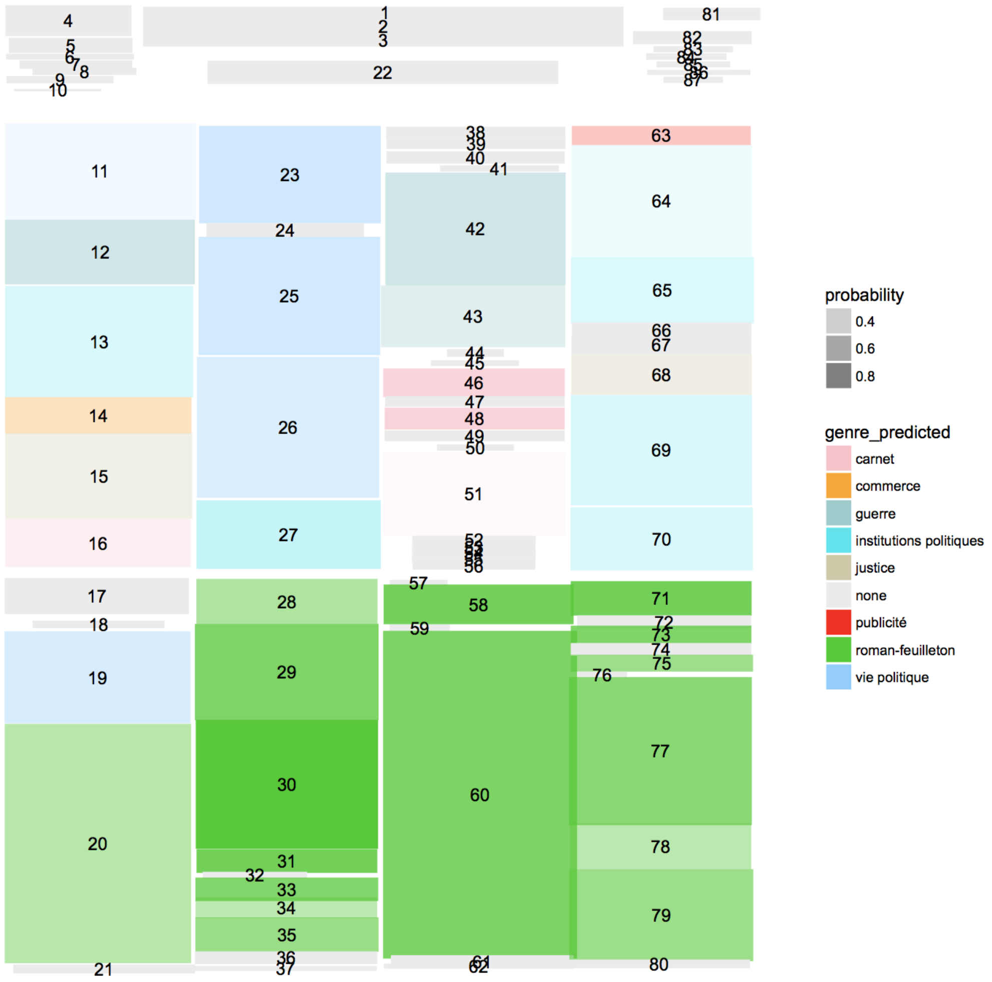
Grande-Bretagne.

FRANCE.

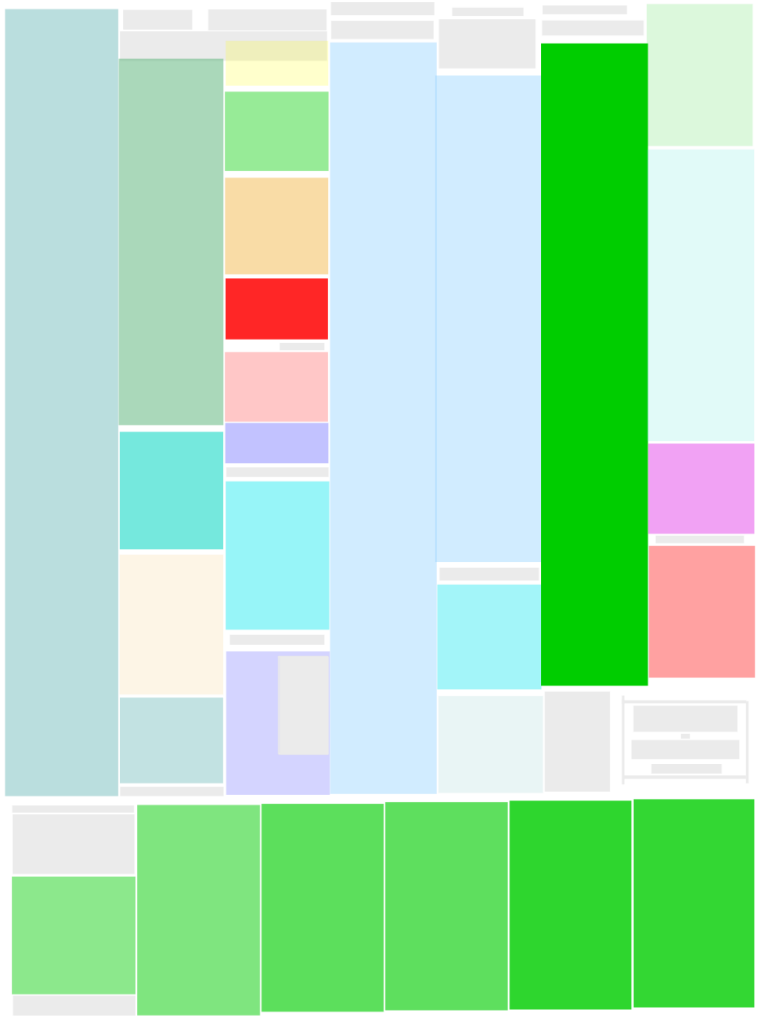
Nouvelles d'Espagne.

Feuilleton du Journal des Débats.

UNE AVERSE.



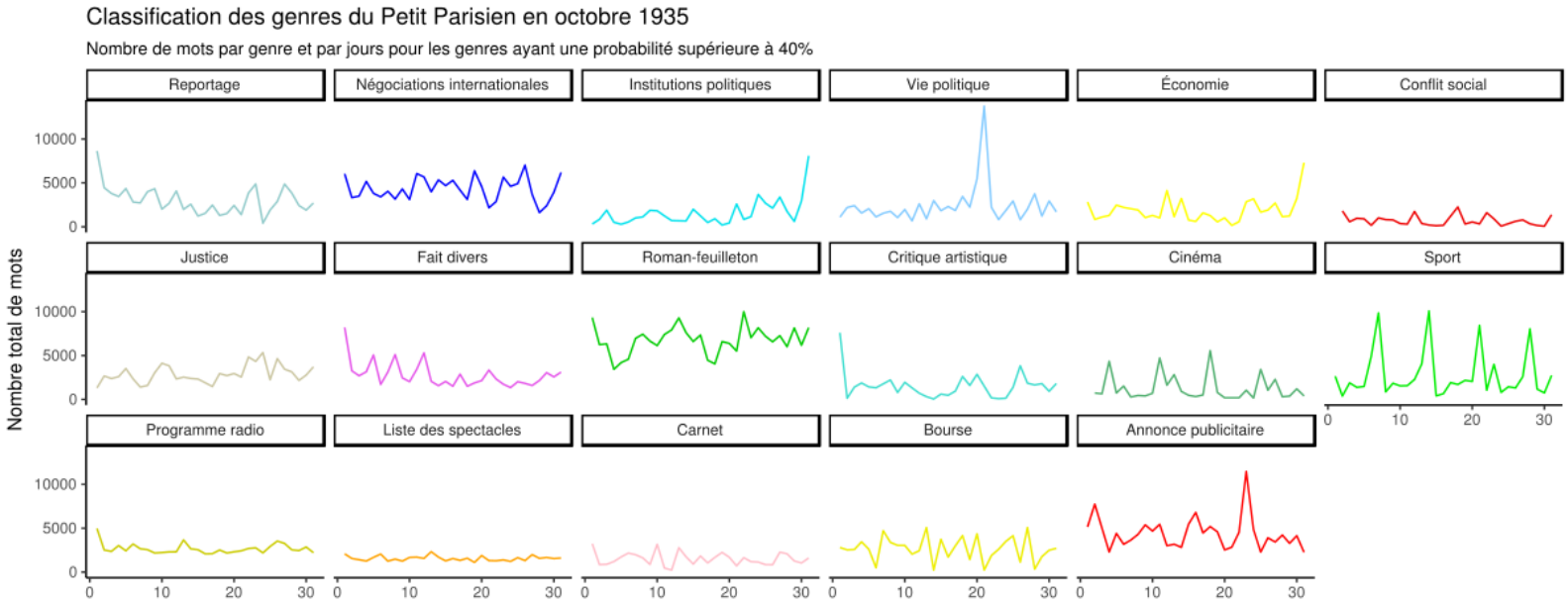

NOS ECHOS
La vie à Shanghaï
Le numéro 10
LAMODE
COURRIER DU SPHINX
Les nuits tragiques de Londres




LA PAGE DU CINÉMA
LA PROPAGANDE A REBOURS
L'activité en France
critiques
L'activité à l'Étranger
ECHOS
SALLES
BABEL
DICTION
RIALTO-CINÉMA
LA VALSE DE L'ADIEU
PARAMOUNT
A MARIVAUX
LA VIE PRIVÉE D'HÉLÈNE DE TROIE
CINÉMA MARBEUF
BEN-HUR
GAUMONT-PALACE
LE GORILLE
AU CAMEO
L'ÉTUDIANT PAUVRE
AU CINÉ MAX LINDER


théâtre : 47.68%
cinéma : 26.14%
critique littéraire et artistique : 15.4%
cinéma
critique littéraire et artistique
économie
justice
liste des spectacles
négociations internationales
none
programme radio
publicité_affiche
publicité_annonce
reportage
sport
théâtre
vie politique

Zooming out the genres



Although imperfects, models still allows to observe genre evolutions on a longer time frame, and to identify structural tendencies and punctual breaks.

Zooming out the genres

DERNIÈRE ÉDITION

Le Petit Parisien

(LE PLUS LU DES JOURNAUX DU MONDE ENTIER)

LUNDI 21 OCTOBRE 1935

LES ÉLECTIONS SÉNATORIALES

M. Pierre Laval l'emporte, dès le premier tour, à la fois dans la Seine et dans le Puy-de-Dôme

67 ANCIENS ET 40 NOUVEAUX DONT 17 DEPUTES

LE CALVAIRE D'UN PÈRE

Le commandant Marescot a pleuré et prié devant l'horrible tombeau de sa fillette

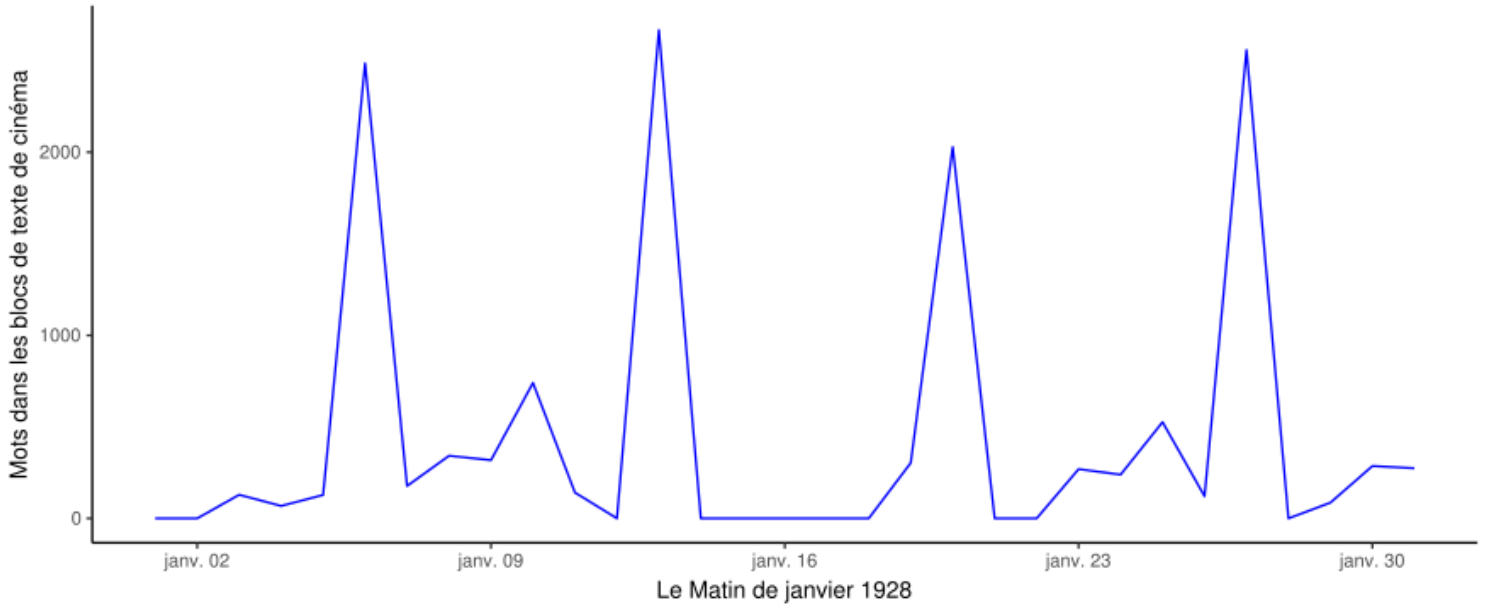
LE DOCTEUR PAUL PROCÉDERA DEMAIN A L'EXAMEN MÉDICO-LÉGAL DE L'INNOCENTE VICTIME

Although imperfects, models still allows to observe genre evolutions on a longer time frame, and to identify structural tendencies and punctual breaks.

Zooming out the genres

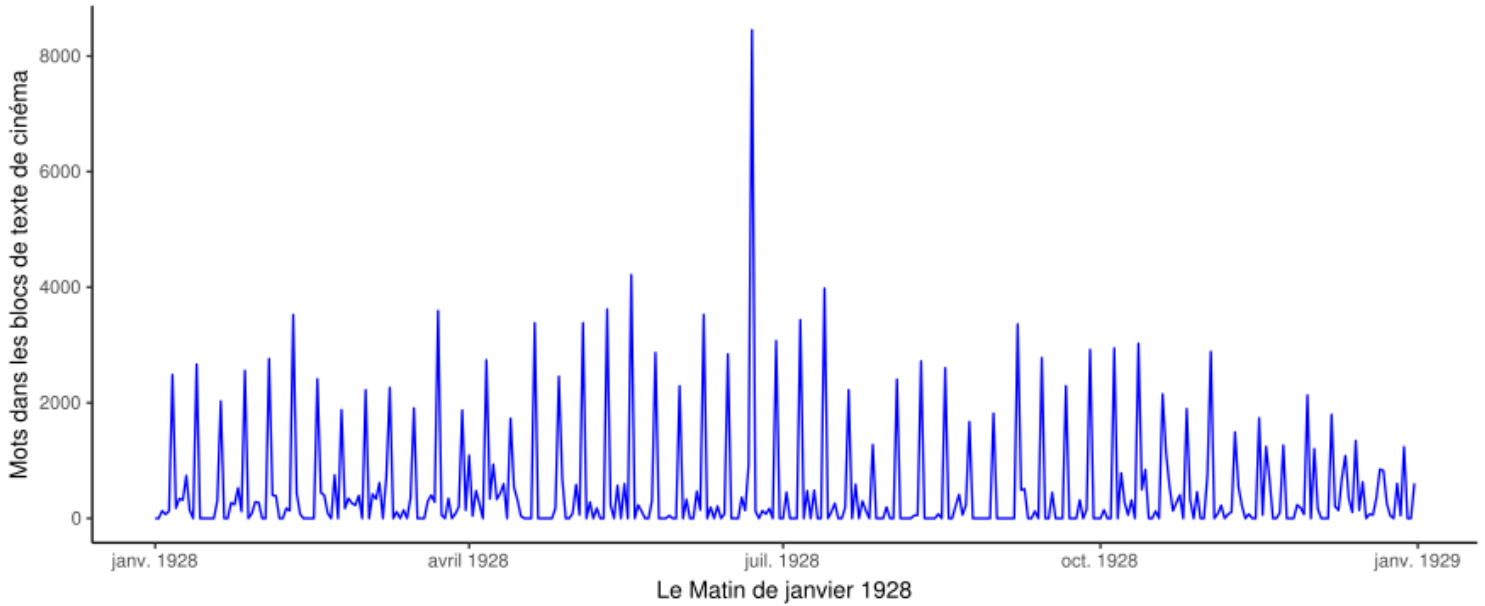
Although imperfects, models still allows to observe genre evolutions on a longer time frame, and to identify structural tendencies and punctual breaks.

Zooming out the genres



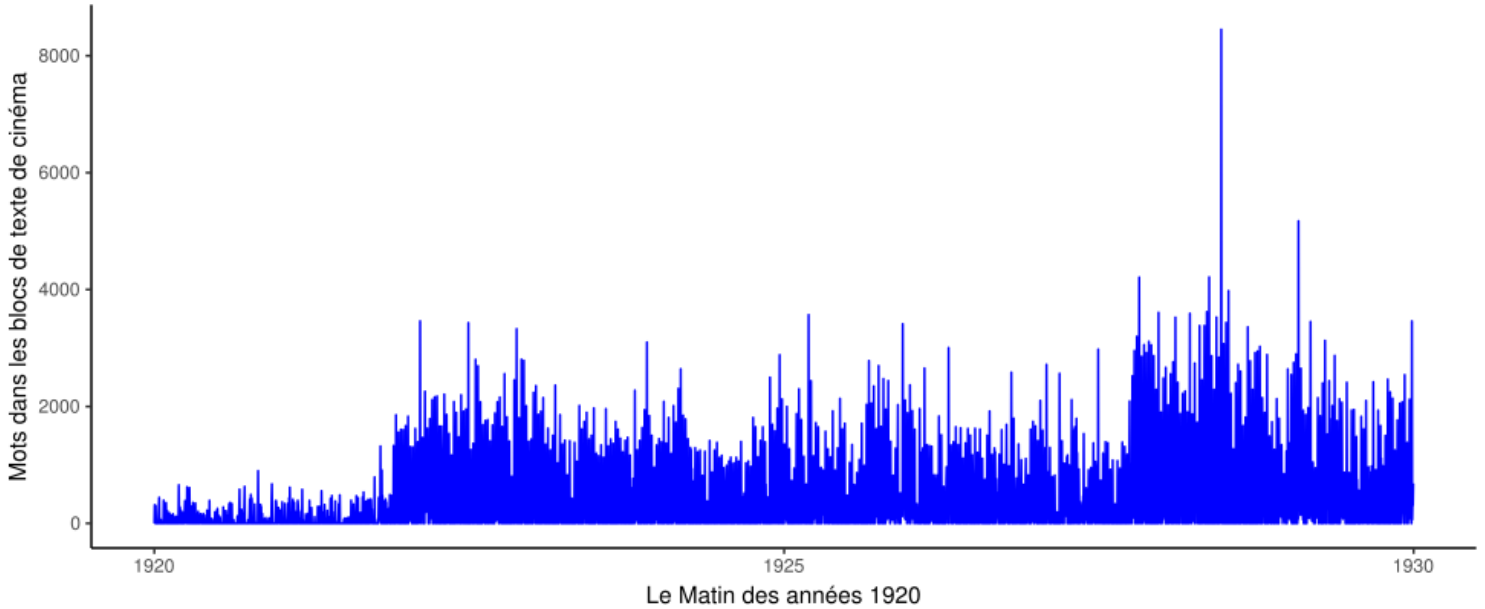
The model highlights the weekly rhythm of movie critics in January 1928, caused by the publication of a supplement on Friday.

Zooming out the genres



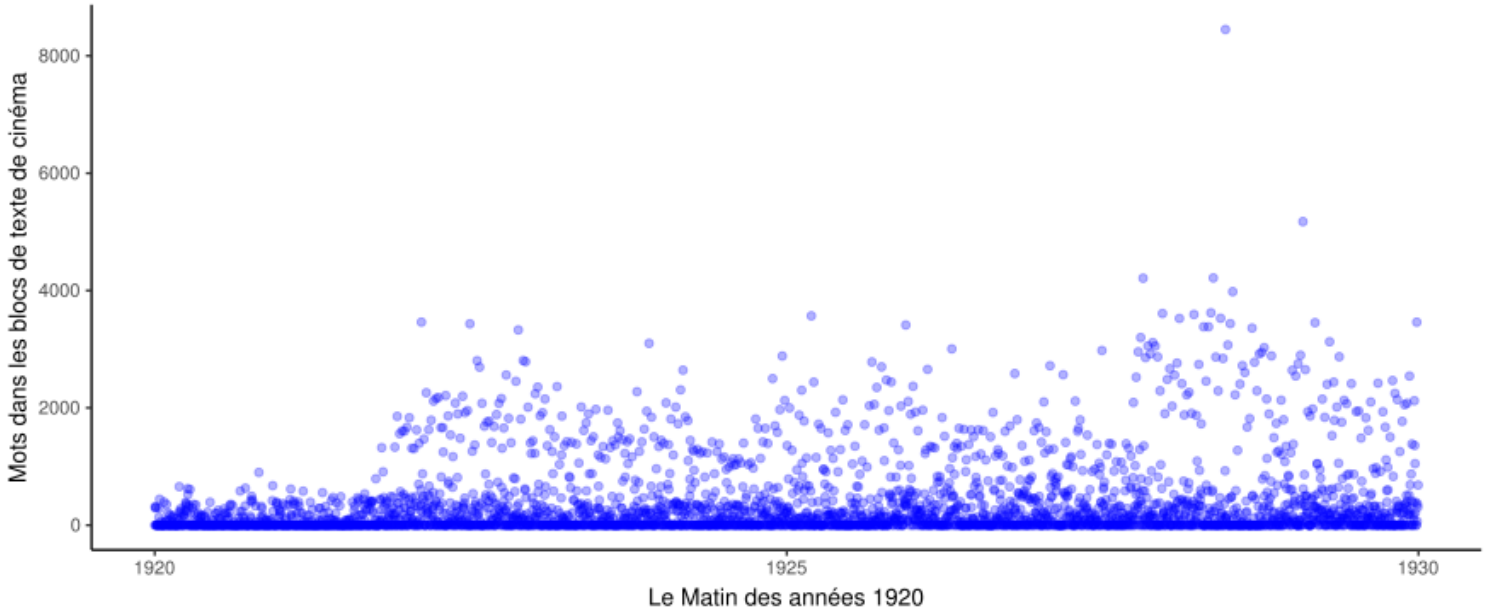
This cycle persists the whole year, with occasional variations of the length of the Friday supplement.

Zooming out the genres



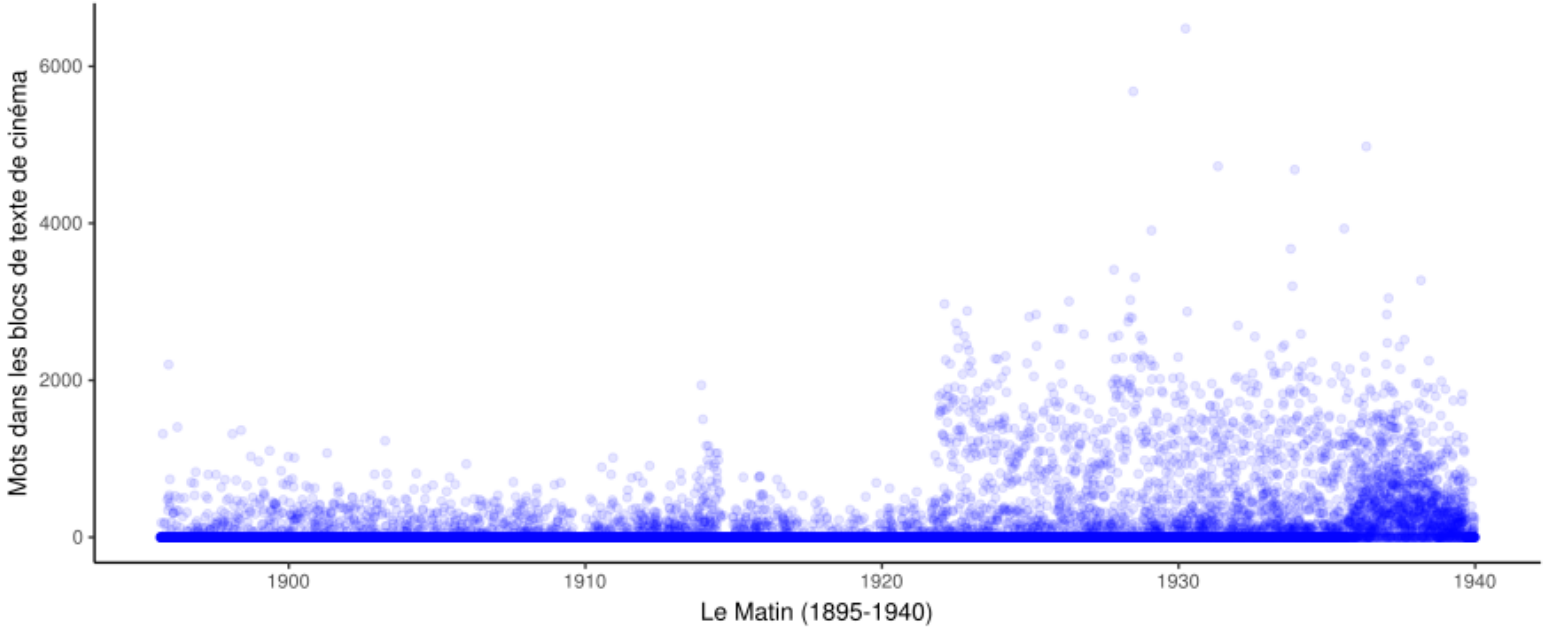
The weekly rhythm is itself part of a general growth in movie coverage across the 1920s

Zooming out the genres



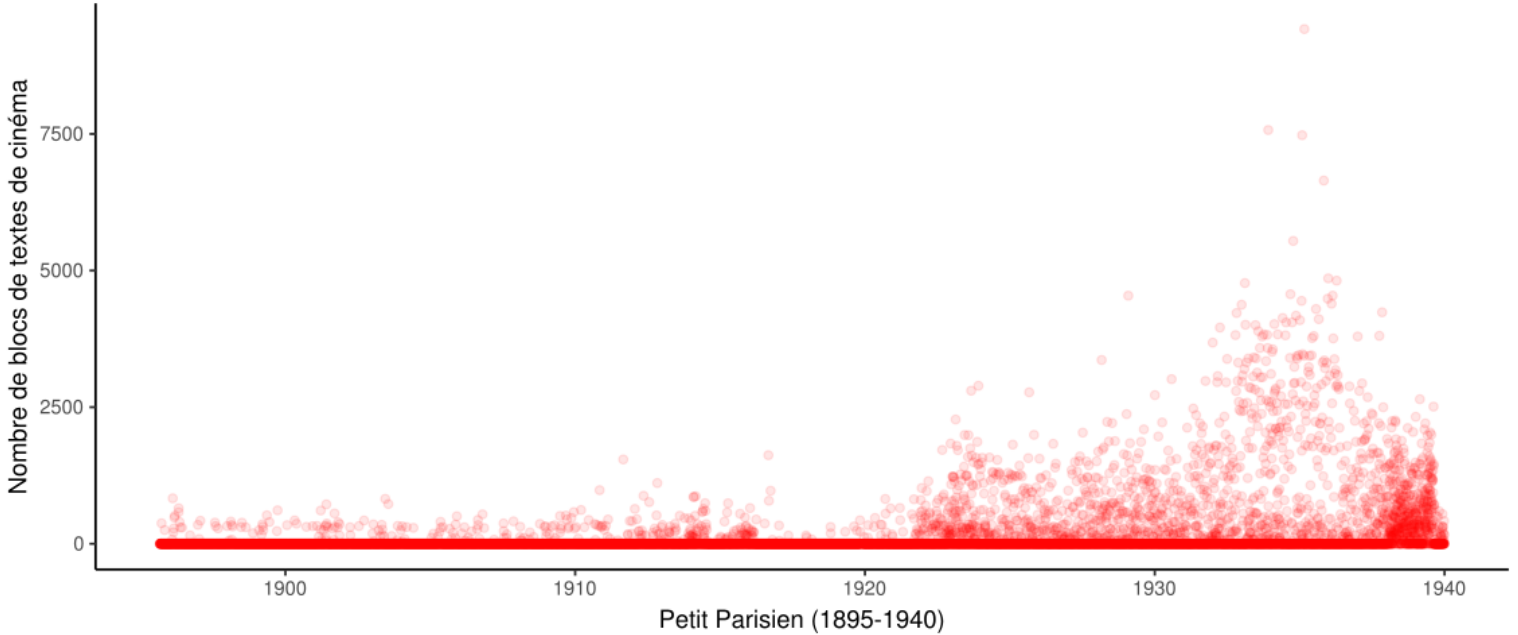
The weekly rhythm is itself part of a general growth in movie coverage across the 1920s

Zooming out the genres



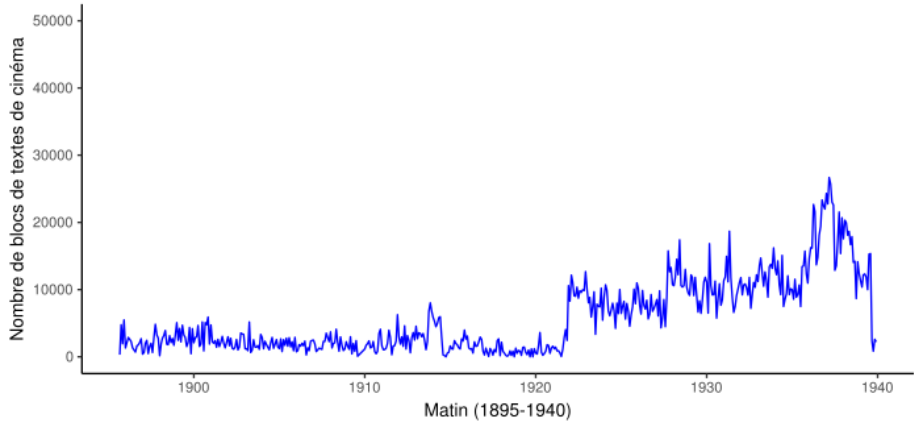
…and across the first half of the 20th century, after a momentary gap during WWI

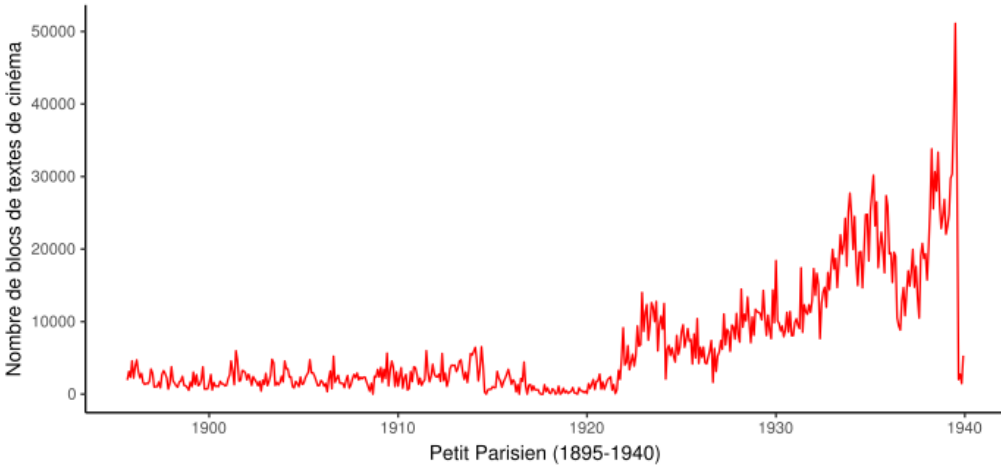
Zooming out the genres



The *Petit Parisien* is marked by a somewhat similar evolution, although, movie coverage extends more dramatically by the end of the 1930s.

