EXTRACTING AND PROVIDING ONLINE ACCESS TO ANNOTATED AND SEMANTICALLY ENRICHED HISTORICAL DATA

The AGODA project

Marie Puren, Pierre Vernus, Aurélien Pellet, Nicolas Bourgeois and Fanny Lebreton

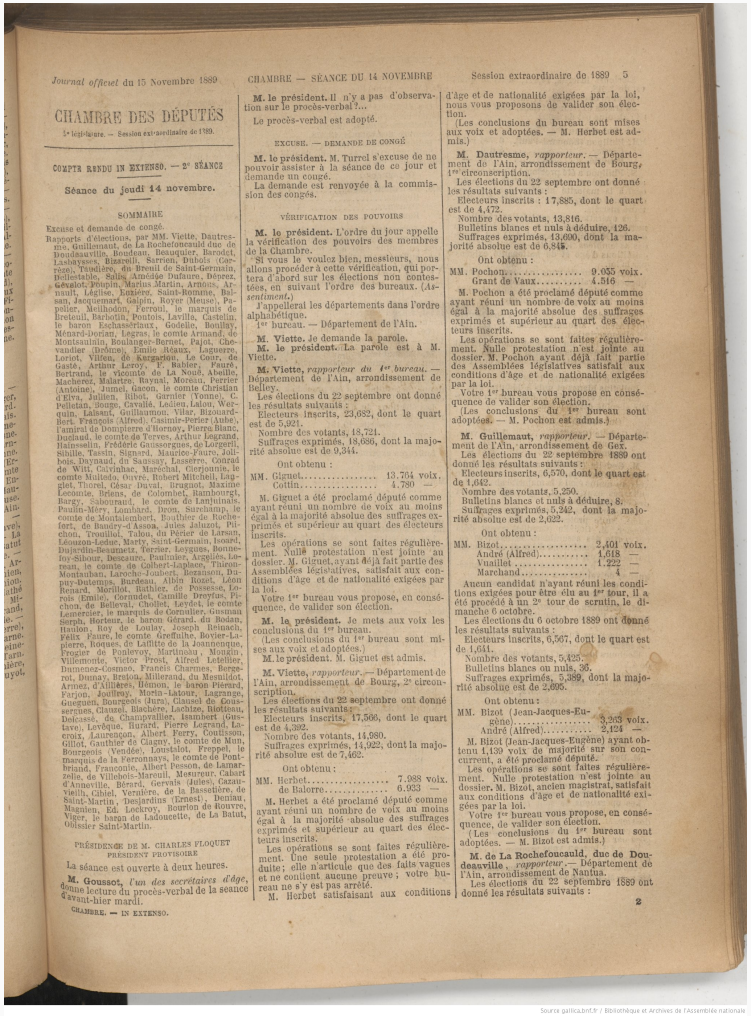
1 June 2022

DH Benelux Conference 2022, University of Luxembourg

THE AGODA PROJECT

- AGODA : **A**nalyse sémantique et **G**raphes relationnels pour l'**O**uverture et l'étude des **D**ébats à l'**A**ssemblée nationale
- Funded by the Bibliothèque nationale de France for a period of one year
- One of the 5 pilot projects supported by the DataLab
- Collaboration between Epitech (MNSHS), Inria (ALMAnaCH) and the University Lumière Lyon 2 (LARHRA)

PARLIAMENTARY DEBATES DURING THE THIRD REPUBLIC



Journal officiel du 15 Novembre 1889 — CHAMBRE — SEANCE DU 14 NOVEMBRE — Session extraordinaire de 1889 — 5

CHAMBRE DES DÉPUTÉS

1re législature. — Session extraordinaire de 1889.

COMPTE RENDU IN EXTENSO. — 2e SÉANCE

Séance du jeudi 14 novembre.

SOMMAIRE

Excuse et demande de congé.

