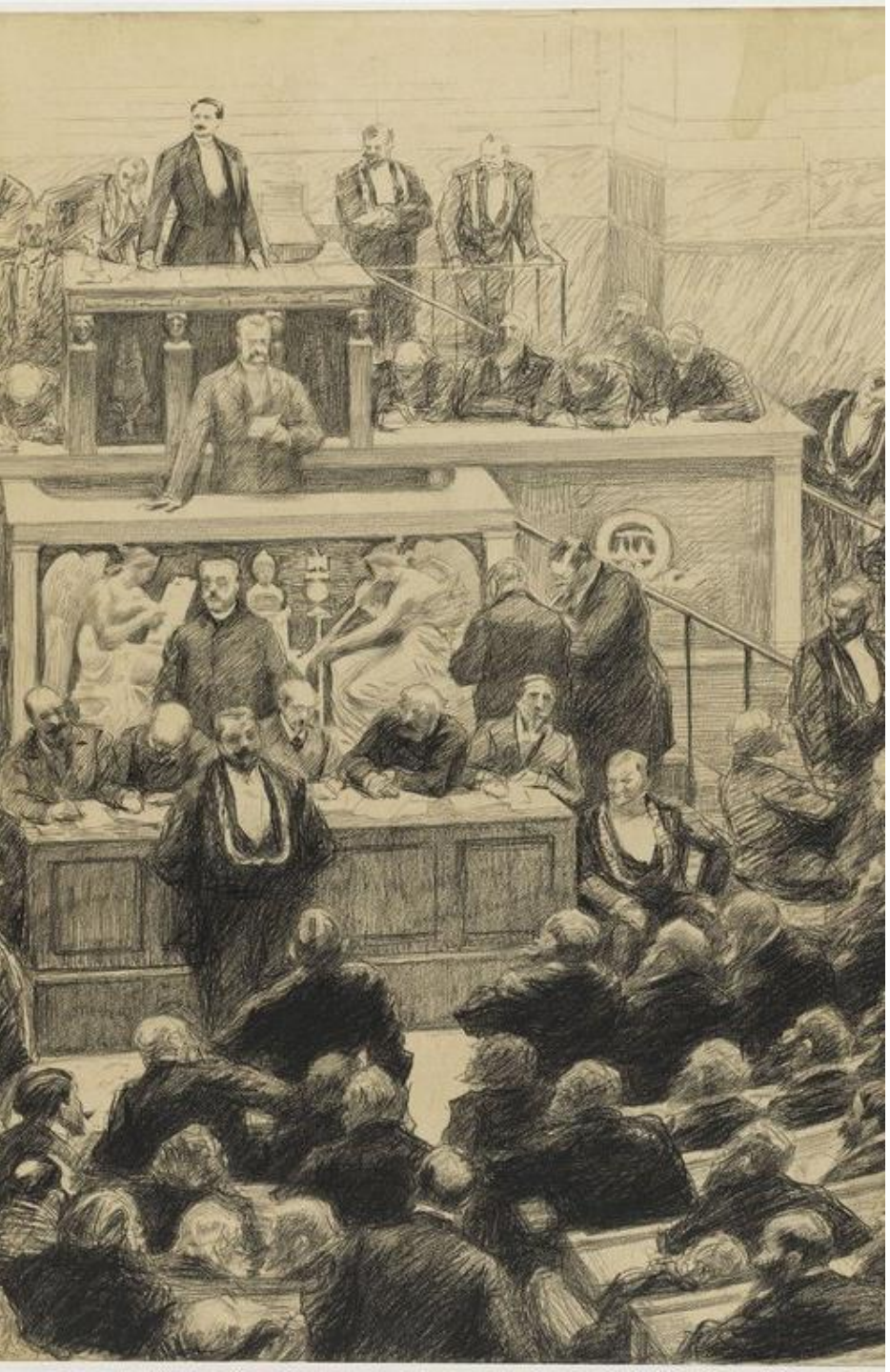


Charles Paul Renouard
Les Présidents Deschanel et Waldeck-Rousseau à la Chambre des Députés
Musée d'Orsay
Sans date (entre 1898 et 1902 ?)