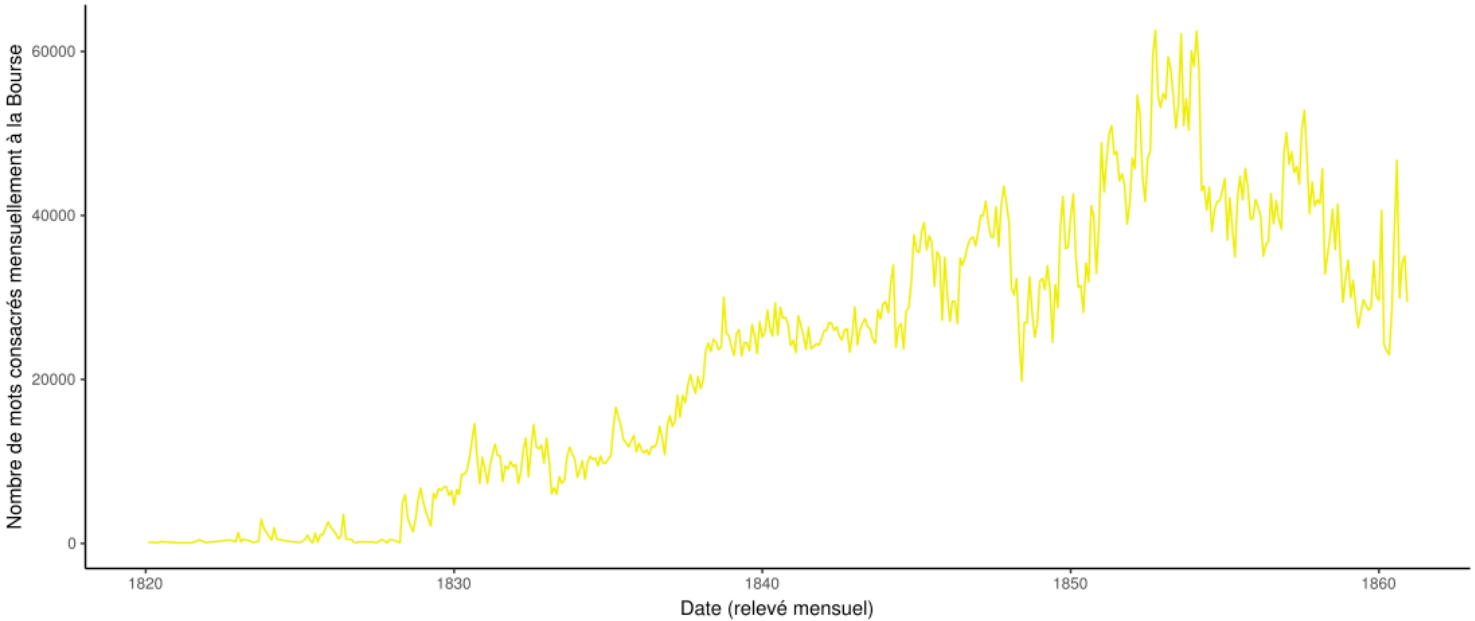
Zooming out the genres





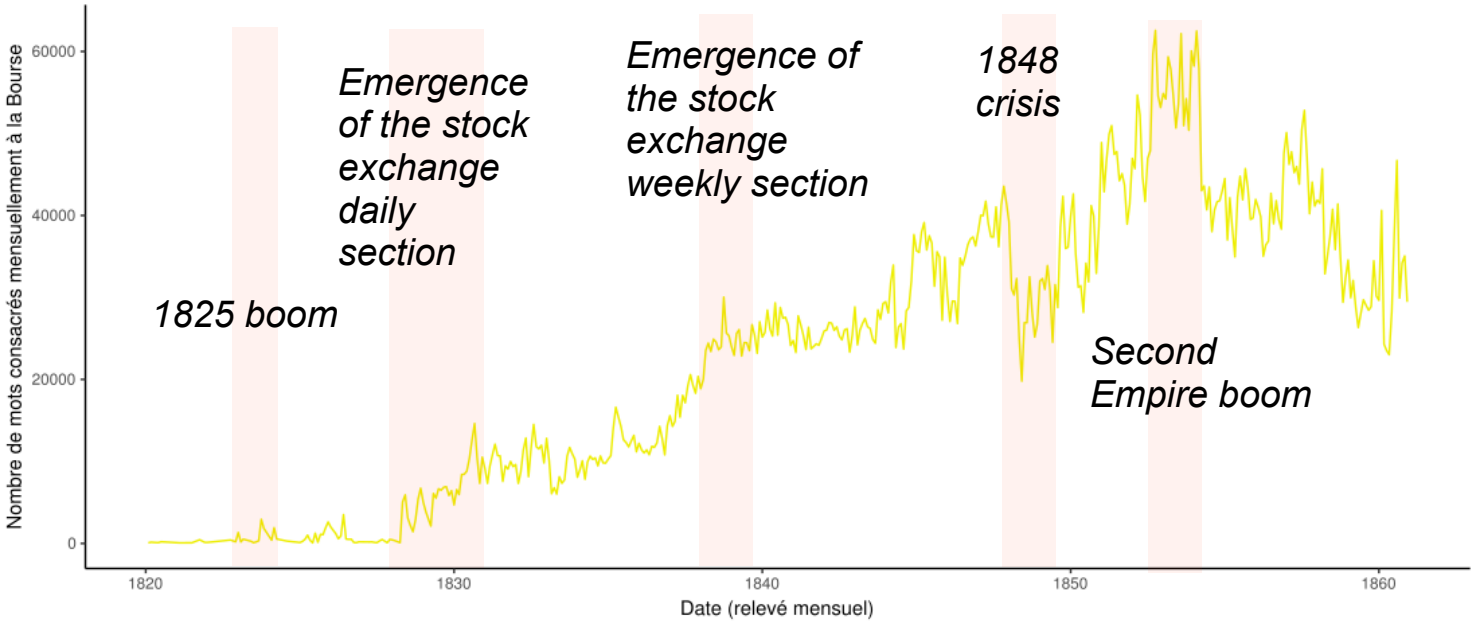
The *Petit Parisien* is marked by a somewhat similar evolution, although, movie coverage extends more dramatically by the end of the 1930s.

Zooming out the genres



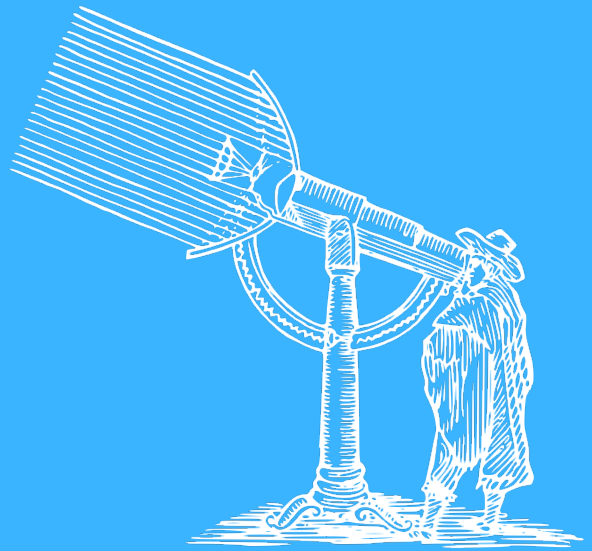
WWI is an exceptional factor, but genre evolution is frequently affected by outward historical events, such as economic crisis and booms with the stock section from 1820 to 1870

Zooming out the genres



WWI is an exceptional factor, but genre evolution is frequently linked to outward historical events, such as economic crisis and booms with the stock section from 1820 to 1870

What is the use of classification?



Generate corpora *on-demand*

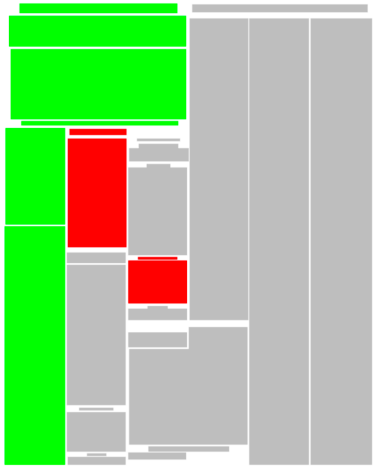
Écritures du sport

Le Matin du 1912-03-18, article n°78, page n°5

- lundi jsj3ox»£JJF

t % Les demi-finales du championnat de France de football rugby 3 FOIMTS COMTRE CE MATCH MÉMORABLE DURA 120 MINUTES

n ne fallut pas moins de deux mi-temps ! de iO minutes chacune, de deux- prolongations, chacune de 20 minutes, pour que le Ciub de France battît le Stade Bordelais, au cours d'une partie mémorable et quelque peu angoissante pour les partisans des deux équipes. Cette partie prendra date dans l'histoire du football français, et l'on dira plus tard : Vous rappelez- vous ? C'était l'année où le Racing et Bordeaux firent deux prolongations. Car il est rare de voir une demi-finale aussi ardemment disputée que le fut celle d'hier. Les- deux équipes se présentèrent ainsi ù Colombes, où vingt mille spectateurs se pressaient autour du terrain de jeu : Ttacino Club : Lagarrlgues ; André, Lane, Burguri et Fâilliot : Cooper et ïequier ; Monniot, Fonsèque, Guillemin. Combemale, Bellen'ger, t erou. Vives et Decamps. Stade Bordelais : Tachoires : Bruneau, Garrétt, Soulan et Pnscarel ; Morgan et Chevalier : ïWonnler. Bona. /"/ de Beyssac. L- euvielle, Blanchard. Boyau, Sainte-Marie et Apolline. Griffilhs, blessé, assistait au match en spectateur et Cooper le remplaçait dans l'équipé du. Racing ; par contre G arrêt t. rétabli, avait repris sa place parmi les Borde- Jais. Par l'absence et la présence de ces 'deux /"/ hommes, l'équilibre de la partie faillit être rompu au profit du Stade Bordelais qui domina nettement son adversaire pendant la première mi-temps, au moment où ie Racing jouait avec le vent qui- soufflait violemment. Le Stade utilisait assez brillamment ses sorties de mêlée, tandis qne.Cooper mettait de la lenteur à profiter des occasions où le ballon sortait vers lui.



Consulter la page sur Gallica

An experimental extraction based on genre classification : every sports article in the *matin* in 1912.

Generate corpora *on-demand*

Écritures du cirque

Le Petit Parisien du 1943-01-09, article n°69 (page n°4)

y «. >. « Importantes décisions àIaF.F, B. Sur le ring du Cirque d'Hiver Les Bayonnais ne seront pas a 1 a se M... Cm. r demain face aux avants staaistes j e France des poids mouche P«r k.- o. à ia 10 'reprise !

Notre dernier match de championnat rentre l'Avirea Rayonnai », s nn * dit I*«s- rmWtnatian aI André Verfer qui dirige le rugby an S lad * Français. remonte à 1031. N#us avions réngai le match nnl en face de ce

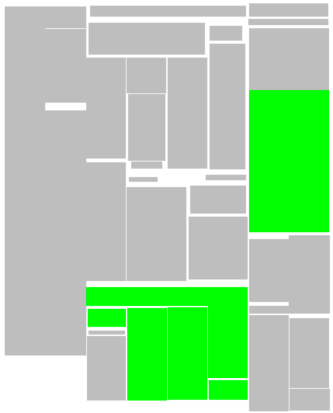
fameux Aviron qui Tenait de battre le Stade Toulcu * » i*. Alors jmtMisi ne pan réaliser dimanche un « ipleit dn intmr(« re ? Ft André Verger lions précisé qn * l'fqnipe dn Stade Français e»t m > mesure Ae « /"/ adapter an jeu allant et mobile des Rayonnais ; tt l'a prenre en battant le C » A. S. C*. par a * • Far aillent*. le * écn « * qni nota * Tiennent de la rét * basque no ne apprennent qaeCelhny, grand ntrqueni d'onipt jouera au centre aux roté » de Danger, le meilleur attaquant de l'henre. Zabalrtta jouer » arrière, Jl a bal en à la mêlée, Alvarez à l'ouverture, et neuf pressentons lS'II J « t » mainte » permutât ion » en coure de Mali il reste à savoir ai tee Bayounato ne aeroat pas dominés par lea avants d * Stade Français qnl,s ■ ans raison, prétend posséder la meilleure troisième ligne do club avec Blond, Henry et Fêsler. • Jl'id dehors de ee- mateb d'envergure, le » trois autres rencontres eomptpvnt également pour leo quarts dp fiai ale dn championnat de France « ont : Stade A. S. G : * Toar*. le Rniiran- I ',. S, Métro à Foltiers « t Biarriti- t'ognac à Bayonne. ■ M, L-

D'importante » dérisions ont été prises hier soir par le comité directeur i dé la F. F. B. En voici l'essentiel : Le 'défi de Besneux à McnozrA pour le titre de champion de la zone occupée, poids mi-lourds, est accepté. Suppression des champions de zone à partir du 15 février, les champions des deux zones devant se rencontrer dans un délmL de trois mois pour le titre de champion de France. Le boxeur Médina rst proclamé charrtpion de France poids mouche. La Fédération accordera des subventions aux olubs pour leurs professeurs à condition que cet professeurs soient diplômés de la F, N. E. Organisation, directe par lu Fédé ration de toute * les compétitions amateurs. Création d'iutlé caisse autonome pour aider- les Amateurs. 10 % des bons pour chaussures, gants et équipements seront remis aux prisonniers par l'intermédiaire des clubs.

CM OS. i ROGER Mit H FI.OT B era « ppasé au champion de Belgique de » poids moyen » Al Baker le 30 janvier i Bruxelles.

r—-•-- /"/-r l ! Hier soir, au Cirque d'fJivrr. le\Xord- Africain Ben Omar, cueUii d froid par- un « wihfr du gauche de Sli- ! « final, qui l'Atteignit au visage, alla au. tapis pour neuf secondes. Ordre à son courage et aussi, il fout j bien le dire, à an science du ring, Brn I Omar put terminer le round. Puis, j avant récupéré, il combla son handicap j et, au moment oit il allait rejoindre au décompte des points le Xord- Africain fut de nouveau envoyé au sol pour huit secondes au huitième round, Cette fois, c'en était fini. Jl subtil encore sept knock- downs et fut mis hors de combat néant la fin du dixième et dernier round. pour sa première défaite, Ben Omar fut battu par knockttrut. Mais quelle ardente bataille ! Xouvelle victoire du classique Jean Mougin et pi té usé exhibition de Momber qui battit Dunand aine joints.

! UEMAIX DIMANCHE, autour de 1 » pj*ee d'eau des Suisses, k Versailles. ! seront dispute * le challenge de efosaicountry Bené Poîncetet et. la eu ope du ; Xombre. Ton » les grand * elnb » de la, 'région parisienne sont engages. i



Consulter la page sur Gallica

Proximité au thème

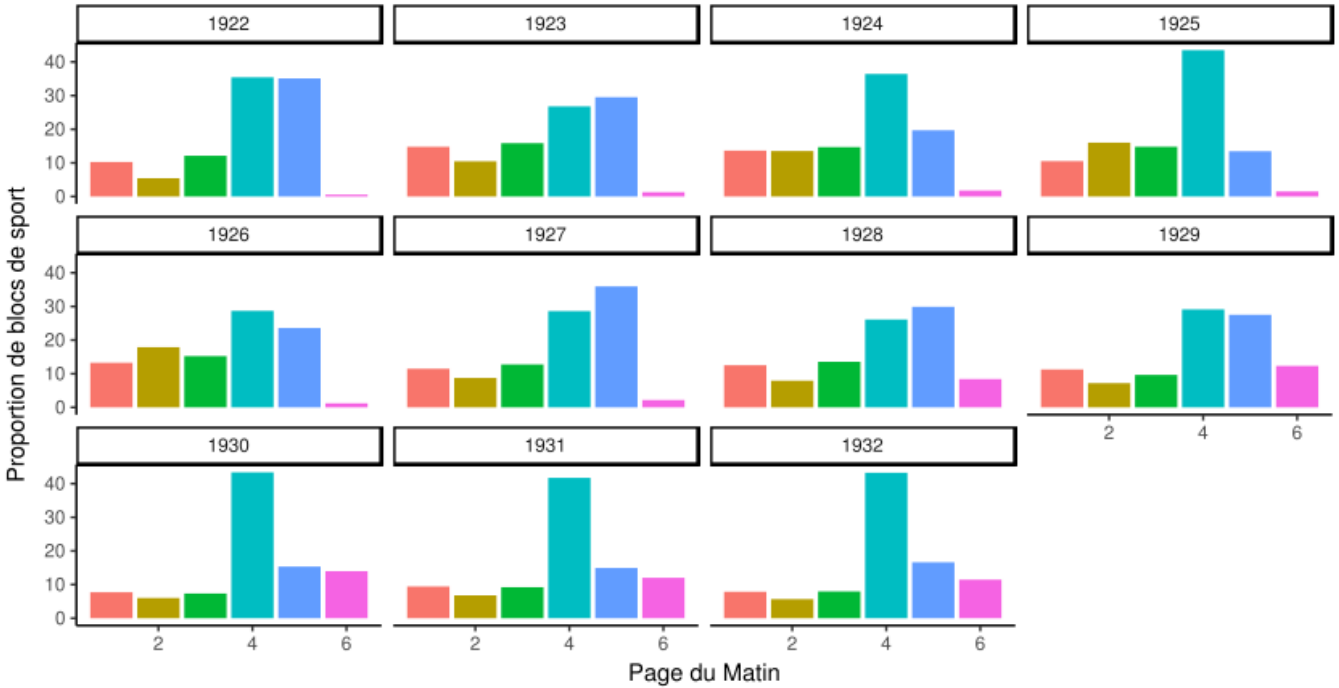
Indice de proximité au thème de **0.68** à partir des mots « *Cirque*, *Cirque* »

Classification

Sport (47.75 %), Publicité_affiche (10.65 %), Bourse (5.95 %), Économie (5.59 %)

Classification can also serve to document a pre-existing corpora (here the circus in occupied France during WWII) and ease the reading process.

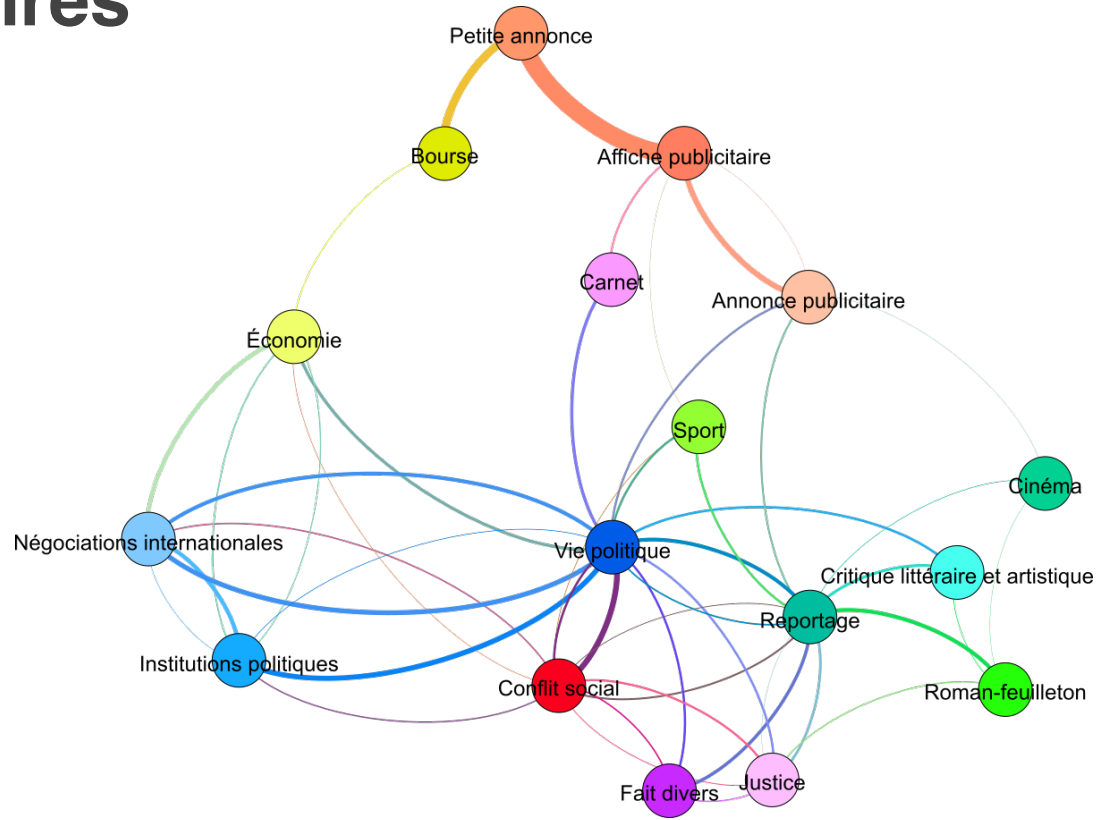
Place the genres



During the 1920s, the place of the sports article within the newspaper is more and more specific: on page 4.

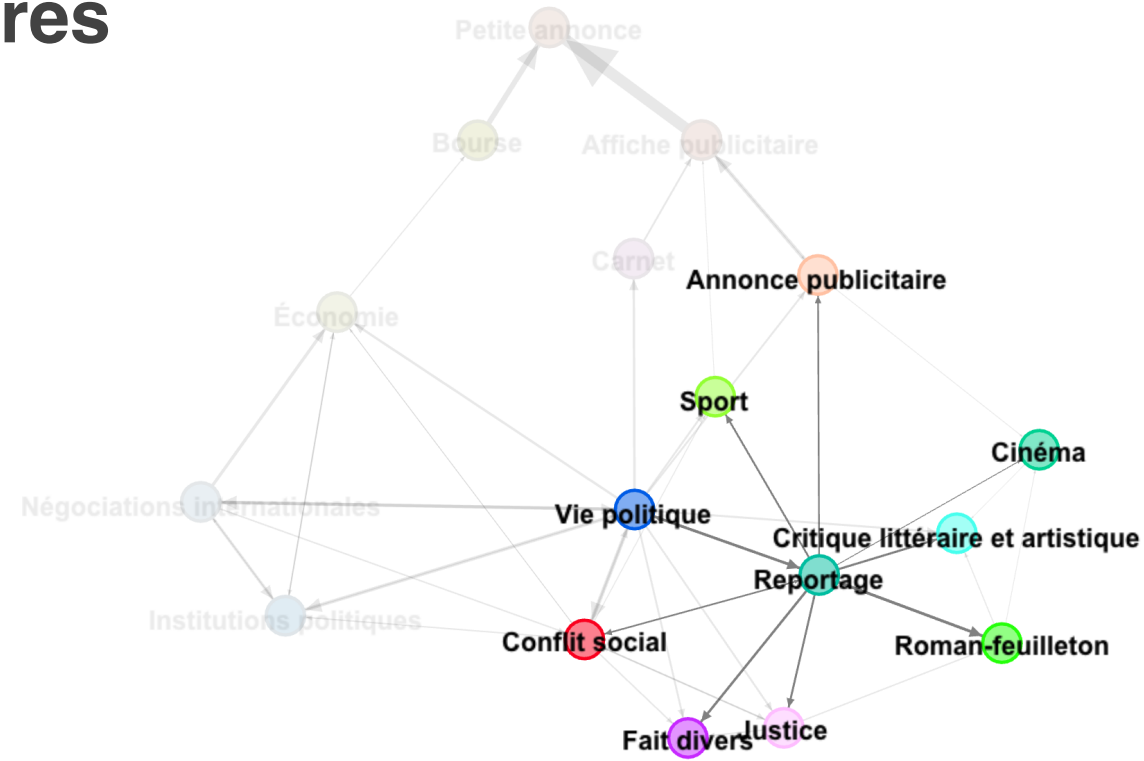
Place the genres

As each text block can belong to several genres (defined by probabilities), we can map all the frequent relations between genres into a network, with each link indicating the frequency of cross-classification…



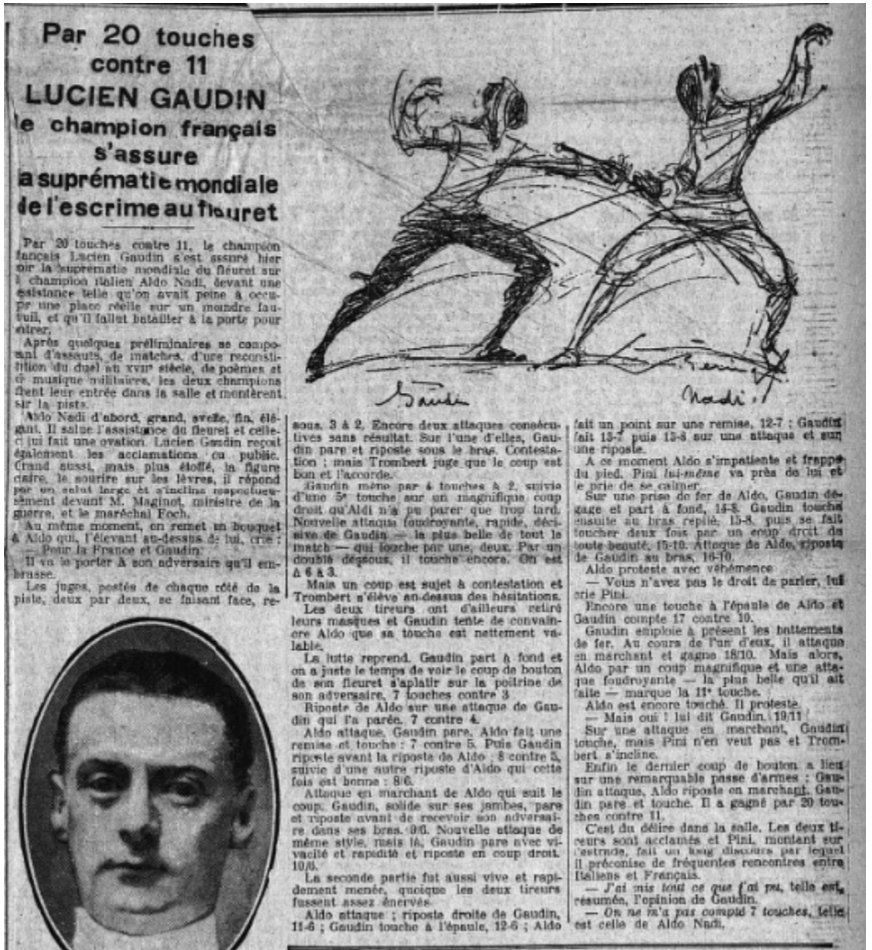
Place the genres

This visualization highlights “crossroads genre”, such as *reportage* which connect political, social and cultural sections of the newspaper…



Place the genres

A case of intertextuality: an account of fencing written like a serial novel of Alexandre Dumas (31 January 1922 : serial novel 30%, sport 35%)



Par 20 touches contre 11

LUCIEN GAUDIN

le champion français s'assure a suprématie mondiale de l'escrime au fleuret

Par 20 touches contre 11, le champion français Lucien Gaudin s'est assuré hier soir la suprématie mondiale du fleuret sur le champion italien Aldo Nadi, devant une assistance telle qu'on avait peine à occuper une place réelle sur un moindre fauteuil, et qu'il fallut batailler à la porte pour entrer.