Rapports d'élections, par MM. Viette, Dautresme, Guillemaut, de La Rochefoucauld duc de Doudeauville, Boudeau, Beauquier, Barodet, Lasbaysses, Rizzetti, Sarrien, Dubuis (Cercoe), Taudière, du Breuil de Saint-Germain, Bellestable, Salis, Amédée Dufaure, Dépres, Gévelot, Poupin, Marius Martin, Arnous, Arnault, Lagisse, Lanièrey, Saint-Romme, Bassan, Jacquemart, Galpin, Royer (Meuse), Papolier, Meilhodon, Ferroul, le marquis de Breteuil, Barbetin, Poulois, Lavilla, Castelin, le baron Eschassériaux, Godelle, Boulay, Ménard-Dorian, Legras, le comte Armand, de Montsaulnin, Boulanger-Bernet, Pajot, Chevandier (Drôme), Emile Réaux, Laguerre, Loriot, Vilton, de Kergariou, Le Cour, de Gasté, Arthur Leroy, P. Rabier, Fauré, Bertrand, le vicomte de La Nouë, Abeille, Macherez, Malartre, Raynal, Moreau, Perrier (Antoine), Junet, Gicon, le comte Christian d'Elva, Julien, Ribot, Garnier (Yonne), C. Pelletan, Bouge, Cavaillé, Ledieu-Lalon, Vercquin, Laisant, Guillaumon, Vilar, Bizouard-Bert, François (Alfred), Casimir-Perier (Aube), l'amiral de Bompierre d'Hornoy, Pierre Blanc, Duclaud, le comte de Torres, Arthur Legrand, Hainseelin, Frédéric Gaussorgues, de Lorgeril, Sibille, Tassin, Signard, Maurice-Faure, Jolibois, Duvaud, du Saussay, Lasserre, Conrad de Wit, Calvinhac, Maréchal, Clerjounie, le comte Mathieu, Ouvré, Robert Mitchell, Laglet, Thorel, César Duval, Brugnot, Maxime Lecomte, Brians, de Colombet, Rambourgt, Rargy, Sahorraud, le comte de Lamjuinais, Paulin-Méry, Lombard, Dron, Surchamp, le comte de Montalembert, Boutibez de Rochefort, de Baudry-d'Asson, Jules Jaluzot, Pilchon, Trouillot, Talou, du Périer de Larsan, Léouzon-Leduc, Marty, Saint-Germain, Isoard, Dujardin-Beaumetz, Terrier, Leygues, Bonnefoy-Sibour, Descaure, Paulmier, Argelès, Lerat, le comte de Gilbert-Lapisse, Thiéon-Montauban, Larroche-Joubert, Rezanson, Dupuy-Dutemps, Burdeau, Albin Rozet, Léon Renard, Morillot, Rathier, de Possesse, Lerois (Emile), Cormdet, Camille Dreyfus, Fichon, de Bétteval, Chollet, Leydet, le comte Lemercier, le marquis de Cornulier, Cousmaisorph, Horteur, le baron Gérard, du Bodan, Haulon, Roy de Loulay, Joseph Reinach, Félix Faure, le comte Greffulhe, Bovier-Lapierre, Roques, de Lafitte de la Joannequie, Frezier de Fontevoy, Martinaux, Mongin, Villemonte, Victor Prost, Alfred Letellier, Dumesnez-Cosmos, Francis Charmes, Bergerot, Dunay, Breton, Millerand, du Mesnildot, Arnez, d'Ulmès, Démon, le baron Pierard, Farjon, Joulleoy, Morin-Latour, Lagrange, Gueguen, Bourgeois (Jura), Clausel de Coussergues, Clauzel, Blachère, Lachize, Rotteau, Delcassé, de Champvallier, Isambert (Gustave), Levêque, Rimbert, Pierre Legrand, Lecroix, Laurençon, Albert Ferry, Coutisson, Gillot, Gauthier de Clagny, le comte de Mun, Bourgeois (Vendée), Brunstoit, Freppel, le marquis de la Ferronnays, le comte de Pontbriand, Francoule, Albert Pesson, de Lamarzelle, de Villebois-Mareuil, Mesureur, Cabart d'Anneville, Bérard, Gervais (Jules), Cazauvieilh, Chiet, Verrière, de la Bassetière, de Saint-Martin, Desjardins (Ernest), Deniau, Magnien, Ed. Lockroy, Bourlon de Rouvre, Vigor, le baron de Ladoucette, de La Batut, Odiosier Saint-Martin.

PRÉSIDENCE DE M. CHARLES FLOQUET
PRÉSIDENT PROVISOIRE

La séance est ouverte à deux heures.

M. Gousset, *l'un des secrétaires d'âge,* donne lecture du procès-verbal de la séance d'avant-hier mardi.

CHAMBRE. — IN EXTENSO.

M. le président. Il n'y a pas d'observation sur le procès-verbal ?...

Le procès-verbal est adopté.

EXCUSE. — DEMANDE DE CONGÉ

M. le président. M. Turrel s'excuse de ne pouvoir assister à la séance de ce jour et demande un congé.

La demande est renvoyée à la commission des congés.

VÉRIFICATION DES POUVOIRS

M. le président. L'ordre du jour appelle la vérification des pouvoirs des membres de la Chambre.

Si vous le voulez bien, messieurs, nous allons procéder à cette vérification, qui portera d'abord sur les élections non contestées, en suivant l'ordre des bureaux. (*Assentiment.*)

J'appellerai les départements dans l'ordre alphabétique.

1er bureau. — Département de l'Ain.

M. Viette. Je demande la parole.

M. le président. La parole est à M. Viette.

M. Viette, *rapporteur du 1er bureau.* — Département de l'Ain, arrondissement de Belley.

Les élections du 22 septembre ont donné les résultats suivants :

Electeurs inscrits, 23,682, dont le quart est de 5,921.

Nombre des votants, 18,721.

Suffrages exprimés, 18,636, dont la majorité absolue est de 9,344.