AGODA

ANALYSE SÉMANTIQUE ET **G**RAPHES RELATIONNELS POUR L'**O**UVERTURE ET L'ÉTUDE DES **D**ÉBATS À L'**A**SSEMBLÉE NATIONALE

Marie Puren (Epitech)
Pierre Vernus (LARHRA)

Source : BnF, https://gallica.bnf.fr/html/und/droit-economie/iiie-republique-0?mode=desktop

OBJECTIFS

- Donner accès aux débats parlementaires
- Faciliter la recherche dans le corpus
- Offrir de nouveaux modes de visualisation des documents

OBJECTIFS

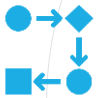
Projet en mode « preuve de concept » sur une partie réduite du corpus (1889-1893)

Exploration

Créer une plateforme de consultation des **débats parlementaires à la Chambre des députés** (1881-1940), retranscrits dans le *Journal officiel de la République française. Débats parlementaires. Chambre des députés : compte rendu in-extenso*, et disponibles sur Gallica

Annotation

Produire des données textuelles structurées et sémantiquement enrichies à partir de ces débats numérisés



Réutilisation

Contribuer à la conception d'un *workflow* adapté à l'analyse de gros corpus de documents historiques (initié par les travaux réalisés au cours de l'ANR TIME US 2018-2021)

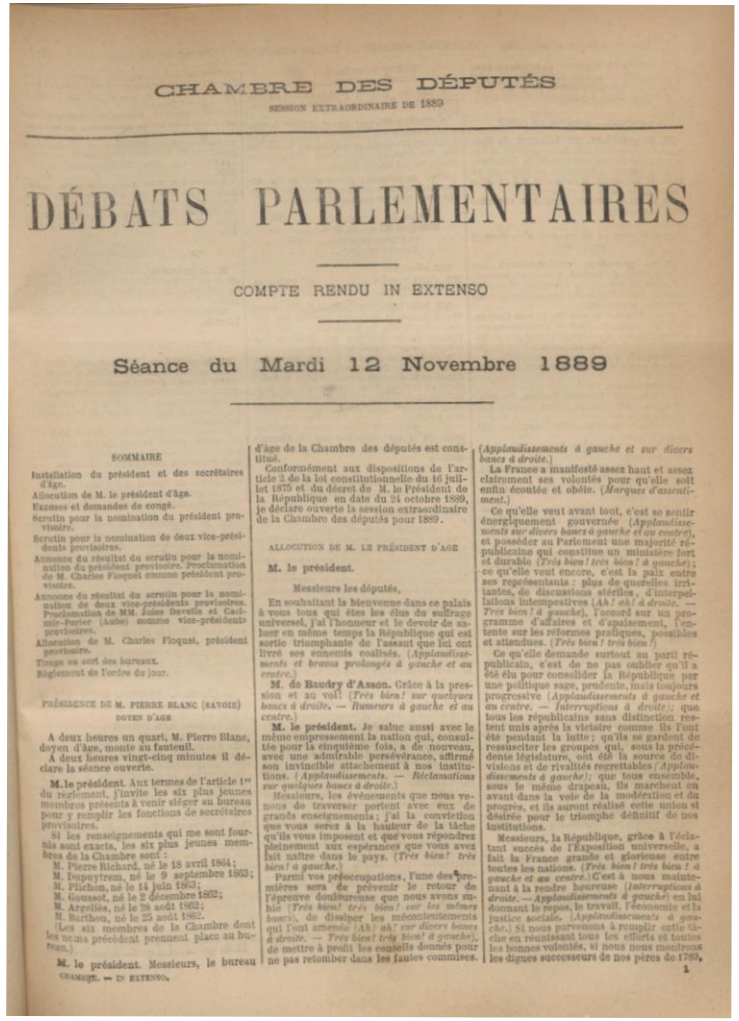
COMPOSITION DE L'EQUIPE

Pierre Vernus – LARHRA

Marie Puren – Epitech (MNSHS)