Après quelques préliminaires se composant d'assauts, de matchs, d'une reconstitution du duel au XVIIIe siècle, de poèmes et de musique militaires, les deux champions font leur entrée dans la salle et montèrent sur la piste.

Aldo Nadi d'abord, grand, svelte, fin, élégant. Il salue l'assistance du fleuret et celle-ci lui fait une ovation. Lucien Gaudin reçoit également les acclamations du public. Grand aussi, mais plus étoffé, la figure ciselée, le sourire sur les lèvres, il répond par un salut large et s'incline respectueusement devant M. Maginot, ministre de la guerre, et le maréchal Foch.

Au même moment, on remet un bouquet à Aldo qui, l'élevant au-dessus de lui, crie :

— Pour la France et Gaudin!

Il va le porter à son adversaire qu'il embrasse.

Les juges, postés de chaque côté de la piste, deux par deux, se faisant face, re-

sous. 3 à 2. Encore deux attaques consécutives sans résultat. Sur l'une d'elles, Gaudin pare et riposte sous le bras. Contestation ; mais Trombert juge que le coup est bon et l'accorde.

Gaudin mène par 4 touches à 2, suivie d'une 5e touche sur un magnifique coup droit qu'Aldi n'a pu parer que trop tard. Nouvelle attaque foudroyante, rapide, décisive de Gaudin — la plus belle de tout le match — qui touche par une, deux. Par un doublé dessous, il touche encore. On est à 6 à 3.

Mais un coup est sujet à contestation et Trombert n'élève an-dessus des hésitations.

Les deux tireurs ont d'ailleurs retiré leurs masques et Gaudin tente de convaincre Aldo que sa touche est nettement valable.

La lutte reprend. Gaudin part à fond et on a juste le temps de voir le coup de bouton de son fleuret s'aplatir sur la poitrine de son adversaire. 7 touches contre 3.

Riposte de Aldo sur une attaque de Gaudin qui l'a parée. 7 contre 4.

Aldo attaque, Gaudin pare, Aldo fait une remise et touche : 7 contre 5. Puis Gaudin riposte avant la riposte de Aldo : 8 contre 5, suivie d'une autre riposte d'Aldo qui cette fois est bonne : 8/6.

Attaque en marchant de Aldo qui suit le coup. Gaudin, solide sur ses jambes, pare et riposte avant de recevoir son adversaire dans ses bras. 9/6. Nouvelle attaque de même style, mais là, Gaudin pare avec vivacité et rapidité et riposte en coup droit. 10/6.

La seconde partie fut aussi vive et rapidement menée, quoique les deux tireurs fussent assez énervés.

Aldo attaque ; riposte droite de Gaudin, 11-6 ; Gaudin touche à l'épaule, 12-6 ; Aldo fait un point sur une remise, 12-7 ; Gaudin fait 13-7 puis 13-8 sur une attaque et sur une riposte.

A ce moment Aldo s'impatiente et frappe du pied. Pini *lui-même* va près de lui et le prie de se calmer.

Sur une prise de fer de Aldo, Gaudin dégage et part à fond, 14-8. Gaudin touche ensuite au bras replié. 15-8, puis se fait toucher deux fois par un coup droit de toute beauté, 15-10. Attaque de Aldo, riposte de Gaudin au bras, 16-10.

Aldo proteste avec véhémence

— Vous n'avez pas le droit de parler, lui crie Pini.

Encore une touche à l'épaule de Aldo et Gaudin compte 17 contre 10.

Gaudin emploie à présent les battements de fer. Au cours de l'un d'eux, il attaque en marchant et gagne 18/10. Mais alors, Aldo par un coup magnifique et une attaque foudroyante — la plus belle qu'il ait faite — marque la 11e touche.

Aldo est encore touché. Il proteste

— Mais oui ! lui dit Gaudin. 19/11

Sur une attaque en marchant, Gaudin touche, mais Pini n'en veut pas et Trombert s'incline.

Enfin le dernier coup de bouton a lieu sur une remarquable passe d'armes : Gaudin attaque, Aldo riposte en marchant. Gaudin pare et touche. Il a gagné par 20 touches contre 11.

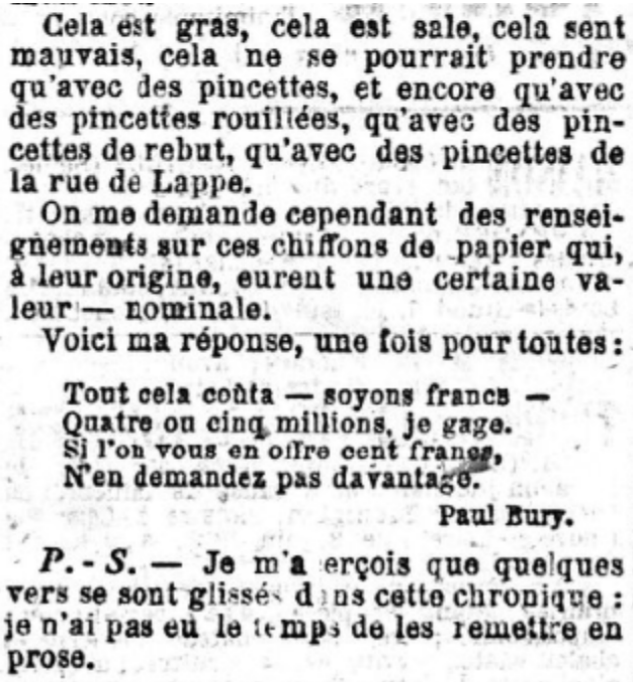
C'est du délire dans la salle. Les deux tireurs sont acclamés et Pini, montant sur l'estrade, fait un long discours par lequel il préconise de fréquentes rencontres entre Italiens et Français.

— *J'ai mis tout ce que j'ai pu*, telle est, résumée, l'opinion de Gaudin.

— *On ne m'a pas compté* 7 touches, telle est celle de Aldo Nadi.

Place the genres

…or a stock exchange section partly written using verses in *Le Figaro*



Cela est gras, cela est sale, cela sent mauvais, cela ne se pourrait prendre qu'avec des pincettes, et encore qu'avec des pincettes rouillées, qu'avec des pincettes de rebut, qu'avec des pincettes de la rue de Lappe.

On me demande cependant des renseignements sur ces chiffons de papier qui, à leur origine, eurent une certaine valeur — nominale.

Voici ma réponse, une fois pour toutes :

Tout cela coûta — soyons francs —
Quatre ou cinq millions, je gage.
Si l'on vous en offre cent francs,
N'en demandez pas davantage.

Paul Bury.

P. - S. — Je m'aperçois que quelques vers se sont glissés dans cette chronique : je n'ai pas eu le temps de les remettre en prose.

An archeology of genres

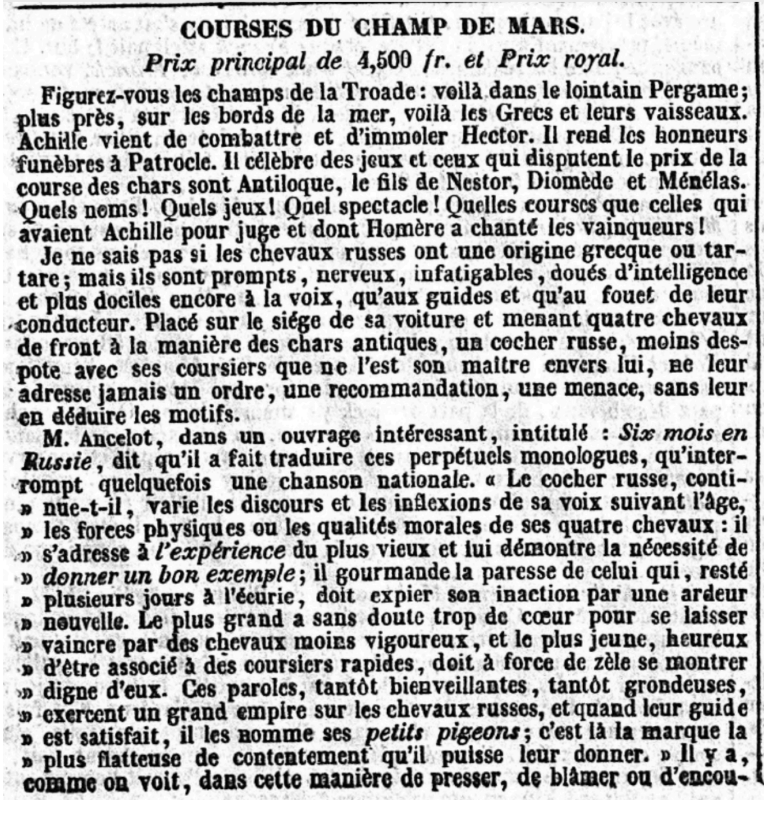
Each model has an expiratory date: we are going to use them on a 20-30 years basis.

Yet, it would still be instructive to use them in an “anachronistic” way. The issue is no longer to classify texts but rather to find the antecedents of a later established genre.

We have made a preliminary test on *Le Journal des Débats* in 1836, at a time where an established sport section did not exist.

An archeology of genres

…and it turns out there were significant, although quite occasional sport articles, such as this lengthy feuilleton (septembre 9th, 1836) on the horse races on the *Champs de Mars*



COURSES DU CHAMP DE MARS.

Prix principal de 4,500 fr. et Prix royal.

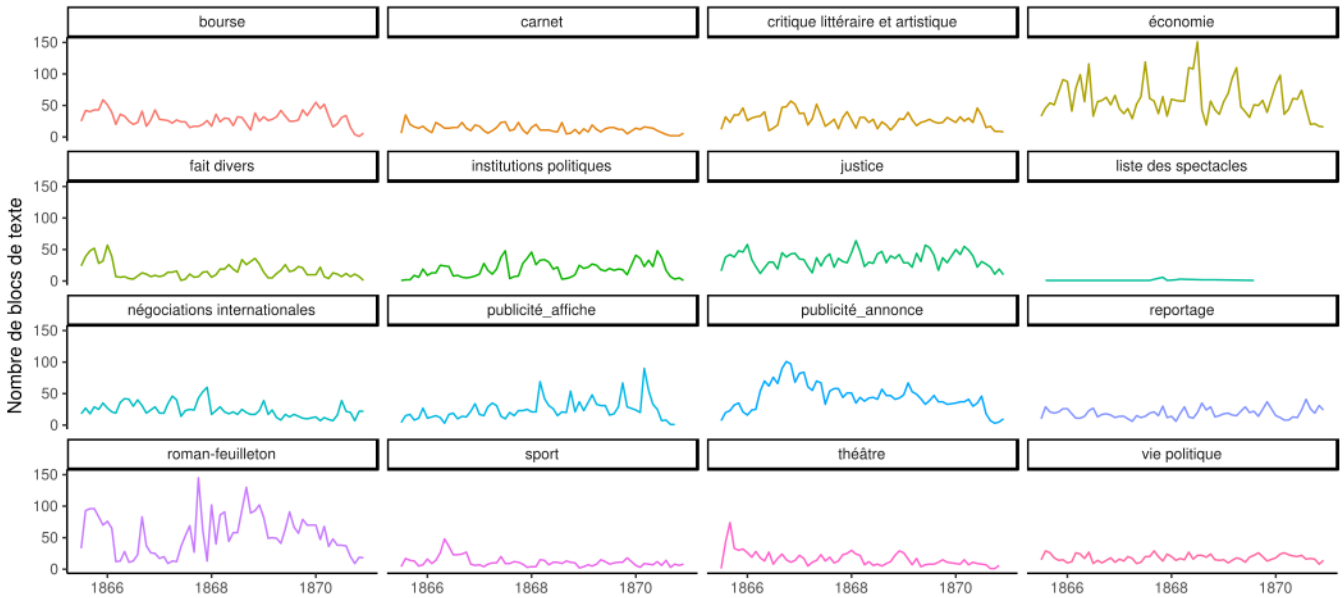
Figurez-vous les champs de la Troade : voilà dans le lointain Pergame; plus près, sur les bords de la mer, voilà les Grecs et leurs vaisseaux. Achille vient de combattre et d'immoler Hector. Il rend les honneurs funèbres à Patrocle. Il célèbre des jeux et ceux qui disputent le prix de la course des chars sont Antiloque, le fils de Nestor, Diomède et Ménélas. Quels noms! Quels jeux! Quel spectacle! Quelles courses que celles qui avaient Achille pour juge et dont Homère a chanté les vainqueurs!

Je ne sais pas si les chevaux russes ont une origine grecque ou tartare; mais ils sont prompts, nerveux, infatigables, doués d'intelligence et plus dociles encore à la voix, qu'aux guides et qu'au fouet de leur conducteur. Placé sur le siége de sa voiture et menant quatre chevaux de front à la manière des chars antiques, un cocher russe, moins despote avec ses coursiers que ne l'est son maître envers lui, ne leur adresse jamais un ordre, une recommandation, une menace, sans leur en déduire les motifs.

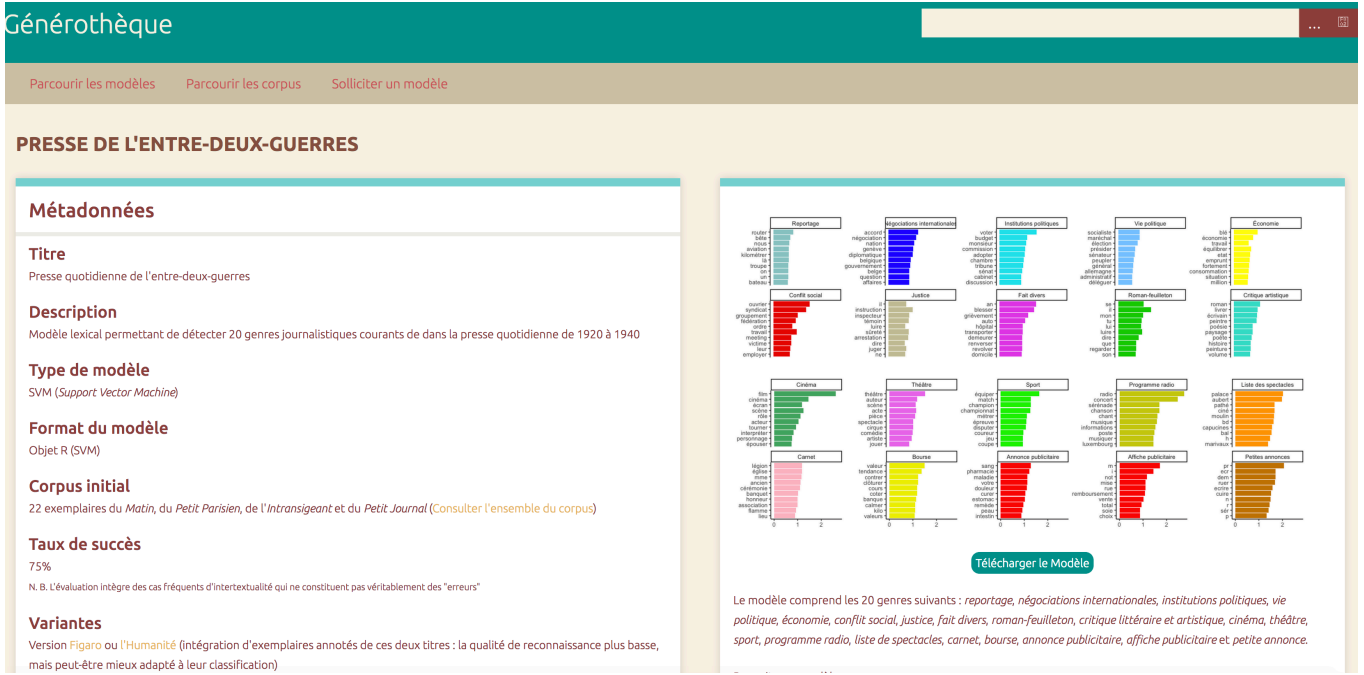
M. Ancelot, dans un ouvrage intéressant, intitulé : *Six mois en Russie*, dit qu'il a fait traduire ces perpétuels monologues, qu'interrompt quelquefois une chanson nationale. « Le cocher russe, continue-t-il, varie les discours et les inflexions de sa voix suivant l'âge, » les forces physiques ou les qualités morales de ses quatre chevaux : il » s'adresse à *l'expérience* du plus vieux et lui démontre la nécessité de » *donner un bon exemple*; il gourmande la paresse de celui qui, resté » plusieurs jours à l'écurie, doit expier son inaction par une ardeur » nouvelle. Le plus grand a sans doute trop de cœur pour se laisser » vaincre par des chevaux moins vigoureux, et le plus jeune, heureux » d'être associé à des coursiers rapides, doit à force de zèle se montrer » digne d'eux. Ces paroles, tantôt bienveillantes, tantôt grondeuses, » exercent un grand empire sur les chevaux russes, et quand leur guide » est satisfait, il les nomme ses *petits pigeons*; c'est là la marque la » plus flatteuse de contentement qu'il puisse leur donner. » Il y a, comme on voit, dans cette manière de presser, de blâmer ou d'encou-

An archeology of genres

This anachronistic outlook can be deployed on a wider scales, such as the evolution of the genres of *La Liberté* from 1865 to 1870 using the model made for the 1920s and 1930s.



Toward a library of models



The « Generothèque » project: a library of classification models

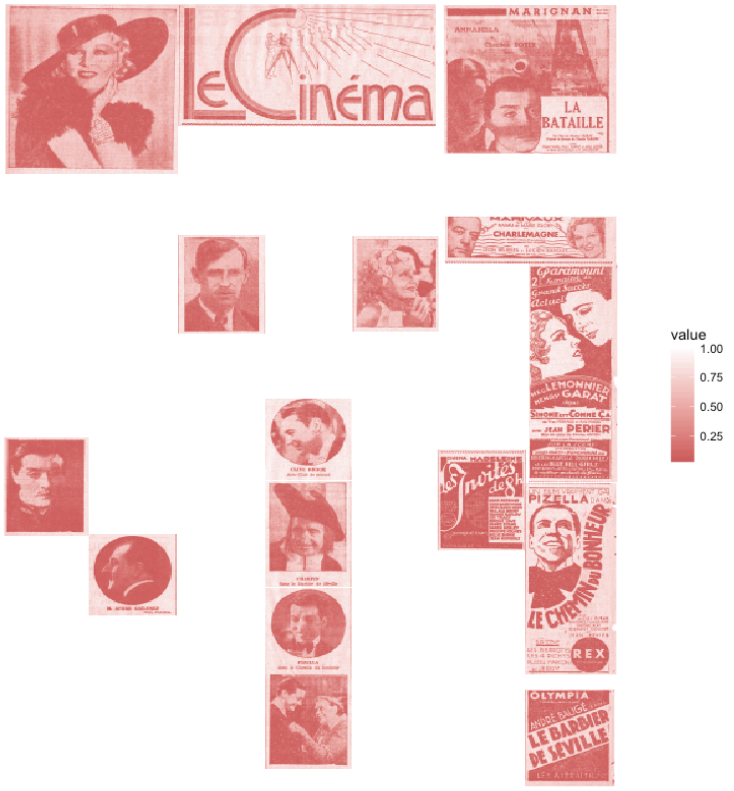
2. Toward a poetic analysis of news images

Tracking aesthetic mutations through *deep learning*.

The news image: the spinal column of the XXth century newspaper

Starting in the 1900s, images became a very strong baseline of news structuration. With the digitized XML files from the BNF it is possible to extract all the illustrations, through their coordinates, and to reconstruct this polyphonic interaction between text and images.

Projection des images de la rubrique cinéma du Matin



From text to image

The position of the image *within* the text allow to easily retrieve the thematic of the image through the automated classification of the surrounding text.

Le Petit Journal du 1934-01-01, article n°9(page n°1)

DU NOUVEAU EN RUGBY

Les équipes nationales d'Angleterre et d'Australie ont donné hier, au stade Perahing, une démonstration fort agréable du rugby à treize joueurs, La partie qu'ils jouèrent-r la première de cette sorte disputée en France fort plaisante, fut mouvementée à souhait, facile à suivre et intéressa vivement les spectateurs. Notre cliché représente un joueur australien qui, s r étant emparé du ballon, va se trouver aux prises avec un Anglais. Les équipes nationales d'Angleterre et d'Australie ont donné hier, au stade Perahing, une démonstration fort agréable du rugby à treize joueurs, La partie qu'ils jouèrent-r la première de cette sorte disputée en France fort plaisante, fut mouvementée à souhait, facile à suivre et intéressa vivement les spectateurs. Notre cliché représente un joueur australien qui, s r étant emparé du ballon, va se trouver aux prises avec un Anglais.

Consulter la page sur Gallica

Classification générale : sport(99%)
Sous-classification : boxe(68%), tennis(29%), football(3%)

From text to image

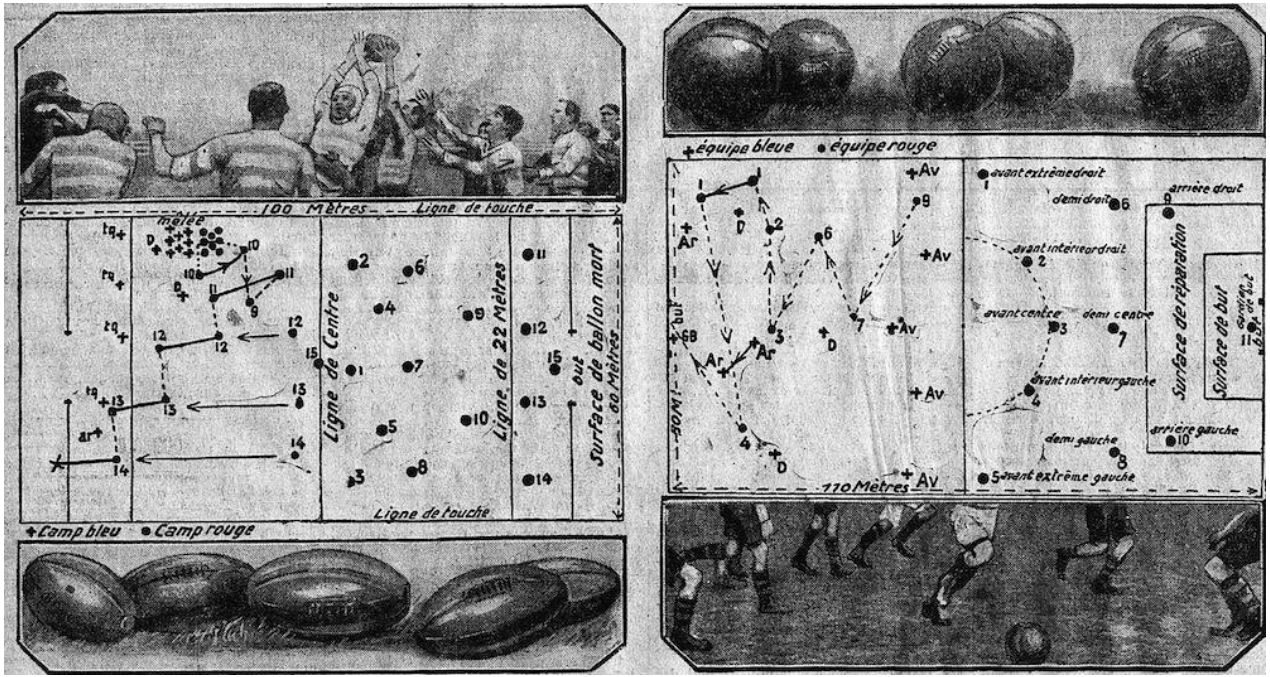
Sport (99%)

Sport (94%)

Sport (34%),
Carnet (30%)

From text to image

This classification by proxy ease significantly the consultation and identification of relevant images — such as this 1912 *dataviz* of a football match.



The imperfect magic of DL models

Image classification has made significant progress during the last few years, through the generalized use of convoluted neural network.

Nevertheless, most comprehensive models (such as ImageNet) remain focused on contemporary classification

=> Utiliser les récurrences de classification plutôt que d'accepter les classifications telles quelles. Si toutes les cartes sont classées comme des mots croisées cela peut être un critère valable.

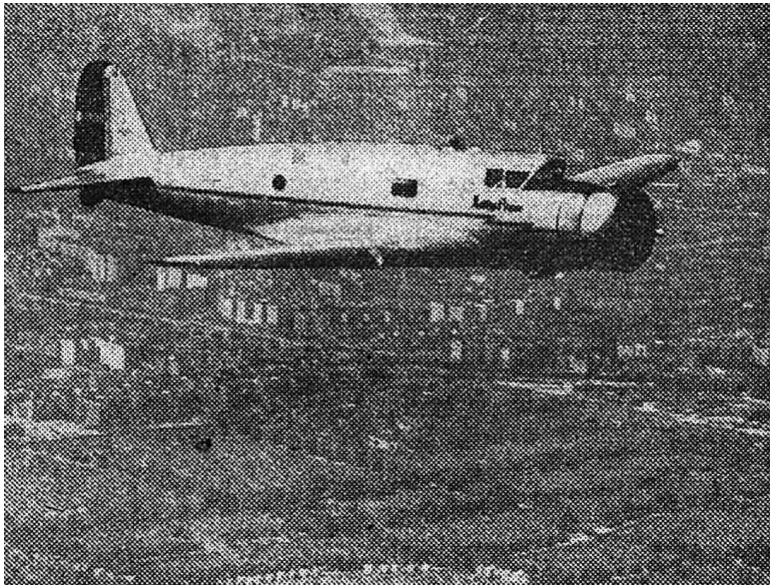
The imperfect magic of DL models

Joueur de baseball (0.18), *Couverture de livre* (0.12), *Télévision* (0.09), *Ballon de foot* (0.08)

Uniforme militaire (0.53), *Gilet pare-balle* (0.20), *Télévision* (0.05)

Couverture de livre (0.76), *Joueur de baseball* (0.13)

The imperfect magic of DL models



Space shuttle

(steampunk for the win)

Crosswords (!)

The (temporary) solution: using more abstract vectors

This approach has been experimented by the Netherland project SIAMESE: identifying cluster of similar images, through the next to last vector generated by the neural network.

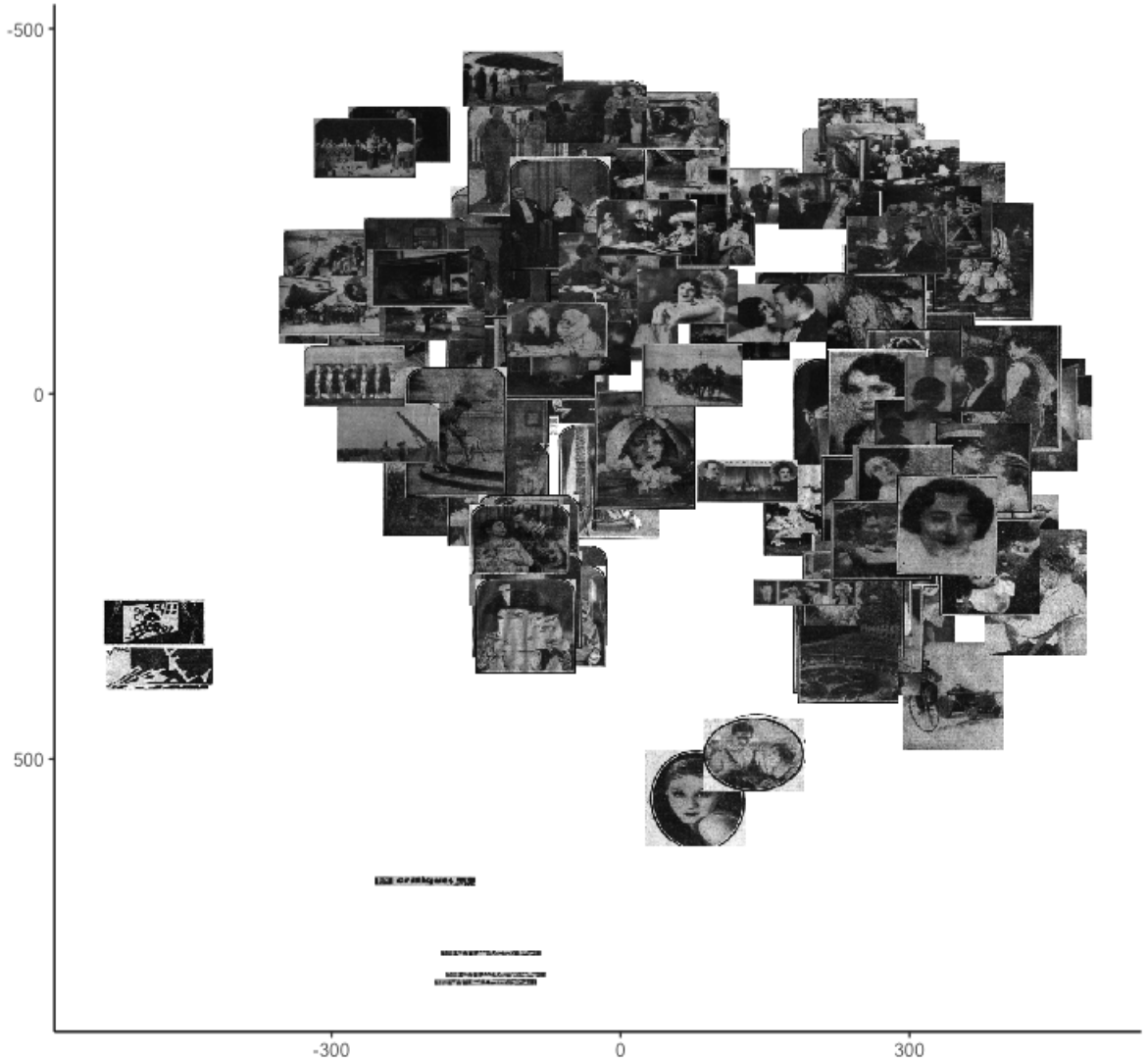
Introduction | Live demo | Instructions | Examples

SIAMESE performs especially well in grouping clearly identifiable objects in advertisements. On the one hand, this is driven by the high number of advertisements for fashion (Fig. 1) and automobiles (Fig. 2). On the other, it also reveals that these advertisements for these products featured a consistent visual trend.

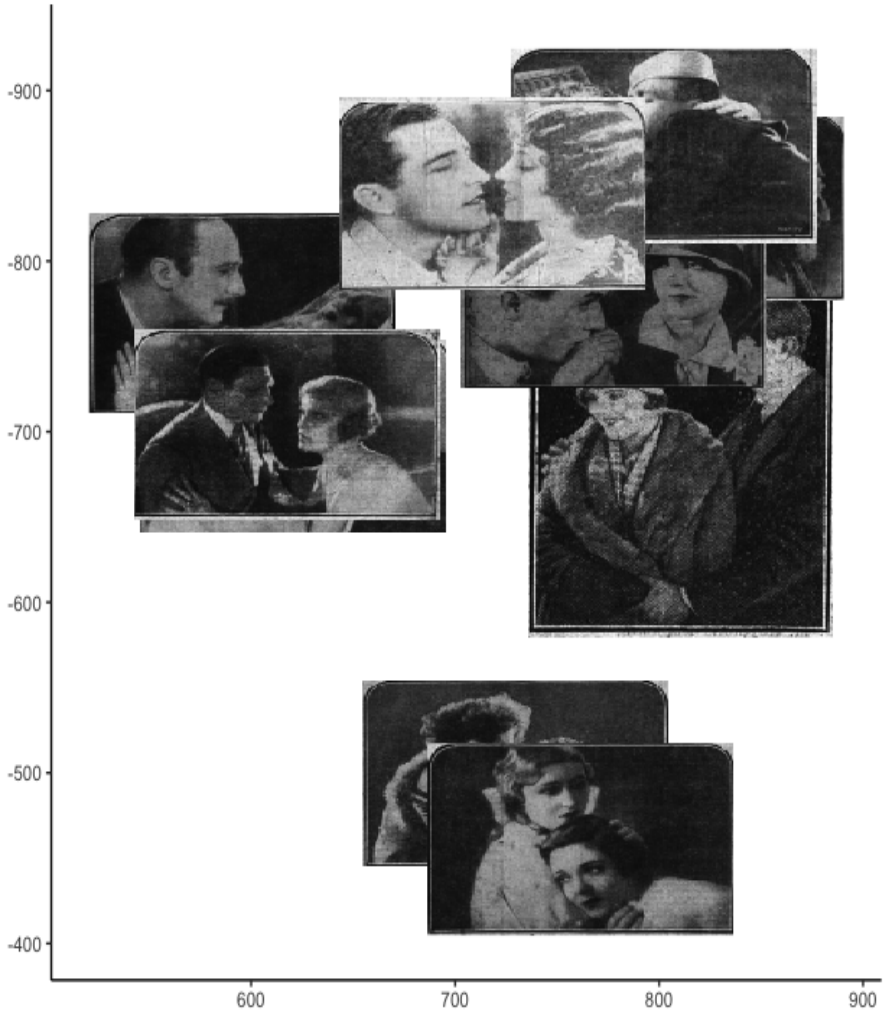
Figure 1. SIAMESE timeline view for fashion advertisements

Figure 2. SIAMESE timeline view for automobile advertisements

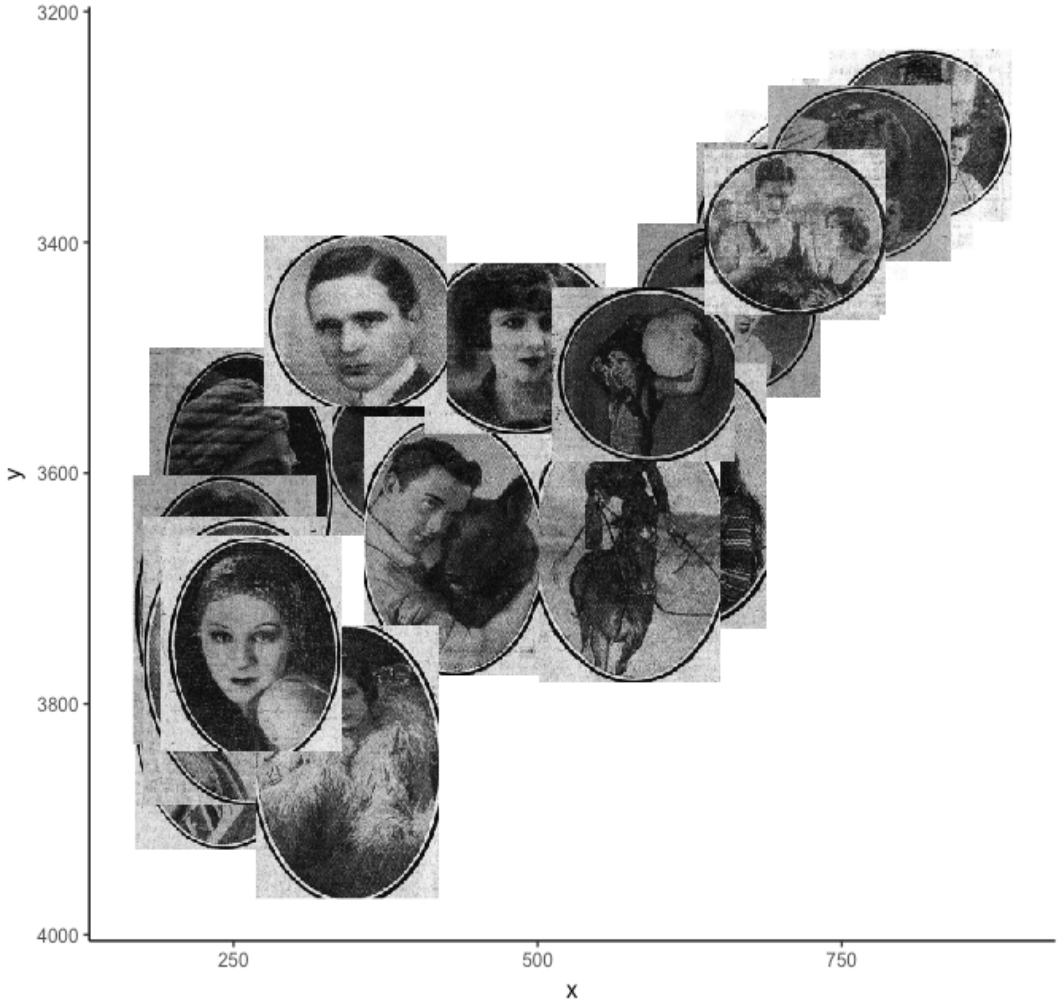
A random sample of all the images in the movie section of *Le Matin* from 1927 to 1929. Clustered images are likely similar.

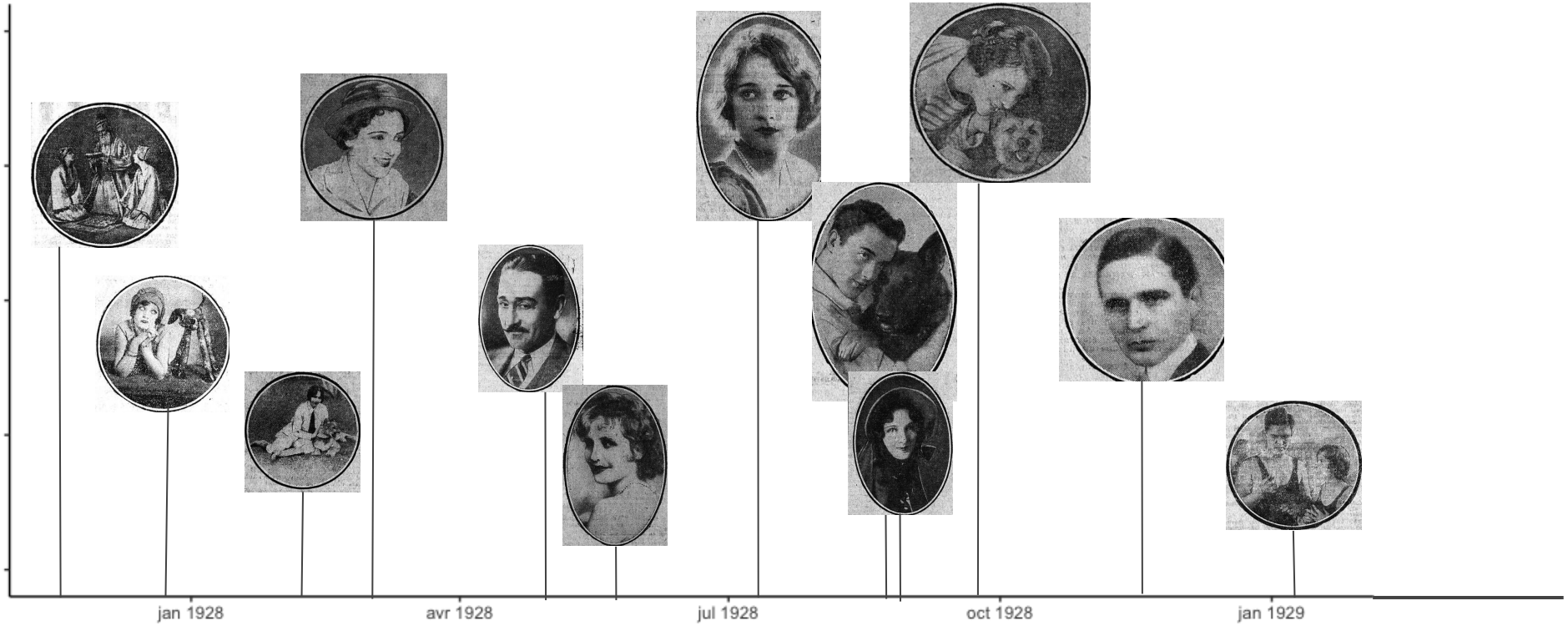


This “similarity“ can be due to the topic represented: here we have a cluster of couples (that are mostly represented through Hollywood stereotype).



The neural network also identifies recurrent form and structure of images, such as the use of an oval medallion.





By crossing deep learning results with metadata we can recover long-time shift in the way images are made, such as the sudden disappearance of the oval medallion

3. Reconstruction of editorial structure

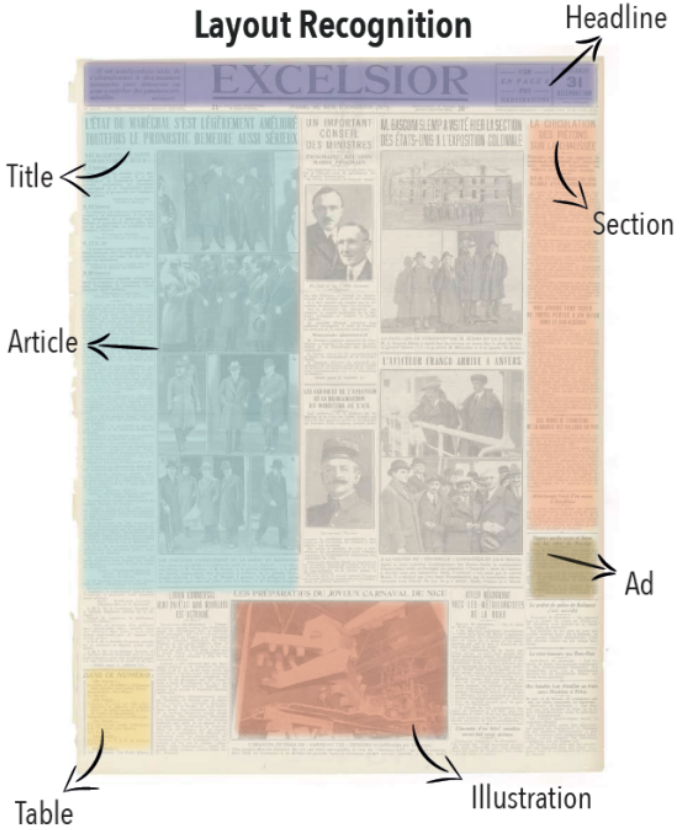
How to sight-read the polyphony of news

Breakthroughs in layout recognition

The Europeana Newspaper (2012-2015) has digitized 2 millions pages using Optical Layout Recognition (OLR), that allows to retrieve article and news components using semi-automatic tools.

The French National Library applies OLR for its current digitization program.

The Newseye projects aims to create advanced automatic tools to generate OLR.

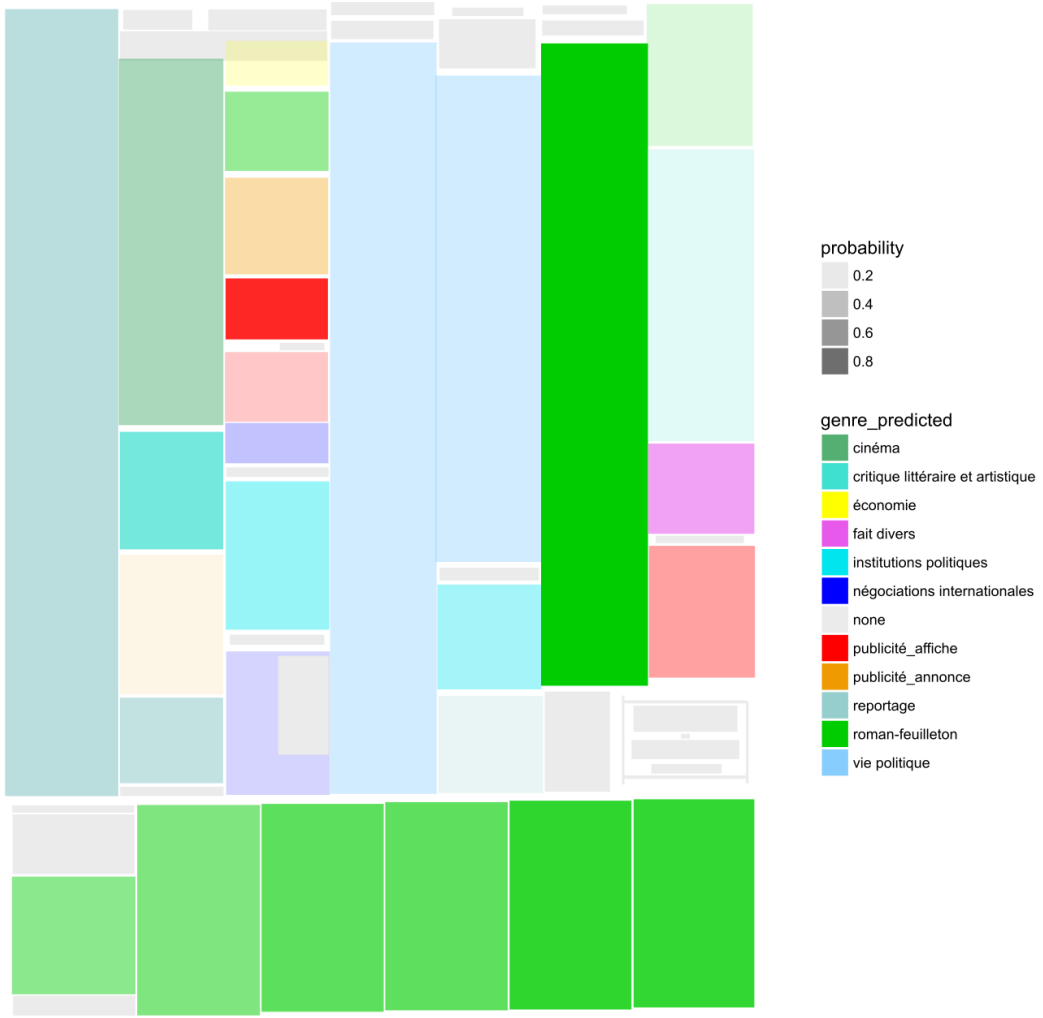


Text as image

A project still experimental: training a model to identify editorial structure, thanks to the pioneering work of DH Segment. The initial tests on the feuilleton are conclusive. Advertisings, images, headers, should come next…

Text as image

This will be especially valuable to further classify literary productions in the newspaper, serial novel being almost by definition in the *feuilleton* while other more fugitive forms (short story, verses…) may be placed in the higher part.



Parse the metadata *in* the news

A wide array of contextual information are actually present *in* the newspaper text, such as the date and place of a dispatch, the name and status of the writer, the communication process, and so forth…

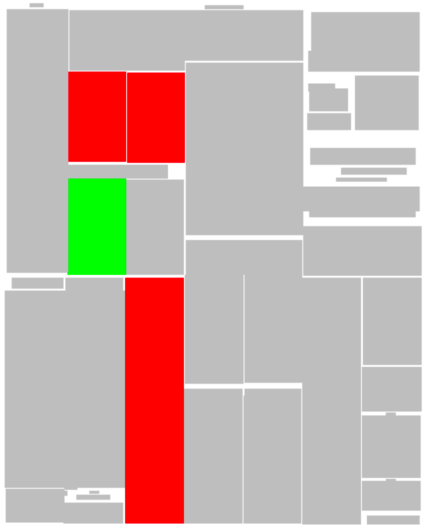
L'Intransigeant du 1935-09-06, bloc de texte n°p3b10(page n°3)

Londres, 5 septembre (de notre corr. part., par téléphone).

— Les événements qui se sont déroulés hier à Genève ont consterné beaucoup d'Anglais, ils n'en ont surpris aucun. Le discours de M. Eden est l' expression même de l'idéalisme de nos amis d'outre- Manche, la traduction fidèle, sur le plan diplomatique et international des véritables sentiments de la démocratie britannique. Certes ; on considère Ici la situation comme fort grave. M. Stanley Baldwin,, premier ministre, a quitté Aix- les- Bains ; M. Mac, Donald a quitté les rlyages de l' Ecosse ; M. Nevllle Chamberlain, chancelier de l' Echiquier, a quitté la Touraine, le jardin de la France, pour revenir à Londres, où sir Samuel Hoare veille.

Il est à remarquer — et certains journaux anglais en font la constatation.— que le discours prononcé hier par le baron Aloisi,, devant le Conseil suprême de Genève, est un peu moihs fort quant aux termes et quant au fond que les rapports et livres de lady Simon, lord Buxton, lord Polwarth, enquêteurs bénévoles en Ethiopie, de M. Bussel, ancien ministre britannique à Addis- Abeba, et de sir Robert Corijndon, ancien gouverneur du Kenya, rapports et livres consacrés aux mœurs et è la civilisation des peuples bienheureux qui vivent sons l' allégeance du Négus.

On pense ici que l' heure n' est pas encore venue de jeter le manche après la cognée. Et cela pour les raisons suivantes :



Consulter la page sur Gallica

Lieu d'expédition : Londres (51.50°, -0.12°)

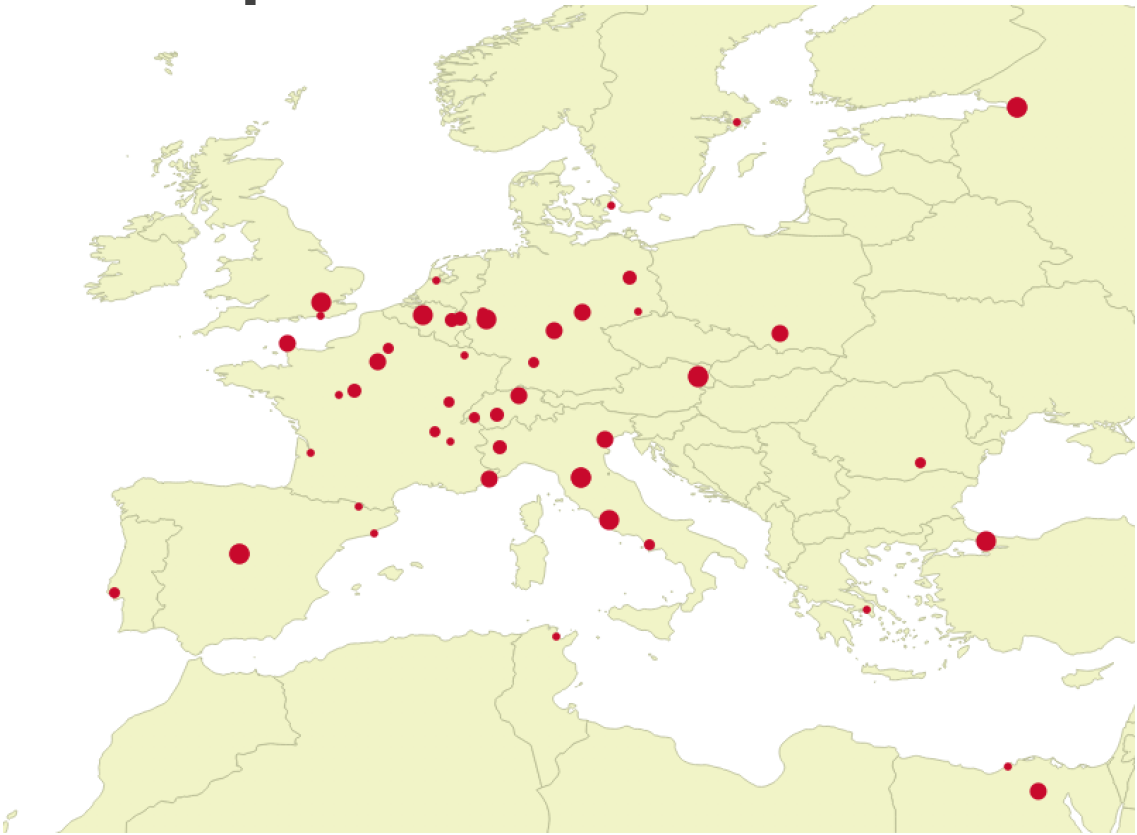
Date d'expédition : 5 septembre 1935 (1 jour avant publication)

Auteur : correspondant particulier du *Matin* (Gérard Boutelleau ?)

Moyen d'expédition : téléphone

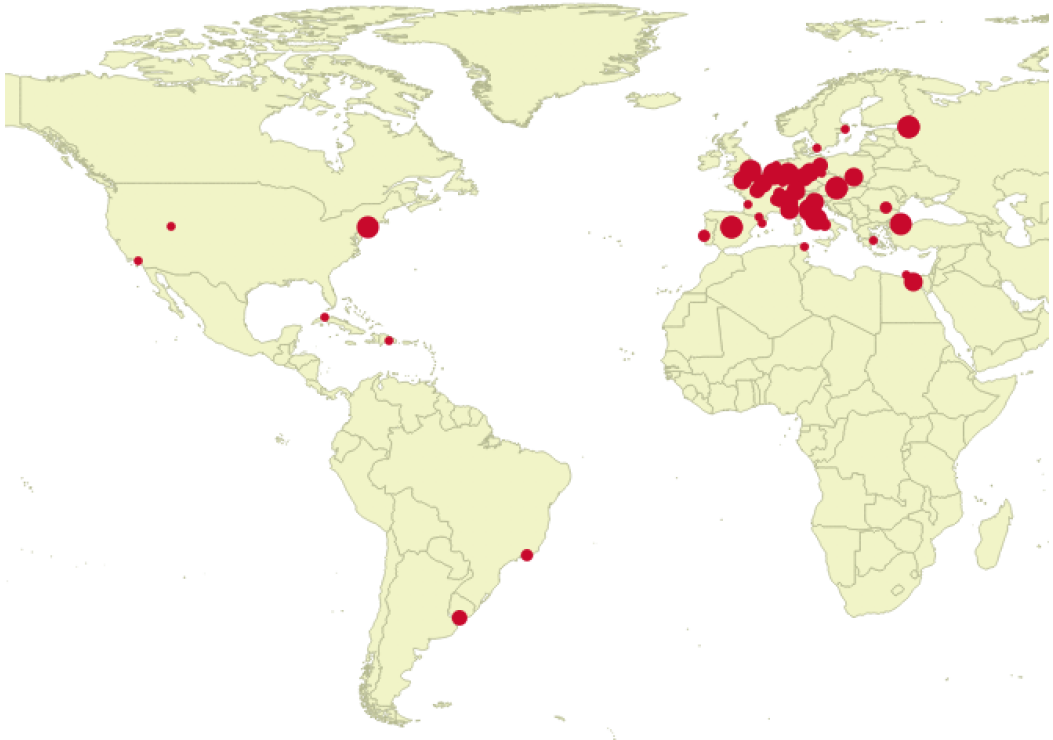
Parse the metadata: places

We have lead an initial case study on the emergence of French reporters in *La Liberté* (1865-1870). All the expedition places have been located and showed that the newspaper already managed a very large network across Europe and the Americas.



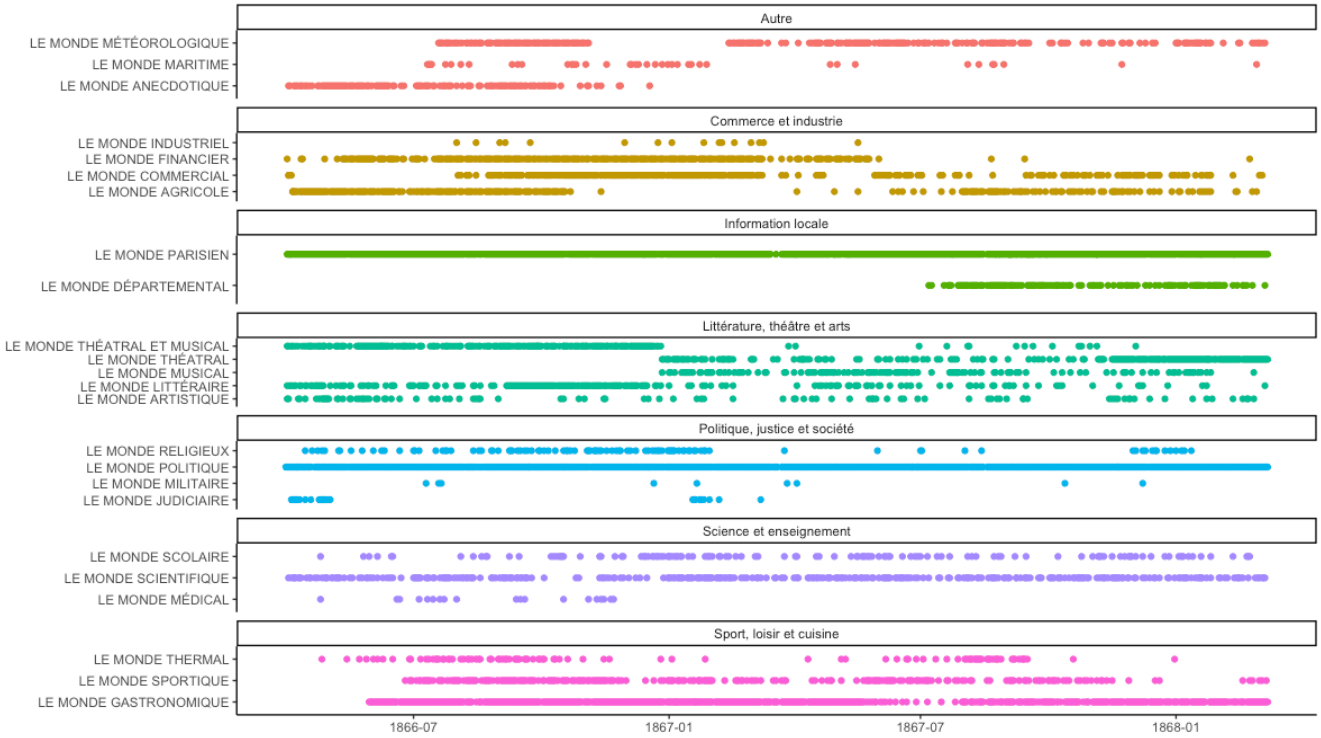
Parse the metadata: places

We have lead an initial case study on the emergence of French reporters in *La Liberté* (1865-1870). All the expedition places have been located and showed that the newspaper already managed a very large network across Europe and the Americas.