Nicolas Bourgeois – Epitech (MNSHS)

Eric de La Clergerie – Inria (ALMAnaCH)



CHAMBRE DES DÉPUTÉS

SESSION EXTRAORDINAIRE DE 1889

DÉBATS PARLEMENTAIRES

COMPTE RENDU IN EXTENSO

Séance du Mardi 12 Novembre 1889

SOMMAIRE

Installation du président et des secrétaires d'âge.
Allocution de M. le président d'âge.
Excuses et demandes de congé.
Scrutin pour la nomination du président provisoire.
Scrutin pour la nomination de deux vice-présidents provisoires.
Annonce du résultat du scrutin pour la nomination du président provisoire. Proclamation de M. Charles Floquet comme président provisoire.
Annonce du résultat du scrutin pour la nomination de deux vice-présidents provisoires. Proclamation de MM. Jules Develle et Casimir-Perier (Aube) comme vice-présidents provisoires.
Allocution de M. Charles Floquet, président provisoire.
Tirage au sort des bureaux.
Règlement de l'ordre du jour.

PRÉSIDENCE DE M. PIERRE BLANC (SAVOIE)
DOYEN D'AGE

A deux heures un quart, M. Pierre Blanc, doyen d'âge, monte au fauteuil.

A deux heures vingt-cinq minutes il déclare la séance ouverte.

M. le président. Aux termes de l'article 1er du règlement, j'invite les six plus jeunes membres présents à venir siéger au bureau pour y remplir les fonctions de secrétaires provisoires.

Si les renseignements qui me sont fournis sont exacts, les six plus jeunes membres de la Chambre sont :

M. Pierre Richard, né le 18 avril 1864 ;
M. Dupuytren, né le 9 septembre 1863 ;
M. Plichon, né le 14 juin 1863 ;
M. Gousset, né le 2 décembre 1862 ;
M. Argelies, né le 28 août 1862 ;
M. Barthou, né le 25 août 1862.

(Les six membres de la Chambre dont les noms précèdent prennent place au bureau.)

M. le président. Messieurs, le bureau d'âge de la Chambre des députés est constitué.

Conformément aux dispositions de l'article 3 de la loi constitutionnelle du 16 juillet 1875 et du décret de M. le Président de la République en date du 24 octobre 1889, je déclare ouverte la session extraordinaire de la Chambre des députés pour 1889.

ALLOCUTION DE M. LE PRÉSIDENT D'AGE

M. le président.

Messieurs les députés,

En souhaitant la bienvenue dans ce palais à vous tous qui êtes les élus du suffrage universel, j'ai l'honneur et le devoir de saluer en même temps la République qui est sortie triomphante de l'assaut que lui ont livré ses ennemis coalisés. (*Applaudissements et bravos prolongés à gauche et au centre.*)

M. de Baudry d'Asson. Grâce à la pression et au vol! (*Très bien! sur quelques bancs à droite. — Rumeurs à gauche et au centre.*)

M. le président. Je salue aussi avec le même empressement la nation qui, consultée pour la cinquième fois, a de nouveau, avec une admirable persévérance, affirmé son invincible attachement à nos institutions. (*Applaudissements. — Réclamations sur quelques bancs à droite.*)

Messieurs, les événements que nous venons de traverser portent avec eux de grands enseignements ; j'ai la conviction que vous serez à la hauteur de la tâche qu'ils vous imposent et que vous répondrez pleinement aux espérances que vous avez fait naître dans le pays. (*Très bien! très bien! à gauche.*)

Parmi vos préoccupations, l'une des premières sera de prévenir le retour de l'épreuve douloureuse que nous avons subie (*Très bien! très bien! sur les mêmes bancs*), de dissiper les mécontentements qui l'ont amenée (*Ah! ah! sur divers bancs à droite. — Très bien! très bien! à gauche*), de mettre à profit les conseils donnés pour ne pas retomber dans les fautes commises. (*Applaudissements à gauche et sur divers bancs à droite.*)

La France a manifesté assez haut et assez clairement ses volontés pour qu'elle soit enfin écoutée et obéie. (*Marques d'assentiment.*)

Ce qu'elle veut avant tout, c'est se sentir énergiquement gouvernée (*Applaudissements sur divers bancs à gauche et au centre*), et posséder au Parlement une majorité républicaine qui constitue un ministère fort et durable (*Très bien! très bien! à gauche*) ; ce qu'elle veut encore, c'est la paix entre ses représentants : plus de querelles irritantes, de discussions stériles, d'interpellations intempestives (*Ah! ah! à droite. — Très bien! à gauche*), l'accord sur un programme d'affaires et d'apaisement, l'entente sur les réformes pratiques, possibles et attendues. (*Très bien! très bien!*)

Ce qu'elle demande surtout au parti républicain, c'est de ne pas oublier qu'il a été élu pour consolider la République par une politique sage, prudente, mais toujours progressive (*Applaudissements à gauche et au centre. — Interruptions à droite*) ; que tous les républicains sans distinction restent unis après la victoire comme ils l'ont été pendant la lutte ; qu'ils se gardent de ressusciter les groupes qui, sous la précédente législature, ont été la source de divisions et de rivalités regrettables (*Applaudissements à gauche*) ; que tous ensemble, sous le même drapeau, ils marchent en avant dans la voie de la modération et du progrès, et ils auront réalisé cette union si désirée pour le triomphe définitif de nos institutions.

Messieurs, la République, grâce à l'éclatant succès de l'Exposition universelle, a fait la France grande et glorieuse entre toutes les nations. (*Très bien! très bien! à gauche et au centre.*) C'est à nous maintenant à la rendre heureuse (*Interruptions à droite. — Applaudissements à gauche*) en lui donnant le repos, le travail, l'économie et la justice sociale. (*Applaudissements à gauche.*) Si nous parvenons à remplir cette tâche en réunissant tous les efforts et toutes les bonnes volontés, si nous nous montrons les dignes successeurs de nos pères de 1789,

1

https://gallica.bnf.fr/ark:/12148/cb328020951/date

LES DÉBATS PARLEMENTAIRES

- Disponibles sur Gallica : images + textes océrisés
- Source intéressante pour :
  - Histoire
  - Science politique
  - Sociologie
  - Linguistique …
- Mise en ligne du Hansard (retranscription des débats parlementaires des gouvernements de type Westminster) a favorisé de nouvelles recherches

VÈME LÉGISLATURE (1889-1893)

Général Boulanger - KU Leuven, Belgium - Public Domain.
https://www.europeana.eu/fr/item/2024903/photography_ProvidedCHO_KU_Leuven_9983157830101488

Bibliothèque nationale de France
https://gallica.bnf.fr/ark:/12148/bpt6k1206804x/f1

Wikimedia Commons
https://commons.wikimedia.org/wiki/File:ActionPanama.JPG

- Renouvellement partiel du personnel politique (boulangisme et scandale de Panama)
- Premières manifestations du Ralliement des catholiques à la République
- Tournant de la politique douanière (lois Méline)
- Essor du socialisme et du syndicalisme (Fourmies)
- Premiers attentats anarchistes
- …

WORKFLOW

- Extraire le texte / ré-océriser
- Annoter et enrichir sémantiquement le corpus
- Publier un corpus éditorialisé
- Modéliser les sujets

Icone conçue par Freepik

EXTRAIRE LE TEXTE / RÉ-OCÉRISER

- Récupération des textes océrisés via API Document de Gallica
  - Erreurs dues à la courbure de la page au niveau de la reliure
- Deux options pour ré-océriser :
  - Utiliser une solution propriétaire : Abby Fine Reader (excellentes performances sur les corpus de presse)
  - Entraîner un modèle avec une solution libre (Tesseract, OCR4all, Kraken)
- Post-corrections :
  - A la main
  - Emploi d'expressions régulières et librairie Python pyspellchecker

ANNOTER ET ENRICHIR

Annotation des fichiers transcrits en XML-TEI (Text Encoding Inititative)