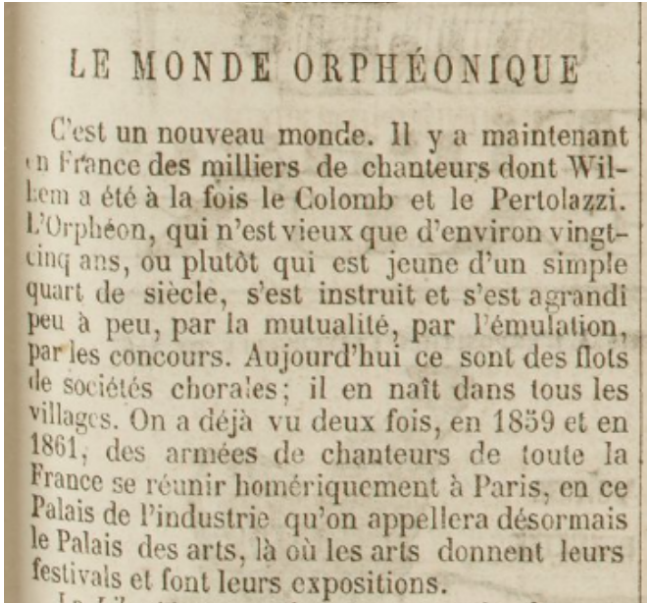
Parse the metadata: titles

Titles can contain valuable information on the way the segmentations of newspaper texts initiated by the editors. For instance, *La Liberté* initiated extensive editorial experiments from 1866 to 1868 can gave birth to a lot of *first* in the history of French journalism (first sport section, first food column…)



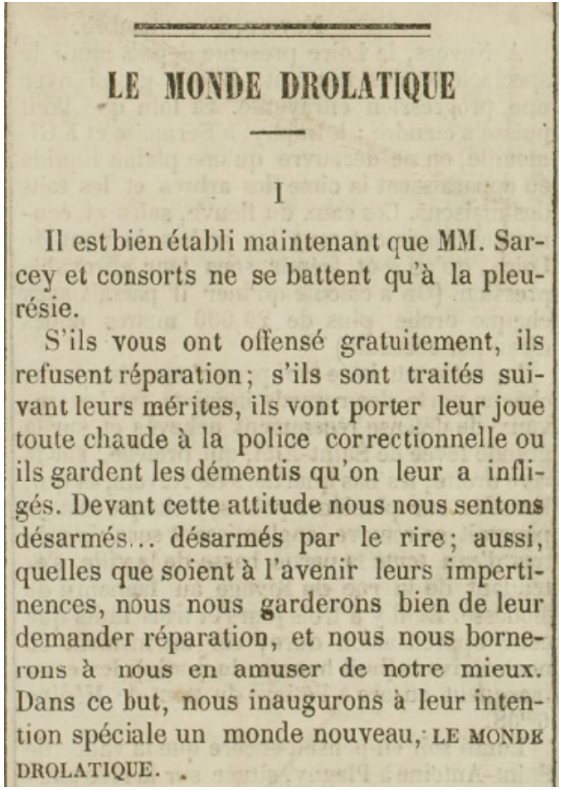
Parse the metadata: titles

We can also zoom to *failed* attempts with specialized section that could have become a regular part of French journalism but didn't (altogether, not for the worst reasons)



LE MONDE ORPHÉONIQUE

C'est un nouveau monde. Il y a maintenant en France des milliers de chanteurs dont Wilhem a été à la fois le Colomb et le Pertolazzi. L'Orphéon, qui n'est vieux que d'environ vingt-cinq ans, ou plutôt qui est jeune d'un simple quart de siècle, s'est instruit et s'est agrandi peu à peu, par la mutualité, par l'émulation, par les concours. Aujourd'hui ce sont des flots de sociétés chorales; il en naît dans tous les villages. On a déjà vu deux fois, en 1859 et en 1861, des armées de chanteurs de toute la France se réunir homériquement à Paris, en ce Palais de l'industrie qu'on appellera désormais le Palais des arts, là où les arts donnent leurs festivals et font leurs expositions.



LE MONDE DROLATIQUE

I

Il est bien établi maintenant que MM. Sarcey et consorts ne se battent qu'à la pleurésie.

S'ils vous ont offensé gratuitement, ils refusent réparation; s'ils sont traités suivant leurs mérites, ils vont porter leur joue toute chaude à la police correctionnelle ou ils gardent les démentis qu'on leur a infligés. Devant cette attitude nous nous sentons désarmés... désarmés par le rire; aussi, quelles que soient à l'avenir leurs impertinences, nous nous garderons bien de leur demander réparation, et nous nous bornerons à nous en amuser de notre mieux. Dans ce but, nous inaugurons à leur intention spéciale un monde nouveau, LE MONDE DROLATIQUE.

Parse the metadata: signatures

Name entity recognition has been greatly boosted by the development of an efficient tool for the French language, *Entity Fishing*. The specific issues raised by the OCR may be partly alleviated thanks to the pioneering work of *Newseye*.

Annotations | Response

Mlle KATE DE NAGY et M. CHARTEA VANEE tiennent simplement de grandi rôles; iLs sont sobres de gestes mal» puissants d'action. M. RAYMOND AIMOS a trouvé un rôle qui lui permet de montrer ses Indéniables qualités d'ACTEUR et nous espérons le revoir AIMAS est un remarquable interprête de CINÉMA PARLANT Il. sont tous excellents, d'ailleurs, les ACTEUR* de ce film magnifique CitocS'lei comme à parade: Mme LINE NORO, superbe toujours MADY BERRY, si simple en son naturel Vers dnt le talent mérite des rôles plus amples, et MM. RAYMOND CORDY, nouveau Jolivet. mécano cette fois, niais rigolard et sentimental H-A. SCIIEFTOW, ANDREW EÛFLEJMAFL, P. GENSCHAW. PIERRE PIÇRADE et RENT BERGERON. Enfin, M. PLERRE BLANCBAR, dur, énergique, .solide et fort comme il sied à Un chef. Belle création Nous n'en sommes plus à louanger M. PIERRE BLANCHAR qui est evec quelque* autres l'honneur de notre ART DRAMATIQUE, mais une coupure malheureuie dans le texte tious troubîe sur la personnalité de son personnage, L'hommer par cette amputation, devient une sorte d'aventurier alors qù'H «t comme beaucoup d'entre nom qui avons combattu, un dégoûté, un écœuré. N'empiche que Au bout du mondt est une de ces oeuvres qui rous redcïïEent une magniSquç espérance dans l'art cinématographique C'est mieux que bleu, c'est beau.

PIERRE BLANCHAR

Type: **PERSON**

Normalized: **Pierre Blanchar**

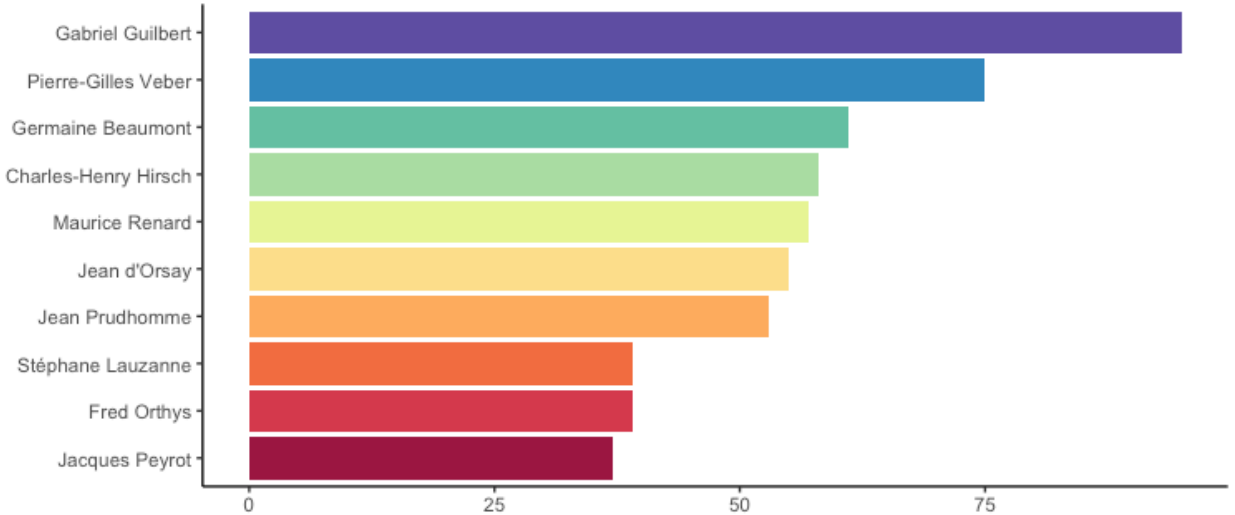
Domains: **Medicine, Psychiatry**

conf: *0.9994*

Pierre Blanchar, né **Gustave Pierre Blanchard** le à Philippeville (aujourd'hui Skikda, en Algérie), et mort le à Suresnes (Hauts-de-Seine) est un acteur et metteur en scène français.

place of interment	Q781858
sex or gender	Q6581097
occupation	Actor
occupation	Film director
occupation	Q10800557
VIAF ID	73772790
ISNI	0000 0001 1936 2114
IMDb ID	nm0087018

Parse the metadata: signatures

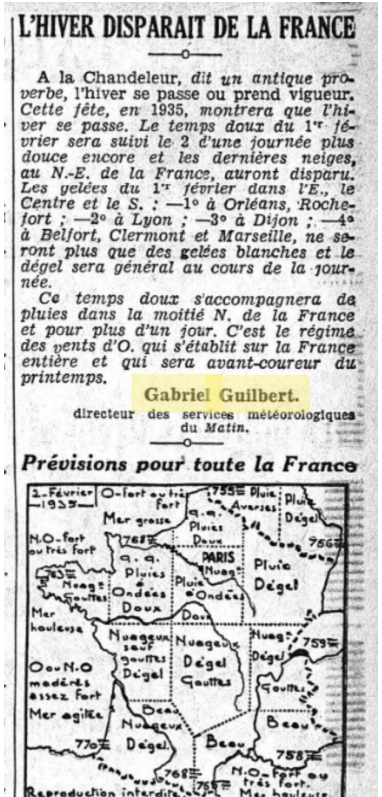


The top ranking of the most productive writers of the *Matin* in 1935 includes well-known *plumes* (Germain Baumont, Maurice Renard) as well as… unexpected profiles

NUMA PRESSE
numapresse.org

Parse the metadata: signatures

Gabriel Guilbert is the more proficient author of 1935 and he acts as… the weatherman of the *Matin* (and he can't be located anywhere in bibliographic databases)



L'HIVER DISPARAIT DE LA FRANCE

A la Chandeleur, *dit un antique proverbe*, l'hiver se passe ou prend vigueur. *Cette fête, en 1935, montrera que l'hiver se passe. Le temps doux du 1er février sera suivi le 2 d'une journée plus douce encore et les dernières neiges, au N.-E. de la France, auront disparu. Les gelées du 1er février dans l'E., le Centre et le S. : —1° à Orléans, Rochefort ; —2° à Lyon ; —3° à Dijon ; —4° à Belfort, Clermont et Marseille, ne seront plus que des gelées blanches et le dégel sera général au cours de la journée.*

Ce temps doux s'accompagnera de pluies dans la moitié N. de la France et pour plus d'un jour. C'est le régime des vents d'O. qui s'établit sur la France entière et qui sera avant-coureur du printemps.

Gabriel Guilbert.
directeur des services météorologiques du *Matin*.

Prévisions pour toute la France

Parse the metadata: signatures

Guibert is not an exceptional case. Both the “garden section“ of Jacques Peyrot of the movie critics of Gilbert Bernard hasn’t been left much traces…

CAUSERIE HORTICOLE

D’autres murs D’autres jardins

Nous enjambons une grande rivière mais nous ne changeons pas de province. Ce n’est plus la même ville, mais c’est une ville du même âge, du temps où elles se tassaient autour du château fort, autour de l’église, maisons contre maisons, avec des jardins comme mis en pénitence et qui en prenaient leur parti.

Le jardin de mon ami de lundi dernier s’offrait à nous dès la porte poussée : logis à droite, jardin à gauche, avec du pavé entre les deux. Pour atteindre le jardin de mon ami d’aujourd’hui, il faut traverser un vestibule, entrer dans un salon puis dans une salle à manger. Ecartons les rideaux, le voici. Le soleil a mis moins de temps que nous ; il est entré par la grande porte du ciel.

Les arbres ont donné leurs fruits. Un parfum de compote d’abricots se mêle au bruit grêle du jet d’eau. La cloche du beffroi sonne lentement et des oiseaux se chamaillent, pour rire, dans le jasmin de Virginie (alias *bignonia*) qui s’accroche aux vieux murs et semble verser, par les cornes d’abondance de ses fleurs couleur feu, les plus belles prémices de l’avenir.

Du jasmin, les yeux se portent vers la tour qui le domine et qui fait partie du vénérable logis. C’est la tour d’Agnès Sorel. La belle Agnès habita cette maison, appuya ses jolies mains sur les pierres plates qui terminent le mur de la terrasse, regarda ces vieux quartiers en contre-bas que nous regardons avant de visiter du grenier à la cave (une cave à cinq étages, à cinq glacières) de ce logis solide qui, après tant de siècles ne demande qu’à revivre, qu’à sourire, en écoutant la musique nouvelle de son eau qu’une même source continue d’amener.

Jacques Peyrot.

LES NOUVEAUX FILMS

« Cavalerie légère »

Il y a dans ce film d’excellents condiments. Un cirque en liberté, un puissant dompteur, un clown poète, un garçon d’écurie qui joint à cette fonction le titre de baron, une mise en scène fastueuse et même un superviseur. Mais *Cavalerie légère* ne galope pas comme nous le voudrions parce que cette réalisation, version allemande d’un film français, version calquée pourrait-on dire sur l’original, ne répond pas au goût du jour. Et voici l’erreur, nous sommes aux antipodes de la mentalité allemande, Babelsberg importe d’excellents artistes qui suivent mot à mot le travail de leurs confrères berlinois. On n’a même pas essayé de transposer, on copie et malgré le faste, la technique et un auteur français, le résultat est tout de même allemand. On annonce que cette formule sera abandonnée et qu’il n’y aura plus qu’une seule version, française celle-ci. Souhaitons que cette nouvelle soit vraie.

Cavalerie légère n’est pas un film militaire, mais le titre d’un ballet somptueux inventé par un clown et qui sera dansé par une ballerine hongroise, écuyère à l’occasion que Chérubini, le directeur du cirque, a recueillie par pitié. Naturellement, le dompteur directorial tombe amoureux de sa protégée. La reconnaissance est rare ici-bas et la star foraine s’amourachera d’un valet de litière, qui, au contraire de Ruy Blas, est vraiment noble, et l’amour à la fin triomphera parmi les fastes d’une mise en scène cossue, clinquante, dansante où l’on a gaspillé les danseuses, les nageuses et les hussards d’opérette. Réalisée par W. Hochbaum et photographiée lumineusement, *Cavalerie légère* est interprétée par Gabriel Gabrio, puissant dompteur, heureux avec les tigresses, malheureux avec les danseuses, Constant Remy, clown inspiré et émule de Max Reinhardt, Louis Alibert, valet de cœur, et Mlle Mona Goya, qui danse et fait danser des cavales et des cœurs, caresse des lionceaux et des poncifs, est l’héroïne de ce film, dont les dialogues très sirupeux ont été écrits par M. Vitrac.

Gilbert Bernard.

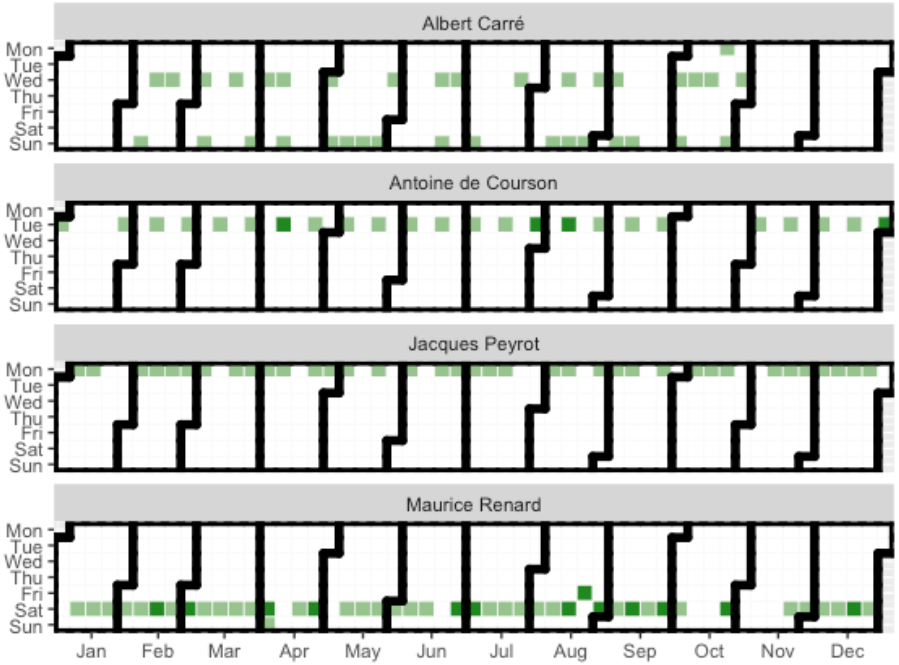
« Veille d’armes »

Le Madeleine Cinéma présente en première exclusivité, un des films les plus importants de l’année : *Veille d’armes*. Marcel L’Herbier a assuré la mise en scène de cette production dont l’émouvant et noble sujet est inspiré d’une œuvre de Claude Farrère, de l’Académie française, et de Lucien Nepoty. Annabella et Victor Francen, avec Signoret et Pierre Renoir, Rosine Deréan, Roland Toutain et Robert Vidalin, de la Comédie-Française, sont les principaux interprètes de *Veille d’armes*.

Des scènes impressionnantes se succèdent au cours de ce film réalisé, en partie, à bord de navires de notre marine nationale. La poursuite d’un bâtiment corsaire dans la brume ; la fin d’un croiseur atteint par une torpille ; l’épisode du tribunal militaire où le capitaine répond de la perte de son vaisseau ont inspiré à Marcel L’Herbier autant d’images qui honorent son talent. La qualité de la réalisation et la valeur des interprètes, tout contribue au succès de cette belle œuvre.

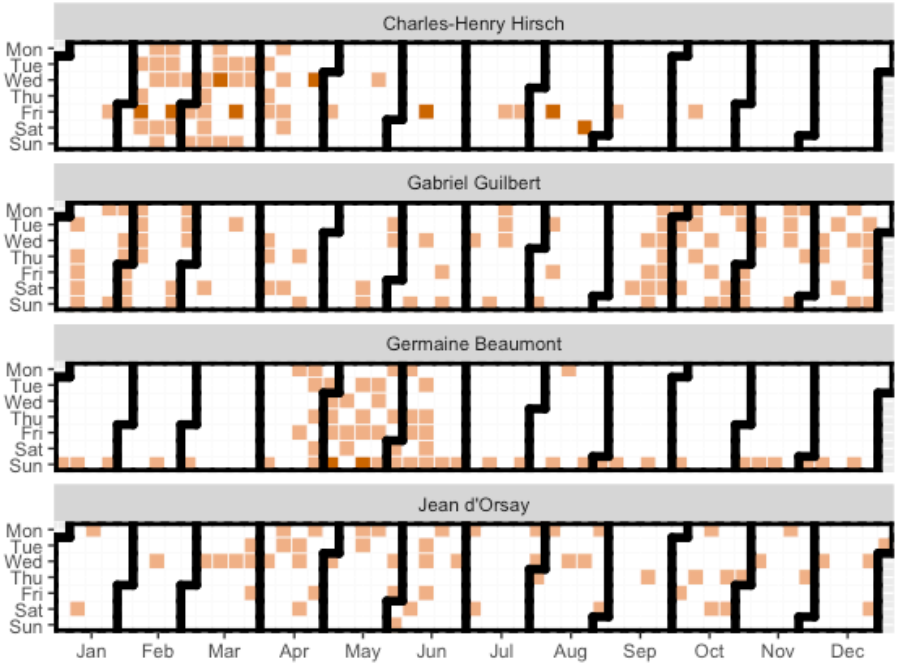
Parse the metadata: signatures

Thanks to signatures, we can infer to some extent the sociological and economic status of the writer within the newspaper. The regular collaborators write constantly on the same weekdays, through the entire year…



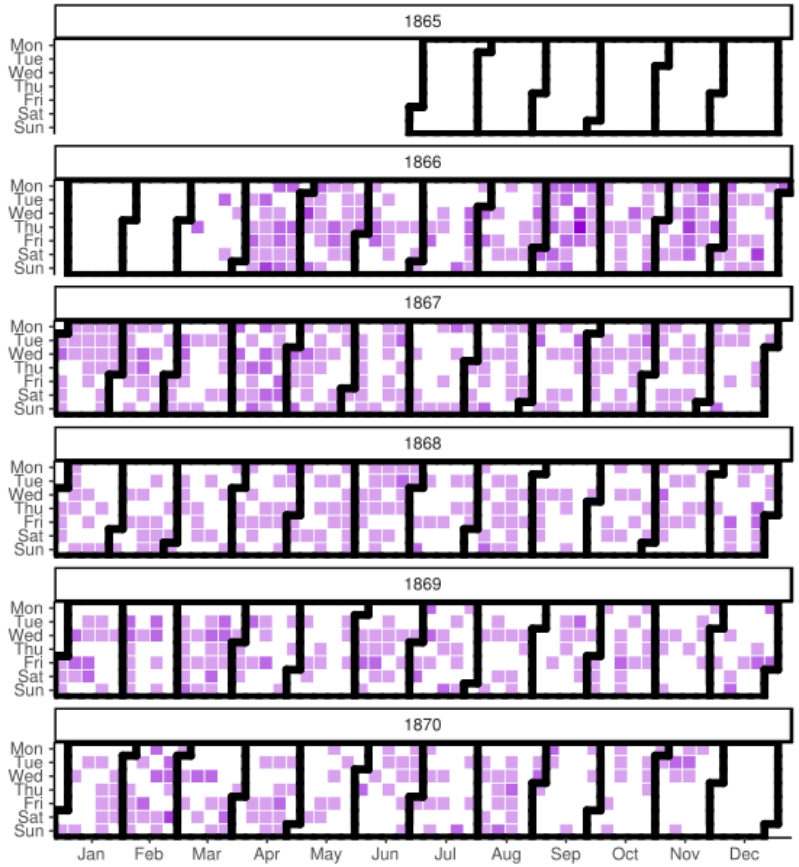
Parse the metadata: signatures

…while some writers tends to have a much more haphazard presence, that may be in some cases caused by a more “contractual” status that may only lasts for several month of intensive activities.



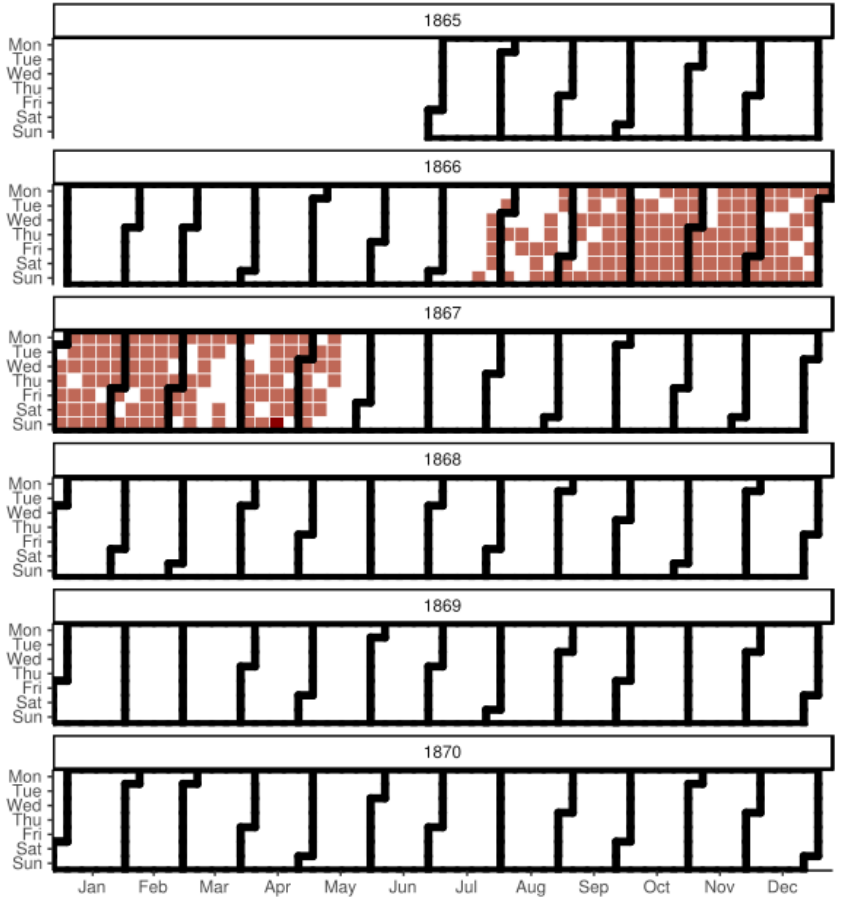
Parse the metadata: signatures

For the *Liberté* we were able to extend this analysis on several years, thanks to the comprehensive annotations created by the French National Library that includes the signature (with METS files). For instance, the scientific journalist Wilfried de Fonvielle had secured a long-term contract.

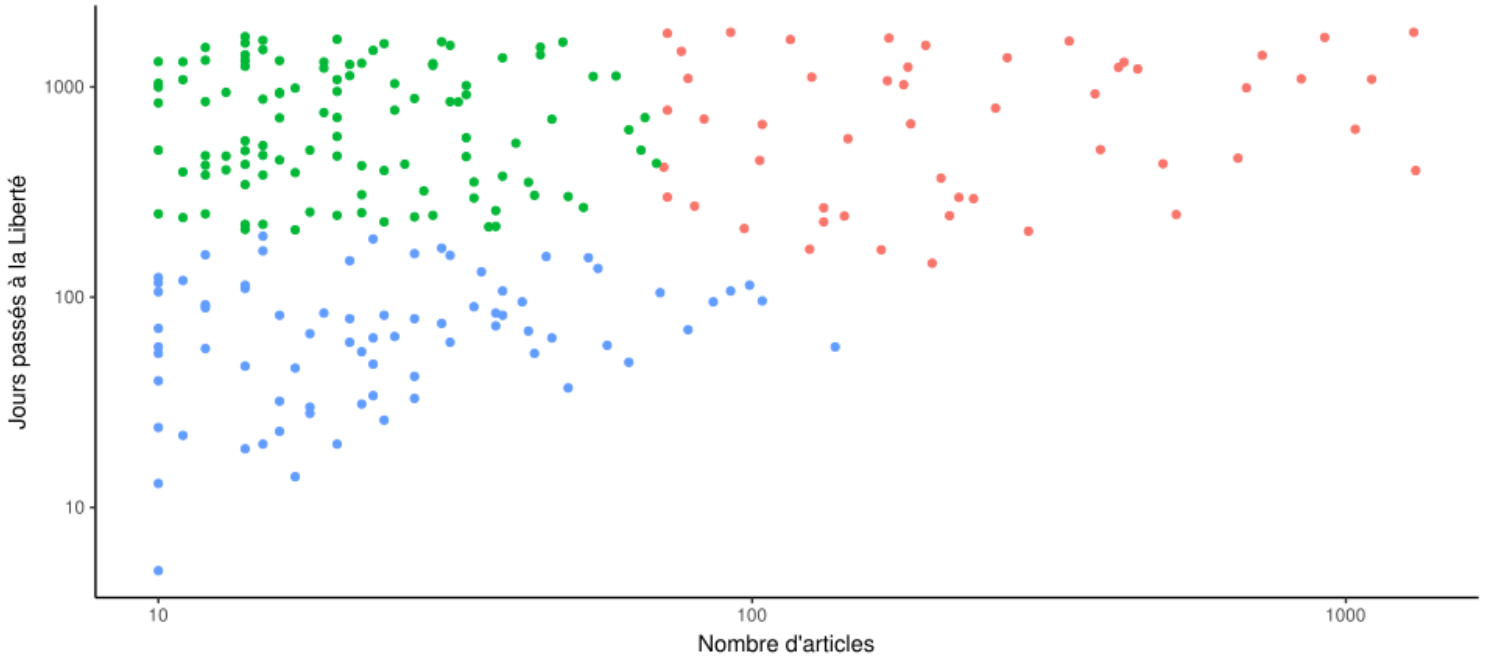


Parse the metadata: signatures

…whereas the food chronicler, Brisse, was quickly fired.



Parse the metadata: signatures



To some extent, theses status could be modeled and categorized (although with significant uncertainties, as signatures can include pseudonyms or some collaboration may not require an explicit signature).