Choix des annotations

- Sections très formalisées
- Entités nommées (Personnes)
- Entités et segments liées aux mouvements, partis et idées politiques
- Perspective « données liées » : identification des annotations avec des URI fournies par data.bnf et Wikidata

Création d'un schéma TEI

- ODD (One Document Does it all)
- Adapté aux débats parlementaires
- Tenant compte des annotations choisies

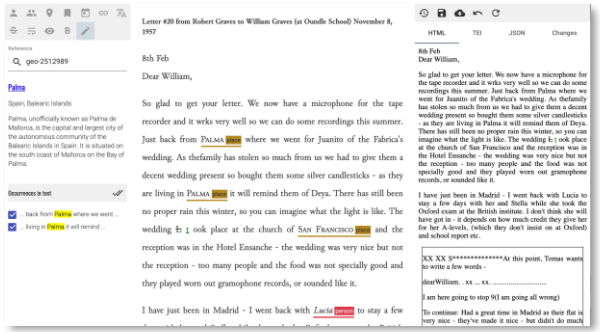
Phase d'annotation

- Scripts Python (par ex. script LSE-OD2M : https://github.com/TimeUs-ANR/LSE-OD2M)
- Annotations d'abord réalisées à la main = données d'entraînement pour l'automatisation de la tâche par l'analyseur syntaxique FrMG
- Annotation des entités nommées avec FastText embeddings et le modèle linguistique contextuel neuronal français CamemBERT dans une architecture LSTM-CRF

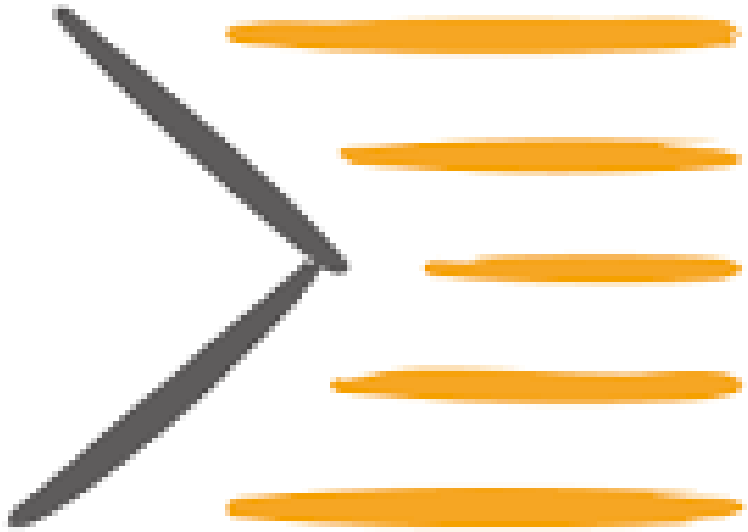
UN CORPUS ÉDITORIALISÉ

Corpus encodé en TEI et stocké dans une base de données XML eXist-db

TEI Publisher : transformer les données stockées dans la base de données en pages Web HTML pour leur publication

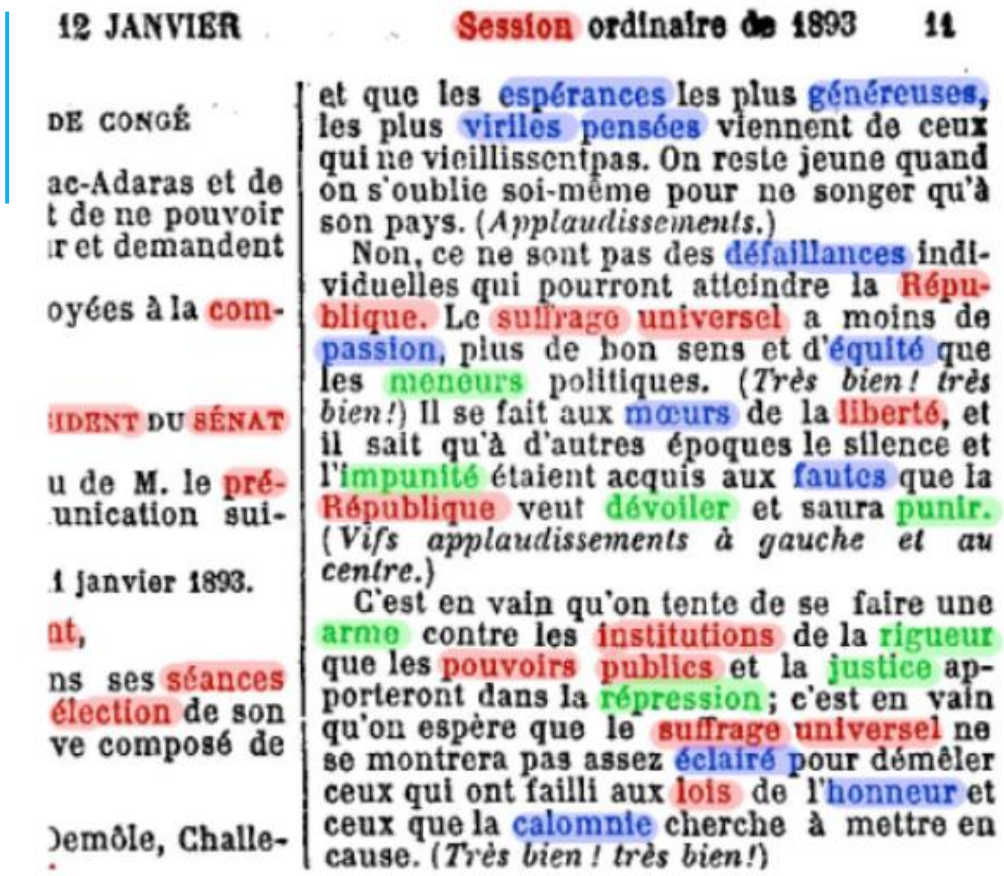


Accès à un corpus éditorialisé, intégré dans un contexte d'application avec navigation, recherche plein-texte et affichage des fac-similés



PUBLIER UN CORPUS ÉDITORIALISÉ

- Technologie *Web Components* → possibilité de développer de nouvelles fonctionnalité
  - Accès aux annotations linguistiques
  - Affichage des entités nommées
  - Exploration avec la modélisation de sujets



12 JANVIER Session ordinaire de 1893 11

DE CONGÉ

ac-Adaras et de
t de ne pouvoir
ır et demandent

oyées à la com-

IDENT DU SÉNAT

u de M. le pré-
unication sui-

1 janvier 1893.

nt,

ns ses séances
élection de son
ve composé de

)emôle, Challe-

et que les espérances les plus généreuses, les plus viriles pensées viennent de ceux qui ne vieillissentpas. On reste jeune quand on s'oublie soi-même pour ne songer qu'à son pays. (*Applaudissements.*)

Non, ce ne sont pas des défaillances individuelles qui pourront atteindre la République. Le suffrage universel a moins de passion, plus de bon sens et d'équité que les meneurs politiques. (*Très bien! très bien!*) Il se fait aux mœurs de la liberté, et il sait qu'à d'autres époques le silence et l'impunité étaient acquis aux fautes que la République veut dévoiler et saura punir. (*Vifs applaudissements à gauche et au centre.*)

C'est en vain qu'on tente de se faire une arme contre les institutions de la rigueur que les pouvoirs publics et la justice apporteront dans la répression; c'est en vain qu'on espère que le suffrage universel ne se montrera pas assez éclairé pour démêler ceux qui ont failli aux lois de l'honneur et ceux que la calomnie cherche à mettre en cause. (*Très bien! très bien!*)

Exemple de détection automatique de topics dans un rapport de l'assemblée (CC-BY-SA Nicolas Bourgeois)