Parse the metadata: portraits

ALFRED EDWARDS qui vient de disparaître si brusquement, non pas de la scène du monde mais de la scène parisienne, fut une de ces personnalités qu'il ne faut pas négliger si on veut comprendre le temps où l'on vit. Il était né àConstantinople d'un père ANGLAIS et d'une mère FRANÇAISE qui s'appelait CAPORAL. Son père avait réa lisé en TURQUIE une fortune considérable, àl'époque des grandes et fructueuses en treprises et il courait sur ses origines les légendes les plus variées. Selon que l'on voulait humilier ALFRED EDWARDS ou le fla gorner ,son père avait été un banquier puissant prêtant àgros intérêts au SULTAN prodigue, ou un coiffeur profitant, pour spéculer des confidences des PACHAS qu'il tenait àsa discrétion sous la lame de son rasoir. Quoiqu'il en fût, M. et Mme EDWARDS s'étaient installés àParis dans un hôtel voisin de celui de M. THIERS, place SaintGeorges, et comme leur fils unique avait à sa disposition beaucoup d'argent, il «in terrompit fort jeune ses études », selon la formule discrète (l'un de ses historio graphes. pour faire tout simplement la fête, comme on disait en ce temps-là. On dit qu'il débuta de bonne heure dans la presse. Il serait plus exact de dire qu'il fut attiré beaucoup moins vers le JOURNALISME que vers les journalistes par le désir de rencontrer des écrivains d'énormément d'esprit comme AURÉLIEN SCHOLL, ALBERT WOLFF et tant d'autres dont la réputation n'a pas survécu. La situation d'ALFRED EDWARDS lui permettait d'être un amateur et le crédit dont il jouissait dans les jour naux lui servait plutôt àglisser dans les échos de théâtre quelques mots élogieux pour les personnes auxquelles il voulait du bien, qu'à faire œuvre de ce qu'on appelle pompeusement le grand reportage. Le boulevard —puisqu'il yavait encore un boulevard —apprit àconnaître ALFRED EDWARDS quand JULES CORNÉLY, qui venait du FIR/ARO ET DU GAULOIS, —le GAULOIS de TARBÉ DES SABLONS qui passait l'été àPourville, —fonda le Clairon, organe royalis te. 11 en était nominalement le SECRÉTAIRE DE RÉDACTION et ces fonctions ne l'occu paient guère. Sa situation n'était pas moins singulière, car il était àcette épo que, —et pour une dizaine d'années en core, —sujet ANGLAIS. Il est vrai que le comte

ALFRED EDWARDS

Type: **PERSON**

Normalized: **Alfred Edwards**

Domains: **Health, Medicine, Enterprise, Commerce, Biology**

conf: *0.9791*

Alfred Charles Edwards, né à Constantinople le 10 juillet 1856 et mort à Paris le 10 mars 1914, est un journaliste et patron de presse d'origine anglaise, fondateur du quotidien *Le Matin*.

place of death	Paris
VIAF ID	31985591
sex or gender	Q6581097
place of birth	Constantinople
date of birth	1856-07-10
date of death	1914-03-10
place of interment	Père Lachaise Cemetery
Léonore ID	LH/891/20
country of citizenship	United Kingdom
instance of	Human

Journalists have always enjoyed to talk about… themselves. Since the Second Empire, journalists portrait becomes a widespread genre, that do not always contain very accurate information but at least document an important process of self-promotion.

Parse the metadata: portraits

A first experimental trial of automated identification of journalist portraits in *Le Figaro* of 1862 (thanks to a metric of *personnalization* of a given texts)

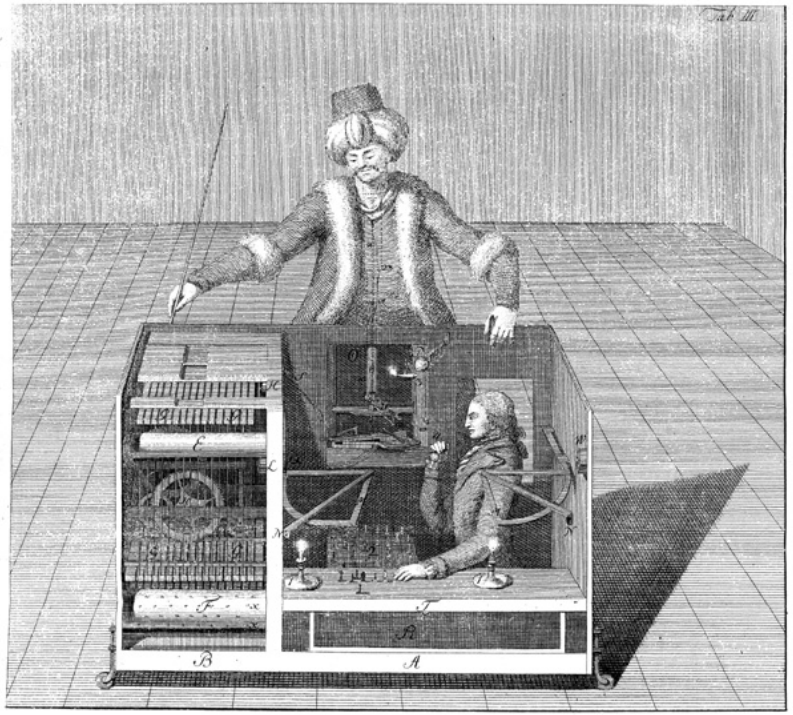
	lemma	mean_tf_idf	count_lemma	pers_entity	page_id
1	Merluchette	0.031210693	32	0.6875000	figaro_1862-08-10_1_2
2	Ducuing	0.021026594	23	0.3913043	figaro_1862-11-23_1_5
3	Noé	0.004966037	84	0.9523810	figaro_1862-11-23_1_1
4	Cuit	0.011995387	17	0.4117647	figaro_1862-09-21_1_6
5	Delaage	0.006437054	27	0.9259259	figaro_1862-10-19_1_6
6	Chasles	0.003253814	49	0.5918367	figaro_1862-01-02_1_1
7	Foucher	0.004767351	26	0.8076923	figaro_1862-06-26_1_5
8	Malibran	0.005936120	53	0.7735849	figaro_1862-01-09_1_2
9	Polynice	0.011182846	31	0.3225806	figaro_1862-11-09_1_2
10	Gounod	0.002410257	48	0.3750000	figaro_1862-02-16_1_5
11	Zabban	0.021500261	10	1.0000000	figaro_1862-12-07_1_5
12	Escudier	0.013043389	29	0.4137931	figaro_1862-04-03_1_5
13	Etéocle	0.017756399	22	0.4545455	figaro_1862-11-09_1_2
14	Buhot	0.011184823	19	0.7894737	figaro_1862-01-02_1_7
15	Sand	0.002032593	173	0.9710983	figaro_1862-10-12_1_3
16	Champfleury	0.003917406	60	0.6833333	figaro_1862-10-09_1_6

The next challenges: automated recognition of the article

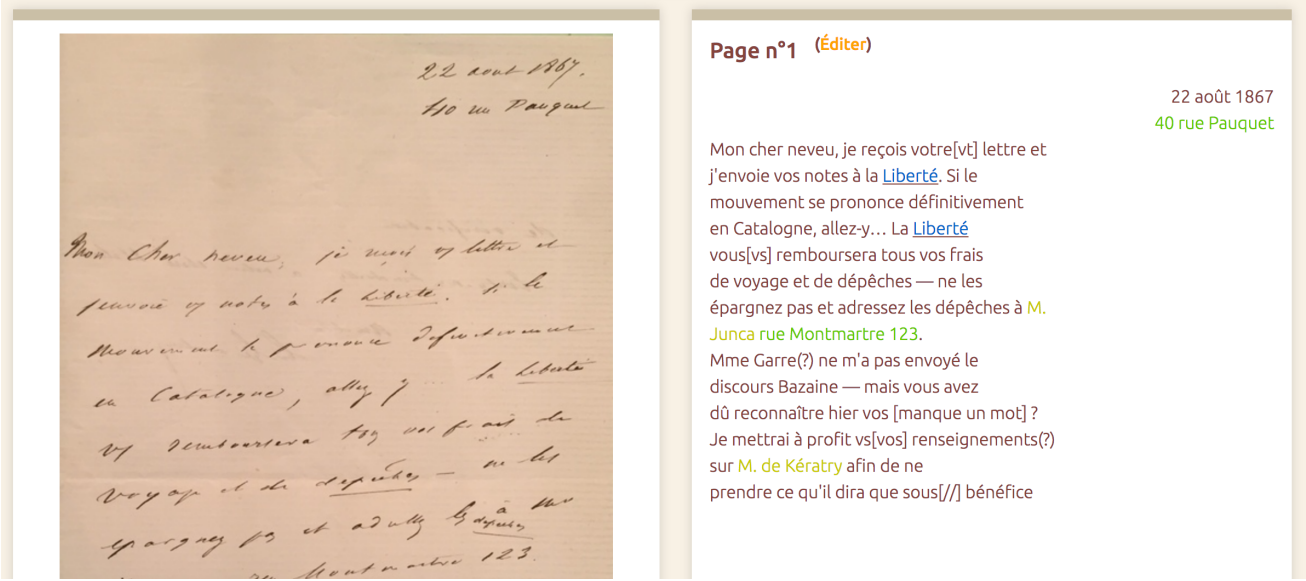
This technical challenge remains unsolved. Contractors for digital libraries mostly use the *digital labor* of developing countries.

A possible approach within Numapresse.

- Map all the information extracted through genre classification and editorial structure analysis (signatures, titles, global page geographies like the *feuilleton*)
- Use several competing definitions for an *article*. In the daily press before 1870, most news are made of free-floating paragraphs heavily recombined, without any titles or headers.



The next challenges: mixing news texts with news archives.



A partnered project of Numapresse, Giranium collects and transcribes the correspondence of the main French media tycoon of the XIXth century, Émile de Girardin. While this work has helped a lot to understand the inner working of a newspaper (*La Liberté*), much remain to be done with archives

4. Mapping news circulation and the media ecosystem

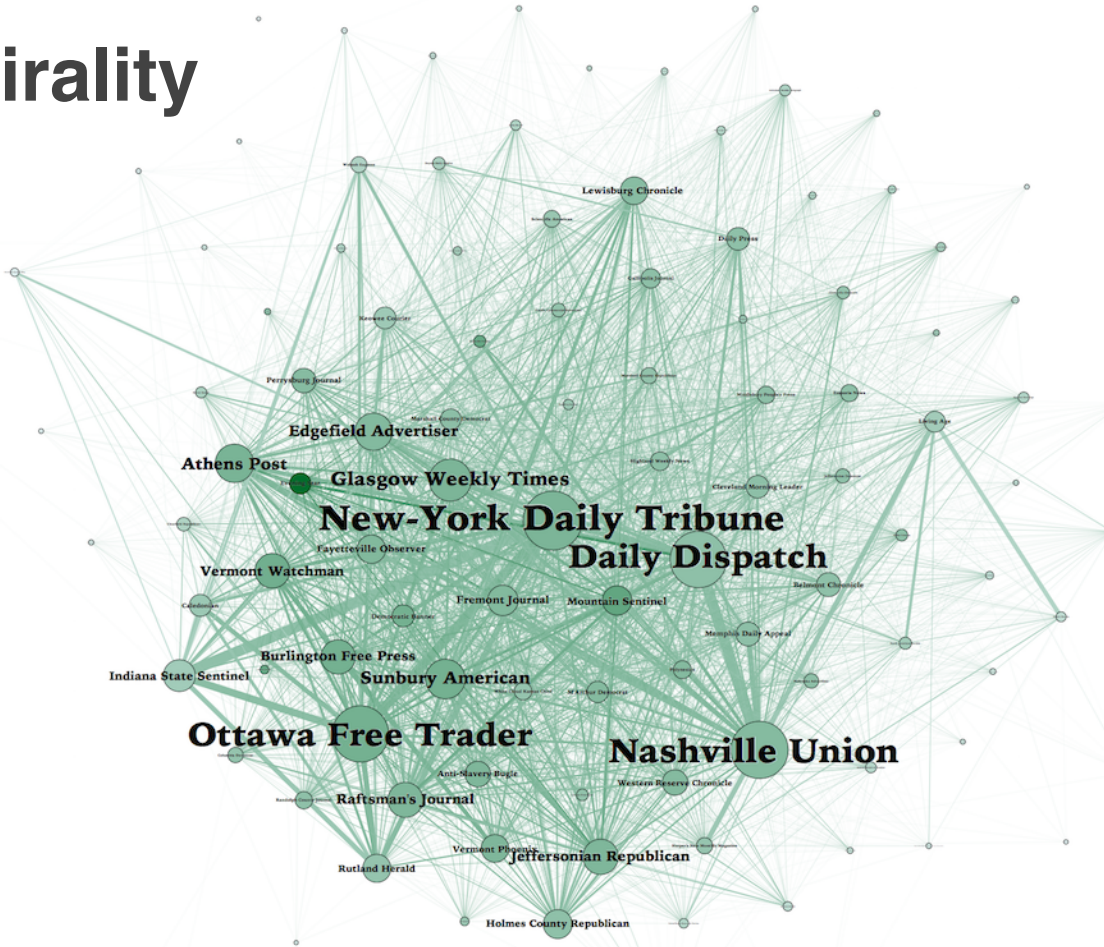
From reprints to networks…

Modeling the virality

The **Viral Text** project lead by Ryan Cordell has initiated a new kind of tools to study XIXth century press: reprinting detection.

At a time where copyright did not apply to newspapers, copies were widespread and their identification can recover an extensive media ecosystem with regular “flows” between periodicals.

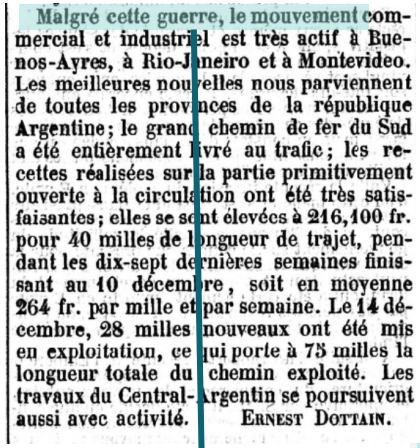
This research is now also extended to other countries through the **Oceanic Exchange** project.



Modeling the virality

We have used the same methods as the Viral Text project coordinated by Ryan Cordell

- On account of OCR mistakes it is unlikely to match “exact” reprints.
- Using wide corpus implies a “combinatory explosion”: for 100 000 paragraphs there is 5 billions possible pairs of text.
- The solution: using a sample of “shared formula” (ngram) that serve to identify if a reprint is probable.


Malgré cette guerre, le mouvement commercial et industriel est très actif à Buenos-Ayres, à Rio-Janeiro et à Montevideo. Les meilleures nouvelles nous parviennent de toutes les provinces de la république Argentine; le grand chemin de fer du Sud a été entièrement livré au trafic; les recettes réalisées sur la partie primitivement ouverte à la circulation ont été très satisfaisantes; elles se sont élevées à 216,100 fr. pour 40 milles de longueur de trajet, pendant les dix-sept dernières semaines finissant au 10 décembre, soit en moyenne 264 fr. par mille et par semaine. Le 14 décembre, 28 milles nouveaux ont été mis en exploitation, ce qui porte à 75 milles la longueur totale du chemin exploité. Les travaux du Central-Argentin se poursuivent aussi avec activité. ERNEST DOTTAIN.

Journal des débats, 26 january 1866

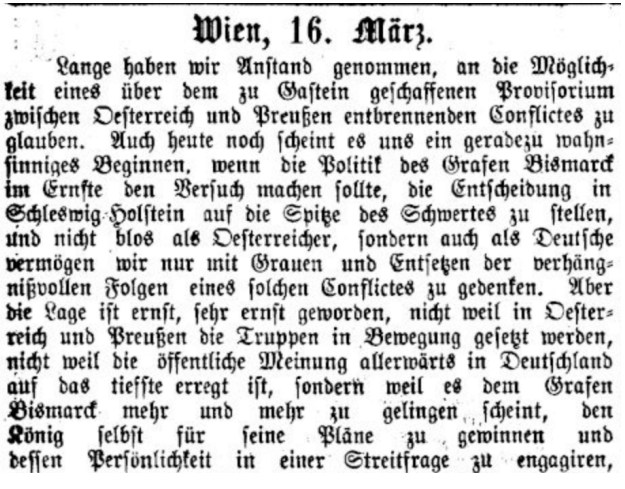
shared ngram: “Malgré cette guerre, le mouvement”


Malgré cette guerre, le mouvement commercial et industriel est très actif à Buénos-Ayres, à Rio Janeiro et à Montevideo. Les meilleures nouvelles nous parviennent de toutes les provinces de la république Argentine ; le grand chemin de fer du Sud a été entièrement livré au trafic ; les récettes réalisées sur la partie primitivement ouverte à la circulation ont été très satisfaisantes ; elles se sont élevées à 216,100 fr pour 40 milles de longueur de trajet, pendant les dix-sept dernières semaines finissant au 10 décembre, soit en moyenne 264 fr. par mille et par semaine. Le 14 décembre, 28 milles nouveaux ont été mis en exploitation ce qui porte à 75 milles la longueur totale du chemin exploité. Les travaux du Central-Argentin se poursuivent aussi avec activité.

Courrier du Canada, 16 February 1866

=> 108 articles were identified in at least a French and a Canadian newspaper

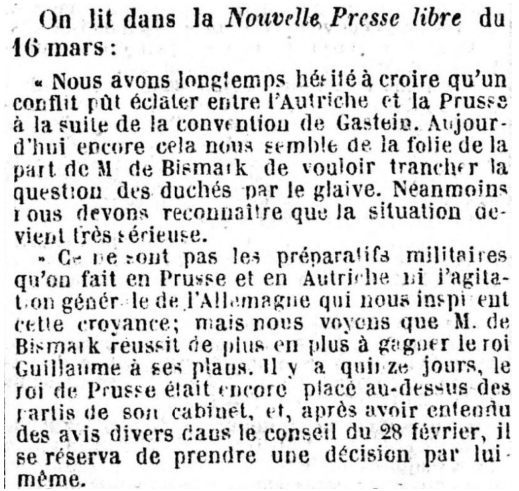
Everything is viral



Wien, 16. März.

Lange haben wir Anſtand genommen, an die Möglichkeit eines über dem zu Gaſtein geſchaffenen Proviſorium zwiſchen Oeſterreich und Preußen entbrennenden Conflictes zu glauben. Auch heute noch ſcheint es uns ein geradezu wahnſinniges Beginnen, wenn die Politik des Grafen Bismarck im Ernſte den Verſuch machen ſollte, die Entſcheidung in Schleswig-Holſtein auf die Spitze des Schwertes zu ſtellen, und nicht blos als Oeſterreicher, ſondern auch als Deutſche vermögen wir nur mit Grauen und Entſetzen der verhängnißvollen Folgen eines ſolchen Conflictes zu gedenken. Aber die Lage iſt ernſt, ſehr ernſt geworden, nicht weil in Oeſterreich und Preußen die Truppen in Bewegung geſetzt werden, nicht weil die öffentliche Meinung allerwärts in Deutſchland auf das tiefſte erregt iſt, ſondern weil es dem Grafen Bismarck mehr und mehr zu gelingen ſcheint, den König ſelbſt für ſeine Pläne zu gewinnen und deſſen Perſönlichkeit in einer Streitfrage zu engagiren,

An article of the *Neue Freie Press* of Vienna (17 March) on the Austro-Prussian war

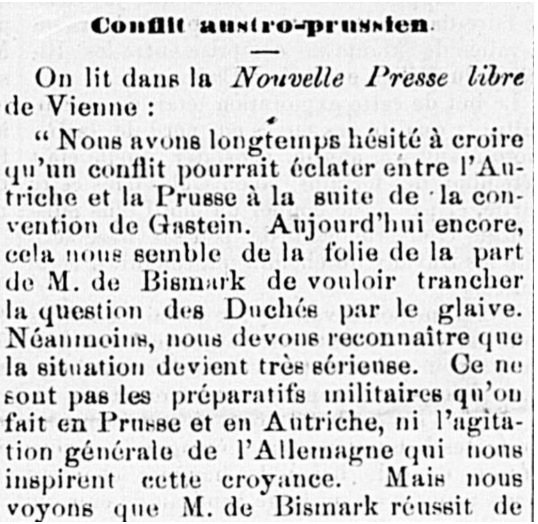


On lit dans la *Nouvelle Presse libre* du 16 mars :

« Nous avons longtemps hésité à croire qu'un conflit pût éclater entre l'Autriche et la Prusse à la suite de la convention de Gastein. Aujourd'hui encore cela nous semble de la folie de la part de M de Bismark de vouloir trancher la question des duchés par le glaive. Néanmoins nous devons reconnaître que la situation devient très sérieuse.

« Ce ne sont pas les préparatifs militaires qu'on fait en Prusse et en Autriche ni l'agitation générale de l'Allemagne qui nous inspirent cette croyance; mais nous voyons que M. de Bismark réussit de plus en plus à gagner le roi Guillaume à ses plans. Il y a quinze jours, le roi de Prusse était encore placé au-dessus des partis de son cabinet, et, après avoir entendu des avis divers dans le conseil du 28 février, il se réserva de prendre une décision par lui-même.

Translated in the *Journal des débats* (21 March)



Conflit austro-prussien.

On lit dans la *Nouvelle Presse libre* de Vienne :

" Nous avons longtemps hésité à croire qu'un conflit pourrait éclater entre l'Autriche et la Prusse à la suite de la convention de Gastein. Aujourd'hui encore, cela nous semble de la folie de la part de M. de Bismark de vouloir trancher la question des Duchés par le glaive. Néanmoins, nous devons reconnaître que la situation devient très sérieuse. Ce ne sont pas les préparatifs militaires qu'on fait en Prusse et en Autriche, ni l'agitation générale de l'Allemagne qui nous inspirent cette croyance. Mais nous voyons que M. de Bismark réussit de

Reprinted in the *Courrier du Canada* (11 April)

International news

Everything is viral

FEUILLETON DU JOURNAL DES DEBATS
DU 4 SEPTEMBRE 1866.

LES BONNES FORTUNES PARISIENNES.

Les Amours d'un notaire.

(*Voir* les Numéros des 28, 29 et 30 août.)

IX.

Une fois dans le lieu saint, Mlle Loulou prit une mine recueillie dont je ne l'aurais pas crue capable. Elle me donna de l'eau bénite, se mit à genoux sur le pavé, et pria avec une singulière ferveur. Cette grande église me reportait à l'église de ma petite ville où j'avais prié si souvent à côté de ma mère. Je m'appuyai sur un pilier, et je restai tout entier dans mes souvenirs.

J'en fus tiré par la main de Mlle Loulou.

—Tu as envie de pleurer; retiens-toi, les hommes doivent avoir du courage.

Et elle ajouta :

— Sais-tu pour qui j'ai prié? C'est pour ta maman.

Dieu me le pardonnera; mais je pris la chère enfant dans mes bras et je l'embrassai devant lui.

— Je crois, me dit-elle, que nous sommes très bons tous les deux, et que nous aurions pu être frère et sœur pour tout à fait.

Quand nous sortîmes de l'église, Mlle Loulou me dit :

— Raconte-moi encore ta maman.

Et quand elle m'eut écouté :

— Je ne sais pas bien pourquoi je n'ai pas eu de maman, me dit-elle, ni de papa. Il paraît que j'ai été orpheline tout de suite; c'est Mme la directrice qui m'a élevée; ce n'est pas une mauvaise femme, mais cela ne doit pas être la même chose. J'ai une tante à Dresden, qui envoie quelquefois de petits cadeaux à madame pour moi, mais elle n'est pas aussi amusante que ton oncle. Tout ce que tu me dis de ton oncle m'amuse beaucoup.

Nous nous promenâmes longtemps dans les rues. Mlle Loulou me raconta son éducation et ses débuts : elle avait eu beaucoup de mal; une fois elle s'était cassé la jambe parce qu'un truc avait manqué sous elle; une autre fois on l'avait sifflée; évidemment le sifflet lui était resté sur le cœur plus que la jambe cassée.

— C'était une cabale d'une grande danseuse jalouse, me dit-elle.

Mlle Loulou s'arrêtait à toutes les boutiques. Elle s'acheta de l'eau de Cologne *chez le vrai Farina* et m'en donna un flacon.

— Il faut que nous sentions bon pendant la route, me dit-elle, et puis après, s'il nous en reste, quand nous ouvrirons nos bouteilles, cela nous rappellera que nous avons été ensemble à Cologne. Il ne faudra jamais l'oublier. Je suis très contente, me disait-elle, d'avoir laissé tomber Cocotte; sans cela, nous ne nous serions peut-être pas parlé, car en voyage je suis très fière, et je ne dis jamais rien à personne; et comme cela n'a pas fait de mal à Cocotte, cela nous a toujours valu notre amitié. Quand vous écrirez à votre maman, il faudra lui demander la permission d'être mon frère. Je vous enverrai ma photographie, celle où j'ai mon costume de sylphide, un rôle où j'ai eu beaucoup de succès. Vous l'enverrez à madame votre maman pour qu'elle voie comment je suis et si cela lui plaît que je sois votre sœur. Vous lui direz aussi que je l'aime bien, et, à votre oncle, que tout ce que vous me dites de lui me fait rire; il est vraiment très drôle.

La nuit était venue. Nous devions nous lever à quatre heures du matin :

— Il faut se coucher de bonne heure, dit-elle; vous êtes un trop bon dormeur, vous, et si nous nous mettions dans nos lits trop tard, vous ne vous réveilleriez pas facilement. Je préviendrai le portier pour qu'il vous réveille. Nous allons souper, mais très peu, car nous avons beaucoup dîné. D'abord, moi, je ne voudrais que des confitures, si vous les aimez aussi. Pour ne pas dépenser beaucoup d'argent, nous entrerons chez le confiseur pour y prendre un pot de gelée de groseille, et après chez un boulanger pour y acheter deux petits pains, et, au lieu de souper en bas, ce qui nous coûterait trop cher, nous

Les Amours d'un notaire of Pierre-Jules Hetzel in le *Journal des débats* (4 September).

J'en fus tiré par la main de Mlle Loulou.

— Tu as envie de pleurer; retiens toi, les hommes doivent avoir du courage.

Et elle ajouta;

— Sais-tu pour qui j'ai prié?— C'est pour ta maman.

Dieu me le pardonnera; mais je pris la chère enfant dans mes bras et je l'embrassai devant lui.

— Je crois me dit elle que nous sommes très bon tous les deux, et et que nous aurions pu être frère et sœur pour tout à fait.

Quand nous sortîmes de l'église, Mlle Loulou me dit :

—Raconte moi encore de tu maman.

Et quand elle m'eut écouté :

— Je ne sais pas bien pourquoi je n'ai pas eu de maman, me dit-elle ni de papa. Il paraît que j'ai été orpheline tout de suite; c'est Mme la directrice qui m'a élevée; ce n'est pas une mauvaise femme, mais cela ne doit pas être la même chose.— J'ai une tante a Dresde, qui envoie quelquefois de petits cadeaux à madame pour moi, mais elle n'est pas aussi amusante que ton oncle. Tout ce que tu me dis de ton oncle m'amuse beaucoup.

Nous nous promenâmes longtemps dans les rues. Mlle Loulou me raconta son éducation et ses débuts: elle avait eu beaucoup de mal; une fois elle s'était cassé la jambe parce qu'un truc avait manqué sous elle; une autre fois on l'avait sifflée; évidemment le sifflet lui était resté sur le coeur plus que la jambe cassée.

— C'était une cabale d'une grande danseuse jalouse, me dit-elle.

Mlle Loulou s'arrêtait à toutes les boutiques. Elle s'acheta de l'eau de Cologne *chez le vrai Farina* et m'en donna un flacon.

— Il faut que nous sentions bon pendant la route, me dit-elle, et puis après, s'il nous en reste, quand nous ouvrirons nos bouteilles, cela nous rappellera que nous avons été à Cologne. Il ne faudra jamais l'oublier. Je suis très contente me disait-elle, d'avoir laissé tomber Cocotte; sans cela, nous ne nous serions peut-être pas parlé, car en voyage je suis très fière, et je ne dis jamais rien à personne; et comme cela n'a pas fait de mal à Cocotte, cela nous a toujours valu notre amitié Quand vous écrirez à votre maman, il faudra lui demander la permission d'être mon frère. Je vous enverrai ma photographie, celle ou j'ai mon costume de sylphide, un rôle ou j'ai eu beaucoup de succès. Vous l'enverrez à madame votre maman pour qu'elle voie comment je suis et si cela lui plait que je sois votre sœur. Vous lui direz aussi que je l'aime bien, et à votre oncle, que tout ce que vous me dites de lui me fait rire; il est vraiment très drôle.

La nuit était venue. Nous devions nous lever à quatre heures du matin :

(A CONTINUER.)

Reprinted (without attribution) by the *Gazette de Sorel* (8 December)

Serial novels

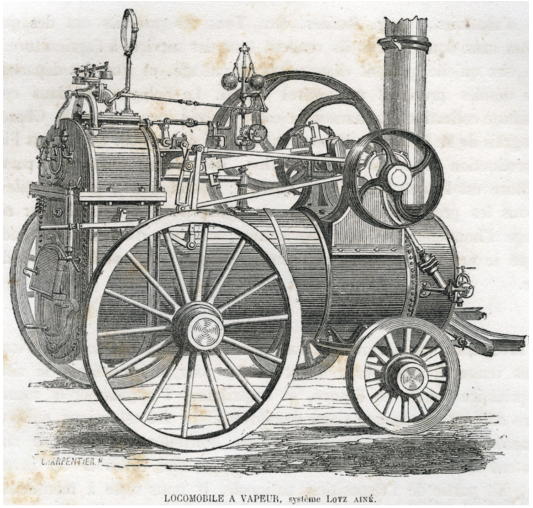
Everything is viral

— Nous avons parlé dernièrement d'une locomotive construite pour marcher sur les routes ordinaires, et qui, partie de Nantes en traînant à sa remorque un certain nombre de voitures, est arrivée à Paris, après avoir effectué son trajet dans les meilleures conditions.

Ce premier essai de locomotion au moyen de la vapeur sur les voies publiques sera bientôt suivi d'essais analogues, car on nous parle, dit *le Moniteur*, d'une nouvelle voiture qu'on verra prochainement circuler dans les rues de Paris concurremment avec nos voitures de place.

Ce véhicule, inventé par un ingénieur auquel on doit un grand nombre de créations nouvelles, a la forme d'un fiacre ordinaire, à part l'avant-train, dans lequel sera placée une petite chaudière; cet avant-train sera précédé d'une cinquième roue, roue directrice, au moyen de laquelle le conducteur pourra tourner ou obliquer à discrétion, et dont l'usage doit enfin donner un démenti à cet ancien dicton qu'on appliquait à quiconque était inutile : « Il sert comme une cinquième roue à un carrosse. »

The trial of an experimental "car" in the *Journal des débats* (22 September)



LOCOMOBILE A VAPEUR, système Lotz aîné.

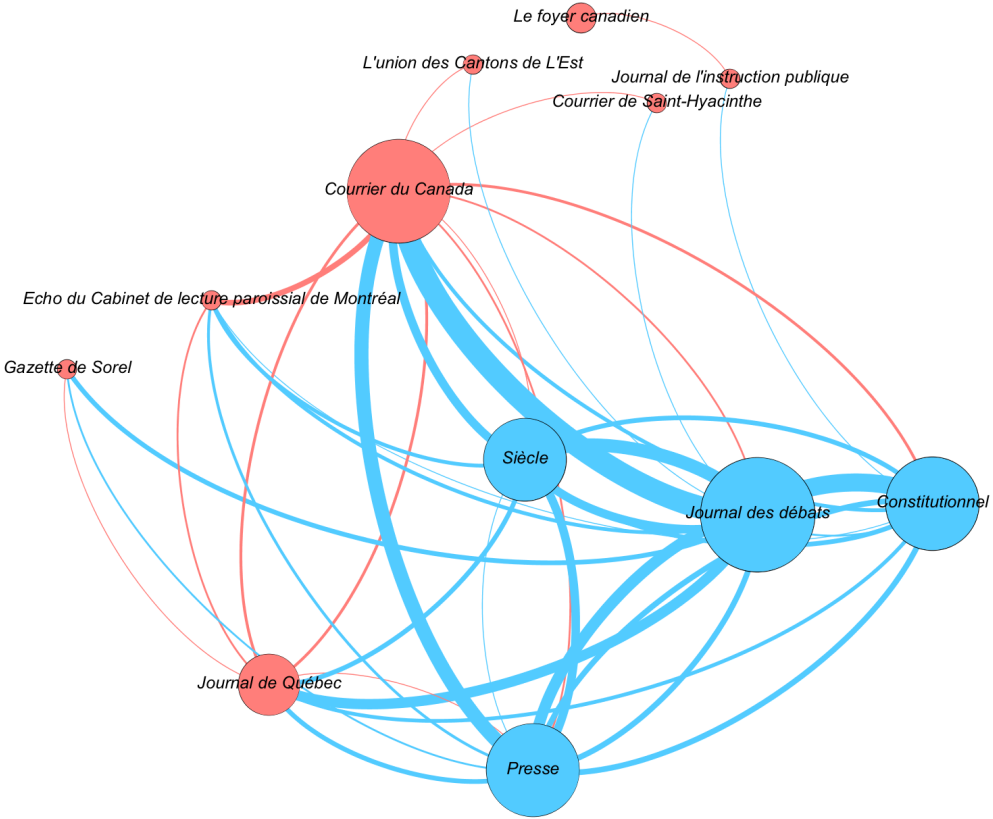
— Les journaux ont parlé d'une locomotive construite pour marcher sur les routes ordinaires, et qui, partie de Nantes, en traînant à sa remorque un certain nombre de voitures, est arrivée à Paris après avoir effectué son trajet dans les meilleures conditions. Ce premier essai de locomotion au moyen de la vapeur sur les voies publiques sera bientôt suivi d'essais analogues, car on parle d'une nouvelle voiture qu'on verra prochainement circuler dans les rues de Paris, concurremment avec les voitures de place. Ce véhicule, inventé par un ingénieur auquel on doit déjà un grand nombre de créations nouvelles, a la forme d'un fiacre ordinaire, à part l'avant-train dans lequel sera placée une petite chaudière ; cet avant-train sera précédé d'une cinquième roue, roue directrice, au moyen de laquelle le cocher pourra tourner ou obliquer à discrétion, et dont l'usage doit enfin donner un démenti à cet ancien dicton qu'on appliquait à quiconque était inutile : *Il sert comme la cinquième roue à un carrosse.*

Reprinted in the *Écho du cabinet de lecture paroissial* (15 October)

Scientific and technical news

Mapping translatlantic networks

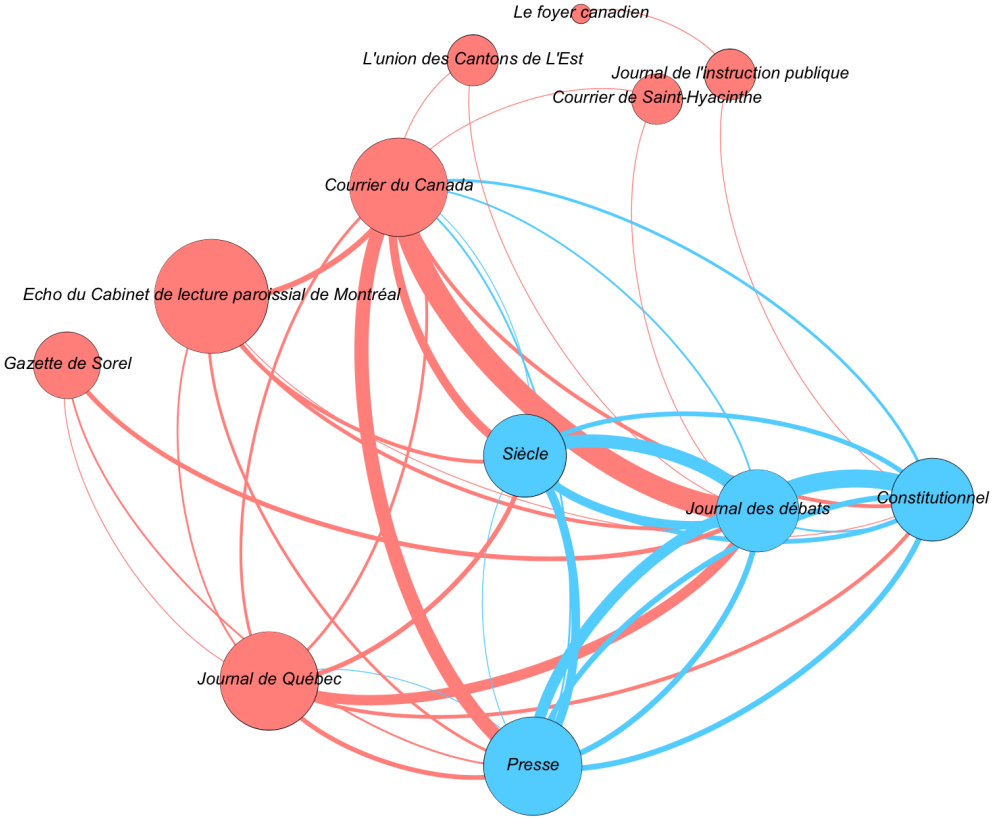
A network visualisation shows that some newspapers are key intermediaries: the *Courrier du Canada* in Québec, the *Journal des débats* in France



Network of “influential” newspapers (those that have been republished)

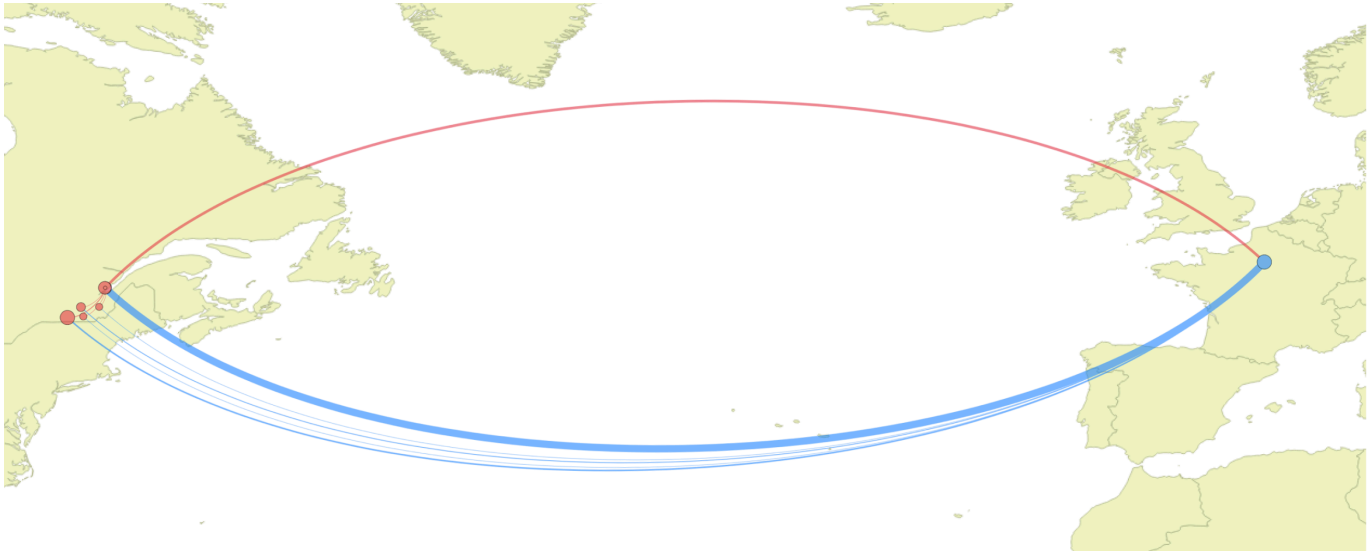
Mapping translatlantic networks

A network visualisation shows that some newspapers are key intermediaries: the *Courrier du Canada* in Québec, the *Journal des débats* in France



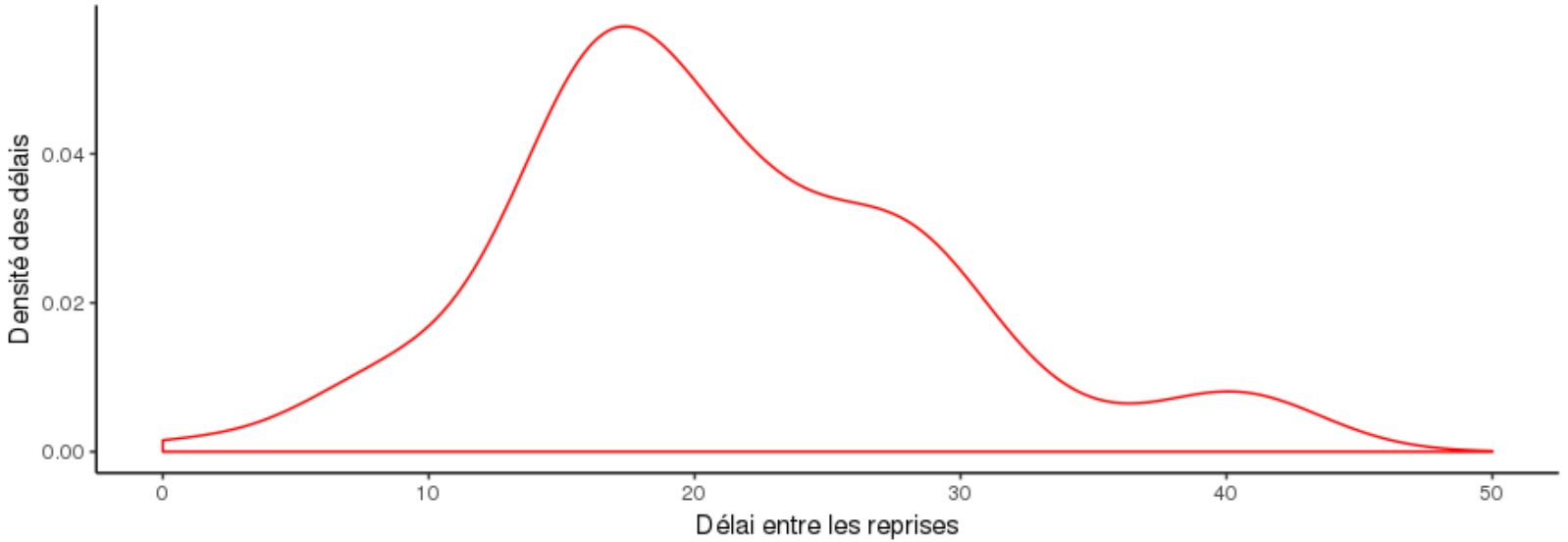
Network of “influential” newspapers (those that have been republished)

Mapping translatlantic networks



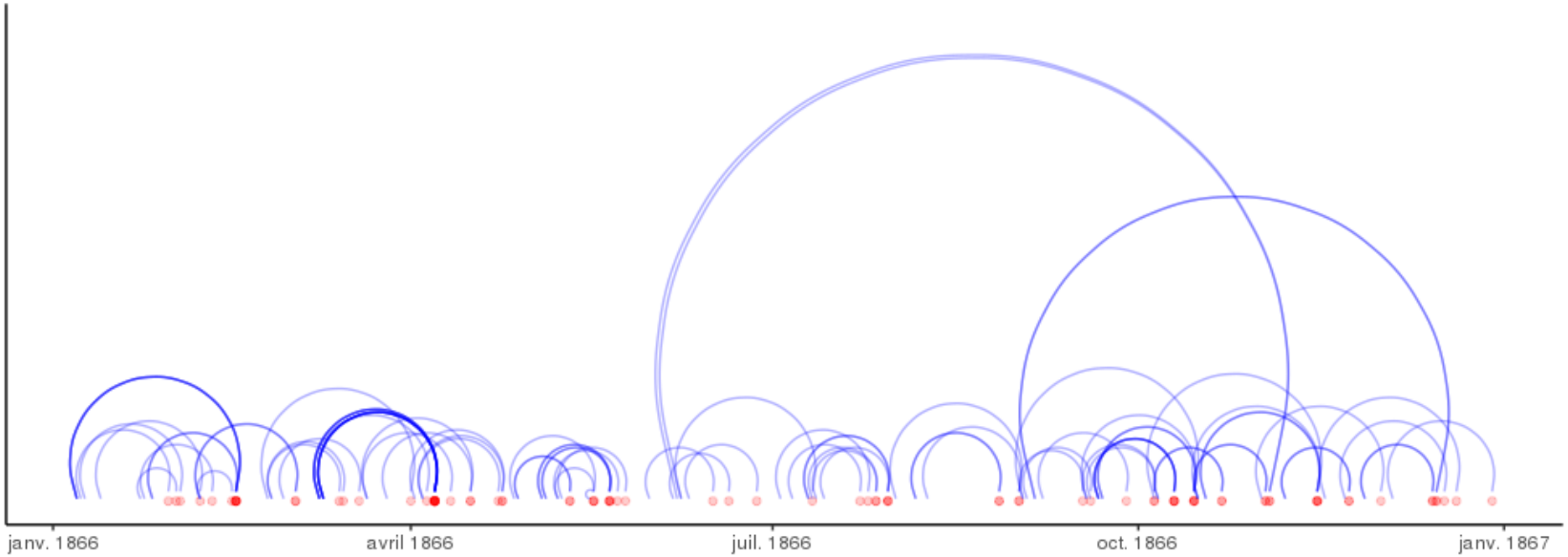
A geographic look at our network: news from France tends to arrive first in Québec and then to ship on the Saint-Laurent all the way till Montréal.

The Mechanics of reprinting “flows”



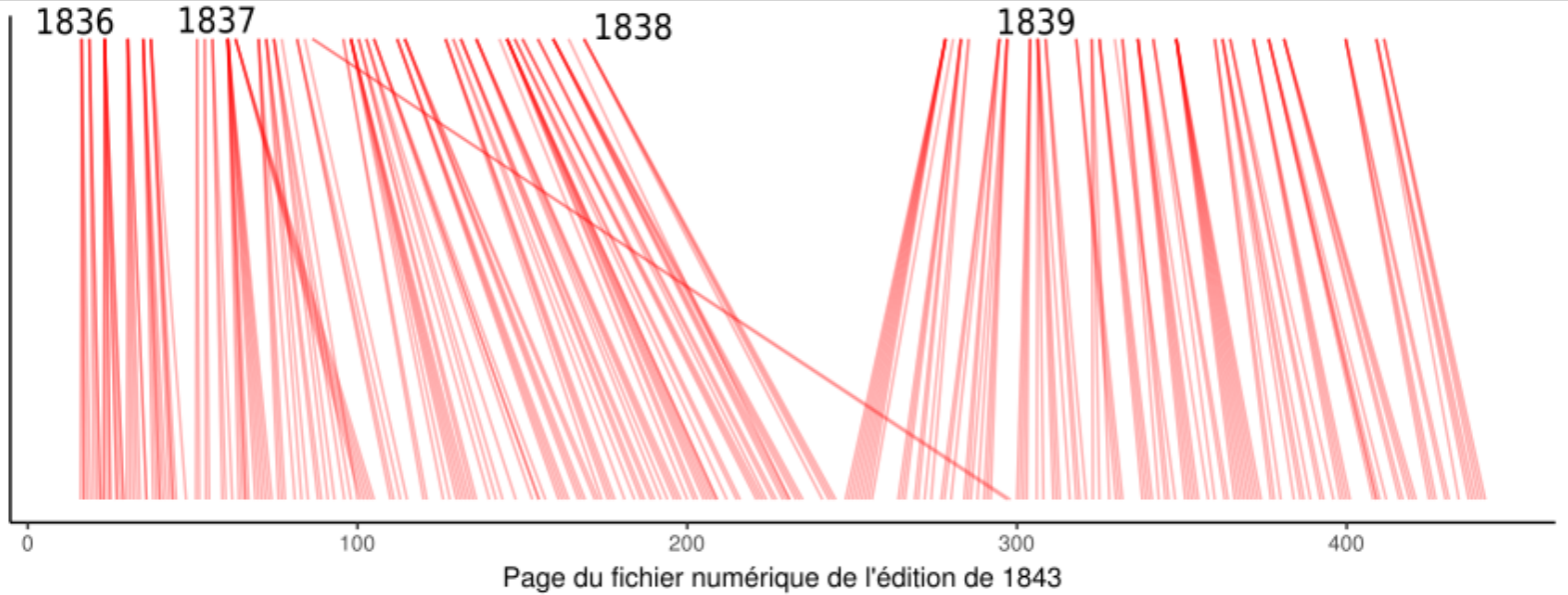
The leading factor of reprinting in 1866: steamers. Most reprinting occur within 12 to 20 days, the standard time of transatlantic crossing.

The Mechanics of reprinting “flows”



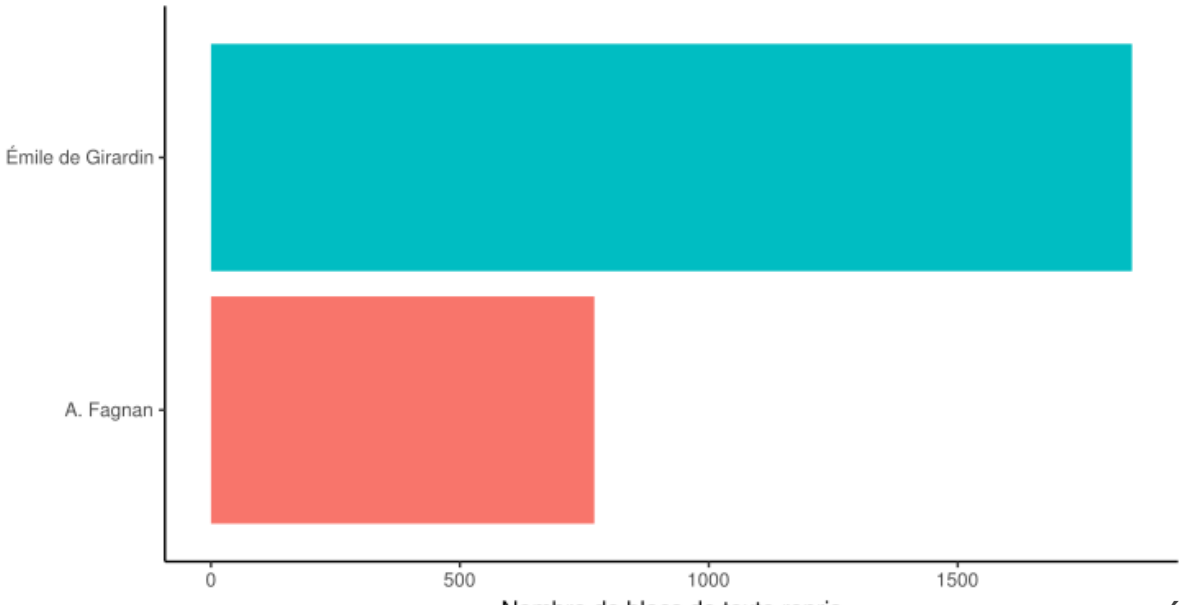
Although they may come from different dates, reprintings usually occur in clusters (red points), which agrees with the maritime rhythms of the 1866 global media system.

Virality beyond newspaper: from articles to book



In the XIXth century French journalists very frequently publish collections of their articles. A variation of *Viral Text* algorithm allows to uncover this mechanism of republication that alter widely the original text.

Virality beyond newspaper: from articles to book



The viral analysis of book publication can also help to identify some pseudonyms. Émile de Girardin reused extensively the texts published under the pen name A. Fagnan.

Virality beyond texts: images as travelling stereotypes

While efficient fingerprinting techniques already exists, *deep learning* classification makes it possible to map a wide range of reproductive practices: reprints, variations, inspirations… With Julien Schuh we has just started working on a much larger project on reprinting and poetics of 6000 images extracted from preeminent literary periodicals of the end of the XIXth century.


72 LA PLUME

MAURICE DENIS. — Illustration pour *Sagesse*.

D'ailleurs le culte que nous avions pour lui et que nous lui gardons, ne peut guère être comparé aux autres. Il est fait d'une reconnaissance toujours nouvelle et infiniment attendrie. Ce n'est pas tant les mots admiration ou sympathie qui le peignent bien, que celui-ci : remercîment perpétuel de nous tous, au seul poète de ce siècle, qui, tout simple et tout ingénu, a chanté.

Edouard Rod. — Ce qui est adorable en Verlaine, ce par quoi il est unique dans notre littérature, trop souvent de convention et d'artifice, c'est son ingénuité. Ah ! qu'elles furent ingénues, ses fautes et qu'ils furent ingénus, ses repentirs ! A l'heure de sa mort, c'est une âme de premier communiant, toute blanche, qui s'est envolée vers Madame Marie, et quand il l'a reçu dans son Paradis — car il y a un Paradis, n'est-ce pas? — le Seigneur a dû dire : « Laissez venir à moi ce petit enfant ! Un enfant qui aime, un enfant qui pèche, un enfant qui se repent, un enfant qui prie, c'est tout le Pauvre Lélian. *Sagesse* corrige *Parallèlement* ; après la folie de la chair, c'est, en toute simplicité, la folie de la croix : « Pardonnez-moi, mon père, parce que j'ai péché. » Et voilà pourquoi, dans l'œuvre de ce douloureux et de ce croyant qui, au moyen-âge, eût écrit *l'Imitation*, je ne sais rien de comparable aux strophes agenouillées de *Sagesse*.

Son rôle dans l'Evolution littéraire ? Mais n'a-t-il pas écrit : « *Et tout le reste est littérature* ! »

Lui mort, deux purs poètes demeurent qu'il faut chérir de tous nos enthousiasmes : Mallarmé et Dierx.

Emile Blémont. — 1° Les meilleures parties de l'Œuvre de Verlaine ? — *Romances sans paroles*, *Fêtes Galantes*, *Jadis et Naguère*, *La bonne Chanson*.

Son rôle dans l'évolution littéraire ? — Il a poursuivi, avec un sens merveilleux du rythme et de la mélodie, l'émancipation totale du vers français entreprise par l'Ecole romantique et suspendue par l'Ecole parnassienne ; il a définitivement enrichi notre poétique d'éléments nouveaux, trouvés dans les prosodies étrangères, dans les essais d'initiateurs tels que Marceline Desbordes et Van Hasselt, ou dans ses propres conceptions ; il a victorieusement revendiqué le droit absolu du poète à l'expression libre de son individualité.

2° On ne remplacera pas plus Verlaine, que Verlaine n'a remplacé Leconte de Lisle.

Pourquoi, diantre ! *la Plume* veut-elle, tout le temps, s'offrir un poète-lauréat ? C'est bon, ça, pour la reine d'Angleterre et pour le Docteur Jameson.

Jules Bois. — Depuis longtemps, Verlaine ne vivait plus dans notre admiration que par ses œuvres anciennes. Déjà *Chansons pour elle* marquait la lassitude de ses nerfs et les défaillances de son âme. Sa langue devenait pâteuse et obscure, l'inspiration le quittait. Tout de même il fut de par *Sagesse* et les *Poètes Galantes* le plus grand poète de ce temps au delà des écoles, dans la Sincérité et dans la Grâce. Je ne sais pas quelle est maintenant l'opinion des jeunes à son égard et quel successeur — puisque successeur il y a — ils lui désigneront. Pour moi, celui dont les vers se rapprochent le plus des beaux vers de Verlaine par leur souplesse fluide et cette âme froissée par la vie jusques aux cimes du rêve, c'est Georges Rodenbach. Il joint à son haut talent l'attitude impeccable de l'honnête homme. Car j'avoue qu'il a fallu l'excuse irréductible du génie pour que nous nous obstinions à admirer jusqu'à sa dernière heure celui qui, après avoir été le sanglot et le sourire de l'humanité, devint le pèlerin bégayeur des hôpitaux et des bars.

Joseph Bouchard. — Permettez-moi de garder le silence sur les deux questions que vous voulez bien me poser.

Il en est assez qui vous diront ce que Verlaine a écrit de meilleur, qui vous définiront son rôle et vous désigneront son successeur.

Dans la circonstance, les jugements les plus sûrs peuvent aisément se tromper, et j'estime que le mien est de me taire.


200 LETTERATI CONTEMPORANEI

M. Denis — Illustrazione per " *Sagesse* ..."

cesa fantasia, trasformante ogni amoretto in passione febbrile, ogni bisogno di preghiera in delirio mistico.

Data tale indole affatto istintiva e così ribelle ad ogni vincolo sociale, non è da sorprendersi che, dopo circa sei anni di esaltazione religiosa e di vita illibata, il Verlaine fatalmente ripiombasse negli antichi peccati, lasciandosi di nuovo attrarre dalle tentazioni carnali, pur ritornando, tratto tratto, ad inginocchiarsi compunto e pentito dinanzi agli altari per implorare la pietà divina.

E, con la spontanea schiettezza con cui si confessava e svelava ogni più recondita piega della sua coscienza, con quell'infallicabile sincerità che forma uno dei maggiori fascini dei suoi scritti, egli, in una caratteristica serie di piccole raccolte poetiche, ci ha mostrato, volta a volta, i due così opposti aspetti della sua nuova vita, in cui ai più virtuosi propositi alternavansi le cadute nei fossati della lussuria, meritando così l'appellativo di *homo duplex* attribuitogli da Anatole France, giustificando così l'affermazione del Retté che in lui vi fossero due anime, quella di un asceta e quella d'un satiro, in continua lotta fra di esse.

Questo parallelismo di opere di carattere così opposto è stato assai ingegnosamente giustificato dal poeta stesso in due pagine di un suo volume di prosa, che verranno certo lette con interesse; eccole:

" Il est certain que le poète doit, comme tout " artiste, après l'intensité, condition héroïque " indispensable, chercher l'unité. L'unité de ton " (qui n'est pas la monotonie) un style recon- " naisable à tel endroit de son œuvre pris in- " différemment, des habitudes, des attitudes, " l'unité de pensée aussi, et c'est ici qu'un débat " pourrait s'engager. Au lieu d'abstractions, " nous allons tout simplement prendre notre " poète comme champ de dispute. Son œuvre " se tranche, à partir de 1880, en deux portions " bien distinctes et le prospectus de ses livres " futurs indique qu'il y a chez lui parti pris de " continuer ce système et de publier, sinon si- " multanément, du moins parallèlement des ou- " vrages d'une absolue différence d'idées, — " pour bien préciser, des livres où le catholi- " cisme déploie sa logique et ses illécébrances, " ses blandices et ses terreurs, et d'autres pu- " rement mondains: sensuels avec une affli- " geante belle humeur et pleins de l'orgueil de " la vie. Que devient dans tout ceci, dira-t-on, " l'unité de pensée preconisée?

" Mais elle y est! Elle y est au titre humain, " au titre catholique, ce qui est la même chose " à nos yeux. Je crois, et je pèche par pensée " comme par action; je crois, et je me repens par " pensée en attendant mieux. Ou bien encore, je " crois et je suis bon chrétien en ce moment; je " crois et je suis mauvais chrétien l'instant d'après. " Le souvenir, l'espoir, l'invocation d'un péché " me délectent avec ou sans remords, quel- " quefois sous la forme même et muni de toutes " les conséquences du Péché, plus souvent, tant " la chair et le sang sont forts, — naturels et " *animals*, tels les souvenirs, espoirs et invo- " cations du beau premier libre-penseur. Cette " délectation, moi, vous, lui, écrivains, il nous " plaît de la coucher sur le papier et de la " publier plus ou moins bien ou mal exprimée; " nous la consignons enfin dans la forme lit- " téraire, oubliant toutes idées religieuses ou " n'en perdant pas une de vue. De bonne foi " nous condamnera-t-on comme poète ? Cent " fois non. Que la conscience du catholique " raisonne autre- " ment ou non, " ceci ne nous re- " garde pas.