MODÉLISER LES SUJETS

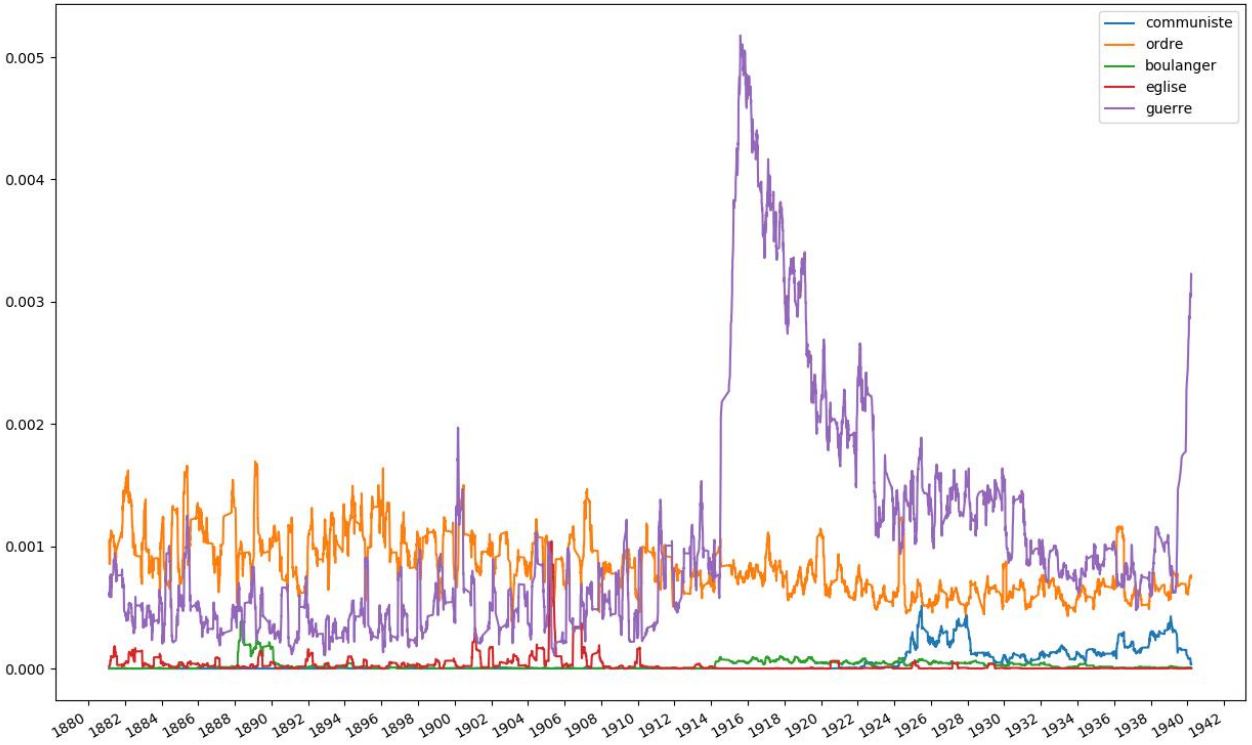
- Modélisation de sujets ou *topic modeling* : une méthode d'apprentissage non supervisée qui permet de découvrir les structures sémantiques latentes d'un corpus de textes, sans faire appel à des ressources sémantiques et lexicales.
- Particulièrement bien adapté pour l'étude de grands corpus, comme les corpus de presse par exemple
- Réalisé avec le Natural Language Toolkit (NLTK) et la librairie Python Gensim

Source : BnF https://gallica.bnf.fr/ark:/12148/btv1b530893305

RÉSULTATS ATTENDUS

- ✓ Une plateforme de consultation
- ✓ Le code source, les données annotées en XML-TEI accessibles
- ✓ Une documentation pour utiliser la plateforme
- ✓ De nouvelles analyses (articles scientifiques, conférences…)

PREMIERS TESTS : FRÉQUENCE DES MOTS



Evolution de la fréquence de mots choisis dans les débats parlementaires (1880-1942)

(CC-BY-SA Nicolas Bourgeois)

MERCI

marie.puren@epitech.eu pierre.vernus@msh-lse.fr

Eugène Atget
Palais Bourbons – Chambre des Députés
Bibliothèque Nationale de France
Entre 1900 et 1927 (d'après négatif 1900)