" Maintenant, " les vers catholi- " ques de Pauvre " L é l i a n cou- " vrent-ils littérairement ses " autres vers? Cent fois oui. " Le ton est le même dans les " deux cas, grave et simple " ici, là fioriture, languide, " enervé, rieur et tout; mais " le même ton partout, com- " me l' *Homme* mystique et " sensuel reste l'homme in- " tellectuel toujours dans les " manifestations diverses d'u- " ne même pensée qui a ses

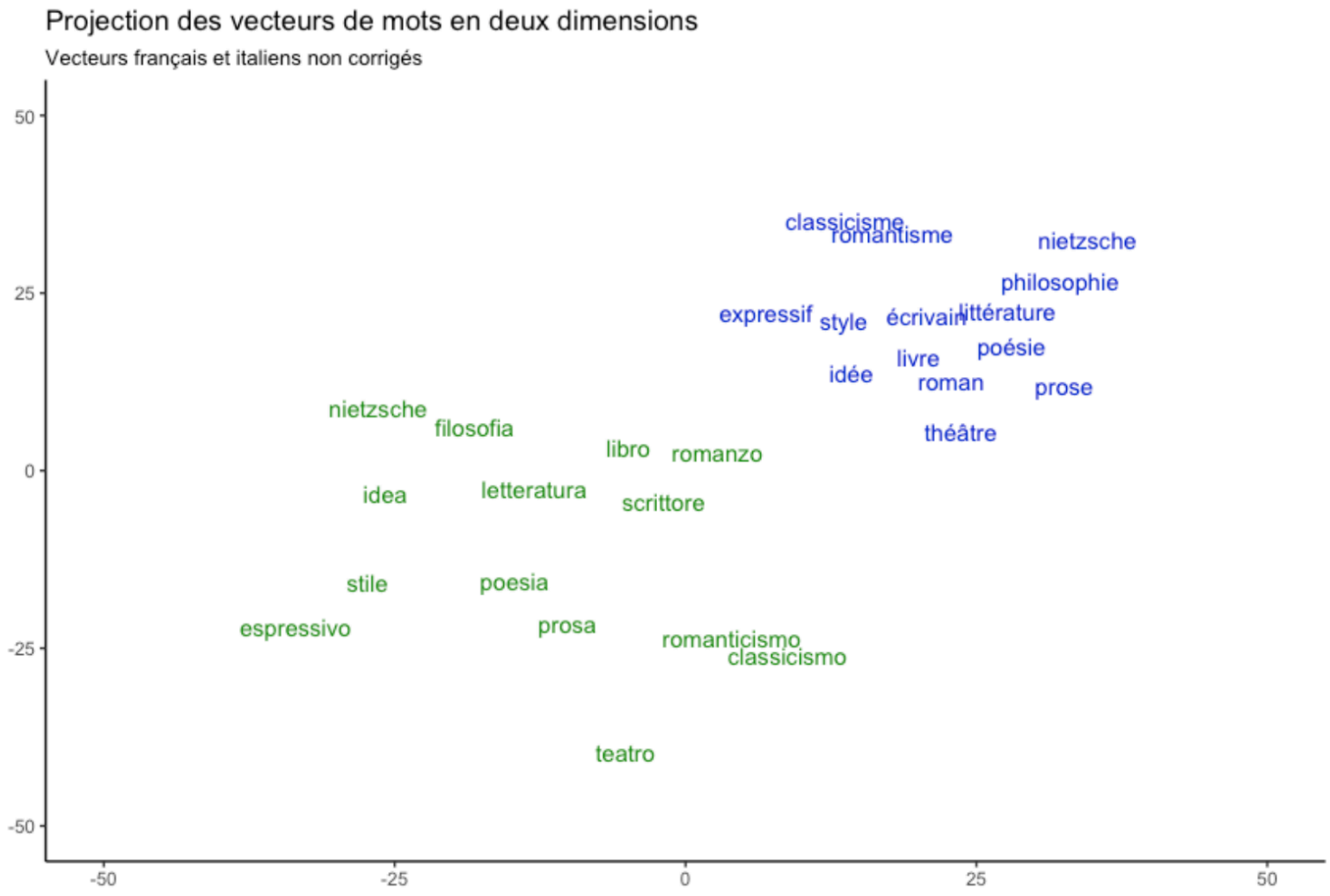
Disegno di Steinlen per *Romances sans paroles* (*Gil Blas illustré*).

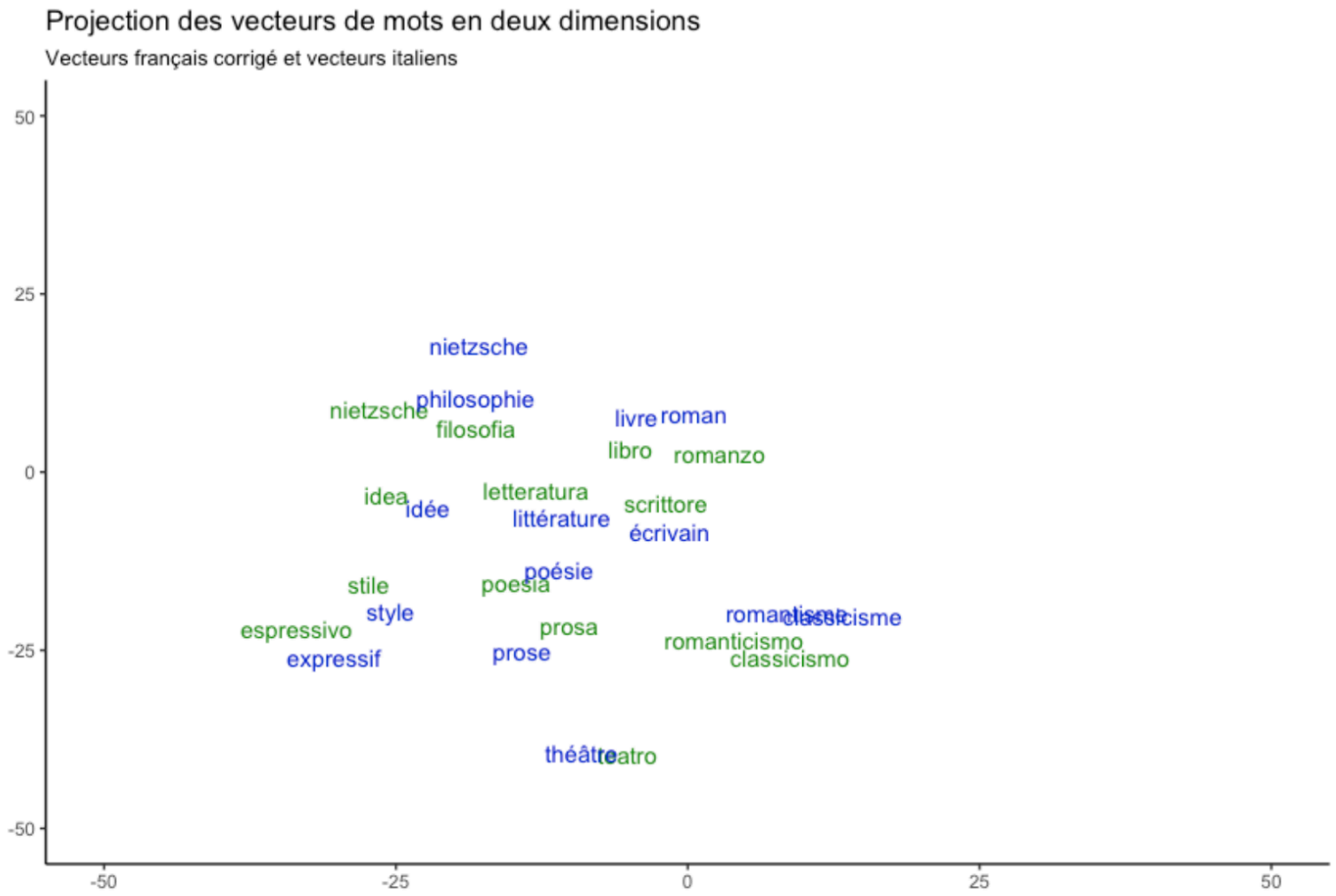
Virality beyond one language

Currently all the virality analysis we lead are limited to French-speaking press, but viral networks are clearly not limited to one langage, especially not in locations with a high level of cross-linguistic contact like Canada or Switzerland.

The key difficulty is methodological: most text mining approaches are based on lexical forms. Yet, between two languages most concepts will be expressed using different forms and, most of the time, they would not even exactly match.

An emerging techniques, word embeddings may partly circumvent this issue, by encoding all the neighbor relationships of each word. Since word embeddings in different languages will likely produce a similar network of relationships, a proper alignment may suffice to bring different languages to the same semantic ground.


Projection des vecteurs de mots en deux dimensions
Vecteurs français et italiens non corrigés
50
25
0
-25
-50
-50
-25
0
25
50
classicisme
romantisme
nietzsche
philosophie
expressif
style
écrivain
littérature
poésie
livre
idée
roman
prose
théâtre
nietzsche
filosofia
libro
romanzo
idea
letteratura
scrittore
stile
poesia
espressivo
prosa
romanticismo
classicismo
teatro


Projection des vecteurs de mots en deux dimensions
Vecteurs français corrigé et vecteurs italiens
50
25
0
-25
-50
-50
-25
0
25
50
nietzsche
nietzsche
philosophie
filosofia
livre
roman
libro
romanzo
idea
idée
letteratura
littérature
scrittore
écrivain
poésie
poesia
stile
style
espressivo
expressif
prosa
prose
romanticisme
romanticismo
classicisme
classicismo
théâtre
teatro

Virality beyond one language

```
> wordVectors::nearest_to(italian_vectors, french_vectors_corrected[["écrivain"]], 30)
   scrittore     letterato    romanziere       artista     prosatore    novelliere         poeta        critico
   0.2621570     0.3841533     0.3859421     0.3908917     0.3972127     0.4074836     0.4565549     0.4661891
 drammaturgo     pensatore  illustratore         opera     sociologo        merito      umorista contemporaneo
   0.4715974     0.4720035     0.4786143     0.4867049     0.4892099     0.4922769     0.5020004     0.5022718
    oltralpe spregiudicato       pittore      filosofo       ingegno    apprezzato   descrittore   divulgatore
   0.5041421     0.5042614     0.5084535     0.5108349     0.5113975     0.5132561     0.5147991     0.5170232
   florilegio    anconitano  diciottesimo  intervistare      tradotto         colto
   0.5207816     0.5213351     0.5219895     0.5220205     0.5253177     0.5259511
```

```
> wordVectors::nearest_to(italian_vectors, french_vectors_corrected[["revue"]], 30)
    rassegna       rivista       mensile    quindicinale    letteraria         giugno    pubblicare     fascicolo
   0.3976879     0.4218748     0.4366481     0.4576961     0.4646905     0.4724801     0.4768679     0.4778350
         apr   preussische   settimanale      periodico pubblicazione       sociali      articolo         maggio
   0.4809541     0.4913233     0.4923484     0.4925884     0.4955149     0.4974714     0.5004570     0.5068495
  ebdomadario         barth         casini       losanna     settembre      rundschau         revue  intervistare
   0.5080657     0.5089233     0.5096743     0.5136988     0.5171027     0.5180243     0.5214868     0.5261398
        ediz           amy       vitalia         marzo        aprile         arie
   0.5300317     0.5301265     0.5304097     0.5304164     0.5363001     0.5386583
```

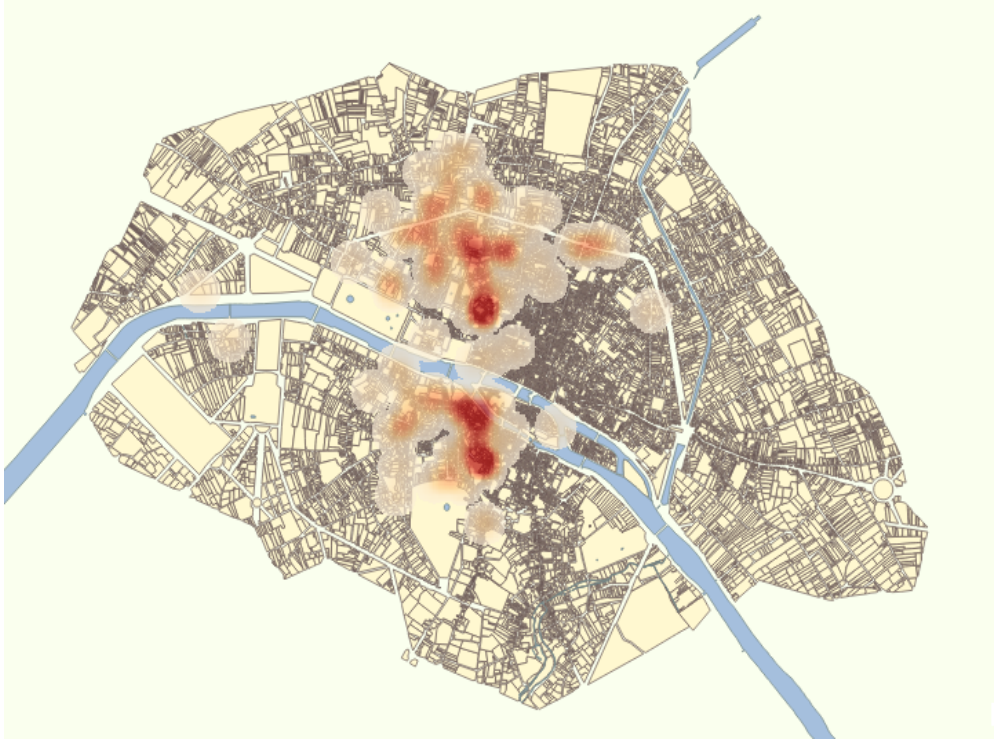
Une utilité immédiate : repérer dans le corpus des mots qui sont appelés de la même manière qu'un concept français (avec des associations trop contextuelles pour figurer dans un dictionnaire)

Virality beyond newspapers: the need for a map.

Networks of reprints raised a whole new series of unresolved issues:

- What are the determining technical factors of circulation in a given context: steamers, telegraphs, trains, horses, pneumatics?
- How to account for the recurrent textual flow between some newspapers? Is this only the incidental result of informal scissors-and-paste? Or, is this the consequence of explicit partnerships?
- How reprinting happens? What pieces of the copied content is kept, permuted, rewritten, joined with other texts? All of this would require innovative tools to deal with automated philological analysis.

Virality beyond newspapers: the need for a map.



5. The Numapresse database…

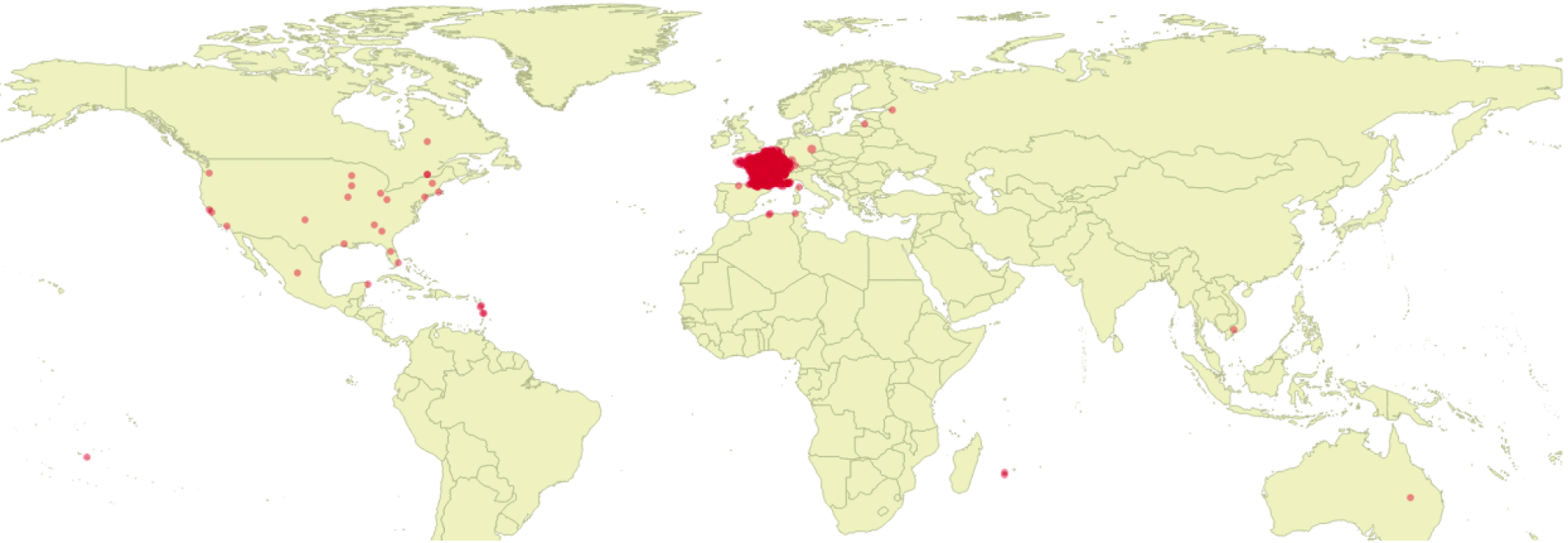
A bridge between the social and textual nature of newspapers?

Building on bibliographic data

The Numapresse database is currently developed by Olivier Lapointe. It initially aims to connect several distinct databases on French media and journalists.

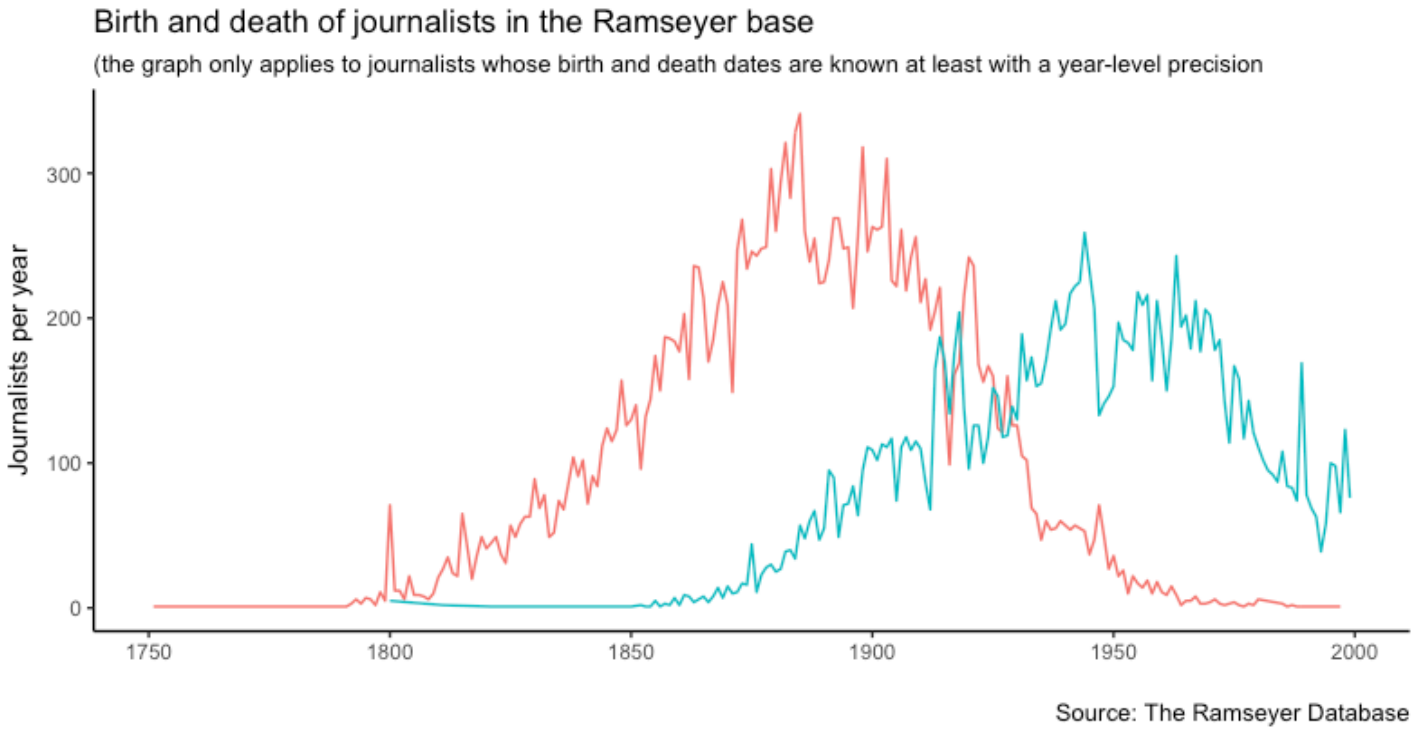
- **The Ramseyer Database**: an unpublished collection of 50 000 pseudonyms used in the French press by Patrick Ramseyer (BNF).
- **Data BNF**: the leading bibliographic database in France. About 40,000 author records (from 2 millions) are explicitly linked to a journalists activities (for instance “journaliste” in the short bio).
- Specialized databases on the history of French news, such as ***Petite Presse*** or ***Medias19***.

Building on bibliographic data



The French-speaking press in 1866

Building on bibliographic data



From a journalist index to a global repository

While initially conceived as a metadata-only service, the Numapresse database is starting to integrate text as well.

- **Metadata is *in* the text**: signatures, reprints and automated parsing of biographical information can extend the database beyond the focus of the existing sources (especially to avoid losing Guibert)
- **The text is significantly enriched**: OCR errors aside, the progress of text mining techniques would probably allow at some point to generate a TEI-like format for each article.
- **A lot of text can be put there**. Hosting by Humanum lifts a lot of concerns over size limitations.

The Numapresse database

In its latest stage, the database host a wide array of interconnect objects (actors, events, places…) and texts (productions, ressources)

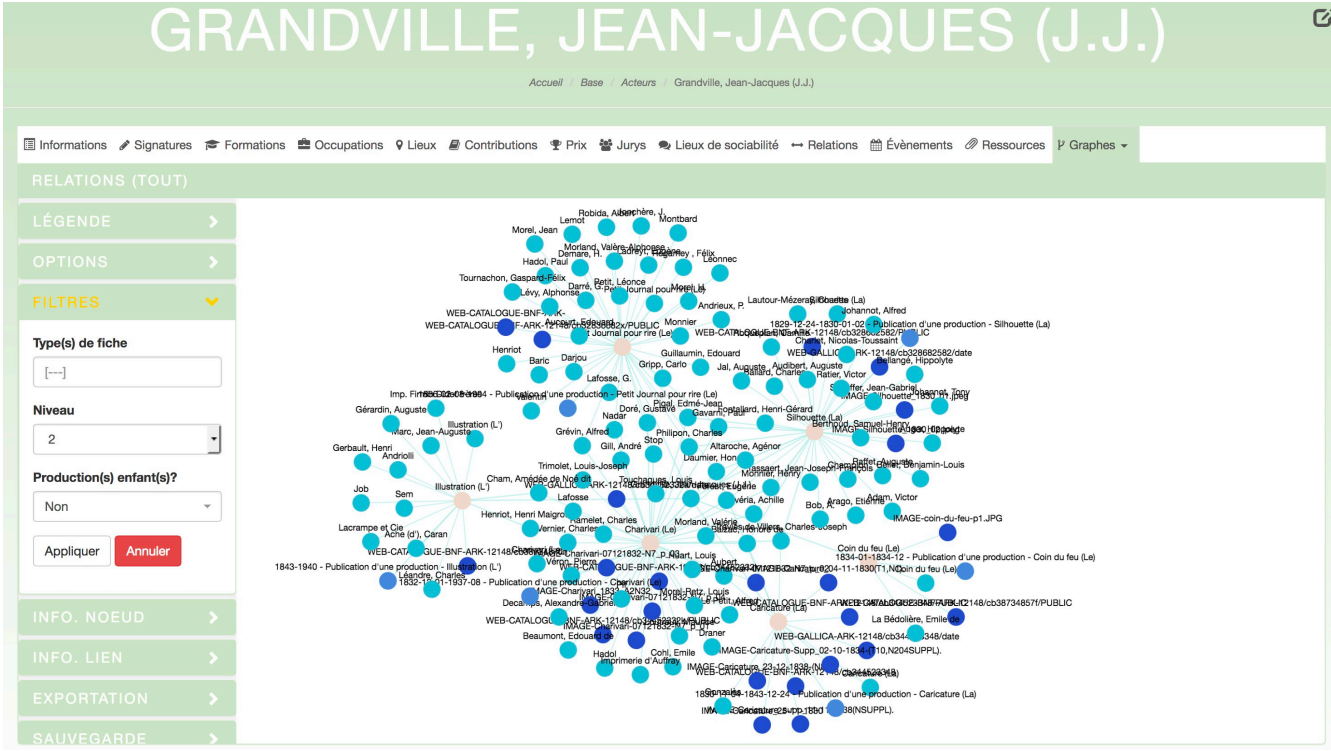
The Numapresse database

In its latest stage, the database host a wide array of interconnect objects (actors, events, places…) and texts (productions, ressources)

Début	Fin	Lieu	Secteur	Sous-secteur	Occupation	Occupation - Détail	Organisme(s)	Actions
AAAA-MM-J.	AAAA-MM-							
[---]	[---]	[---]	Médias	Journaux	Associé	[---]	• Silhouette (La) (Organisme)	
3 oct. 1829	19 janv. 1831	[---]	Médias	Journaux	Directeur	[---]	• Mode (La) (Organisme)	
oct. 1831	1848	[---]	Médias	Journaux	Directeur	[---]	• Journal des connaissances utiles (Organisme)	
3 oct. 1833	juin 1836	[---]	Médias	Journaux	Directeur	[---]	• Musée des familles (Organisme)	
juin 1836	1866	[---]	Médias	Journaux	Directeur de publication	[---]	• Presse (La) (Organisme)	

The Numapresse database

Since the database is bound to include a wide array of interconnections, Olivier Lapointe has created dataviz views, such as networks, that ease significantly the identification of collaborations or sociability links

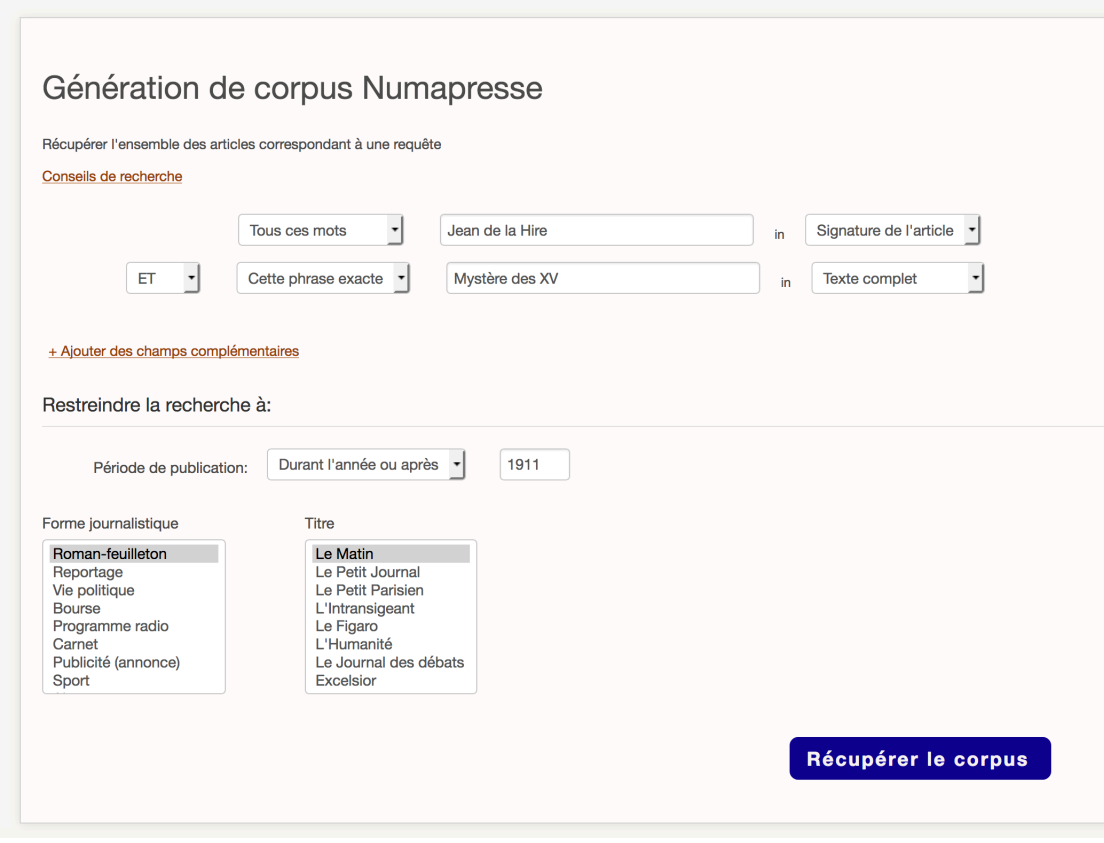


The Numapresse database

Thanks to an extensive collaborative work, the *Chat noir* has served as an experimental test to integrate the complete text metadata of a periodic.

The Numapresse database

All the documents enriched by Numapresse could also serve to generate well-defined corpora — such as all the serial novels published by Jean de la Hire in le *Matin* in 1911



Conclusion