Ont obtenu :

MM. Gignet..... 13,764 voix.
Cottin..... 4,780 —

M. Gignet a été proclamé député comme ayant réuni un nombre de voix au moins égal à la majorité absolue des suffrages exprimés et supérieur au quart des électeurs inscrits.

Les opérations se sont faites régulièrement. Nulle protestation n'est jointe au dossier. M. Gignet, ayant déjà fait partie des Assemblées législatives, satisfait aux conditions d'âge et de nationalité exigées par la loi.

Votre 1er bureau vous propose, en conséquence, de valider son élection.

M. le président. Je mets aux voix les conclusions du 1er bureau.

(Les conclusions du 1er bureau sont mises aux voix et adoptées.)

M. le président. M. Gignet est admis.

M. Viette, *rapporteur.* — Département de l'Ain, arrondissement de Bourg, 2e circonscription.

Les élections du 22 septembre ont donné les résultats suivants :

Electeurs inscrits, 17,506, dont le quart est de 4,282.

Nombre des votants, 14,980.

Suffrages exprimés, 14,922, dont la majorité absolue est de 7,462.

Ont obtenu :

MM. Herbet..... 7,988 voix.
de Balorre..... 6,933 —

M. Herbet a été proclamé député comme ayant réuni un nombre de voix au moins égal à la majorité absolue des suffrages exprimés et supérieur au quart des électeurs inscrits.

Les opérations se sont faites régulièrement. Une seule protestation a été produite; elle n'articule que des faits vagues et ne contient aucune preuve ; votre bureau ne s'y est pas arrêté.

M. Herbet satisfaisant aux conditions d'âge et de nationalité exigées par la loi, nous vous proposons de valider son élection.

(Les conclusions du bureau sont mises aux voix et adoptées. — M. Herbet est admis.)

M. Dautresme, *rapporteur.* — Département de l'Ain, arrondissement de Bourg, 1re circonscription.

Les élections du 22 septembre ont donné les résultats suivants :

Electeurs inscrits : 17,885, dont le quart est de 4,472.

Nombre des votants, 13,816.

Bulletins blancs et nuls à déduire, 426.

Suffrages exprimés, 13,390, dont la majorité absolue est de 6,845.

Ont obtenu :

MM. Pochon..... 9,055 voix.
Grant de Vaux..... 4,816 —

M. Pochon a été proclamé député comme ayant réuni un nombre de voix au moins égal à la majorité absolue des suffrages exprimés et supérieur au quart des électeurs inscrits.

Les opérations se sont faites régulièrement. Nulle protestation n'est jointe au dossier. M. Pochon ayant déjà fait partie des Assemblées législatives satisfait aux conditions d'âge et de nationalité exigées par la loi.

Votre 1er bureau vous propose en conséquence de valider son élection.

(Les conclusions du 1er bureau sont adoptées. — M. Pochon est admis.)

M. Guillemaut, *rapporteur.* — Département de l'Ain, arrondissement de Gex.

Les élections du 22 septembre 1889 ont donné les résultats suivants :

Electeurs inscrits, 6,570, dont le quart est de 1,642.

Nombre des votants, 5,250.

Bulletins blancs et nuls à déduire, 8.

Suffrages exprimés, 5,242, dont la majorité absolue est de 2,622.

Ont obtenu :

MM. Bizot..... 3,401 voix.
André (Alfred)..... 1,618 —
Vuaillet..... 1,222 —
Marchand..... 4 —

Aucun candidat n'ayant réuni les conditions exigées pour être élu au 1er tour, il a été procédé à un 2e tour de scrutin, le dimanche 6 octobre.

Les élections du 6 octobre 1889 ont donné les résultats suivants :

Electeurs inscrits, 6,567, dont le quart est de 1,641.

Nombre des votants, 5,425.

Bulletins blancs ou nuls, 36.

Suffrages exprimés, 5,389, dont la majorité absolue est de 2,695.

Ont obtenu :

MM. Bizot (Jean-Jacques-Eugène)..... 3,263 voix.
André (Alfred)..... 2,124 —

M. Bizot (Jean-Jacques-Eugène) ayant obtenu 1,139 voix de majorité sur son concurrent, a été proclamé député.

Les opérations se sont faites régulièrement. Nulle protestation n'est jointe au dossier. M. Bizot, ancien magistrat, satisfait aux conditions d'âge et de nationalité exigées par la loi.

Votre 1er bureau vous propose, en conséquence, de valider son élection.

(Les conclusions du 1er bureau sont adoptées. — M. Bizot est admis.)

M. de La Rochefoucauld, duc de Doudeauville, *rapporteur.* — Département de l'Ain, arrondissement de Nantua.

Les élections du 22 septembre 1889 ont donné les résultats suivants :

2

Source gallica.bnf.fr / Bibliothèque et Archives de l'Assemblée nationale

Figure – Parliamentary sitting - 14 Nov. 1889

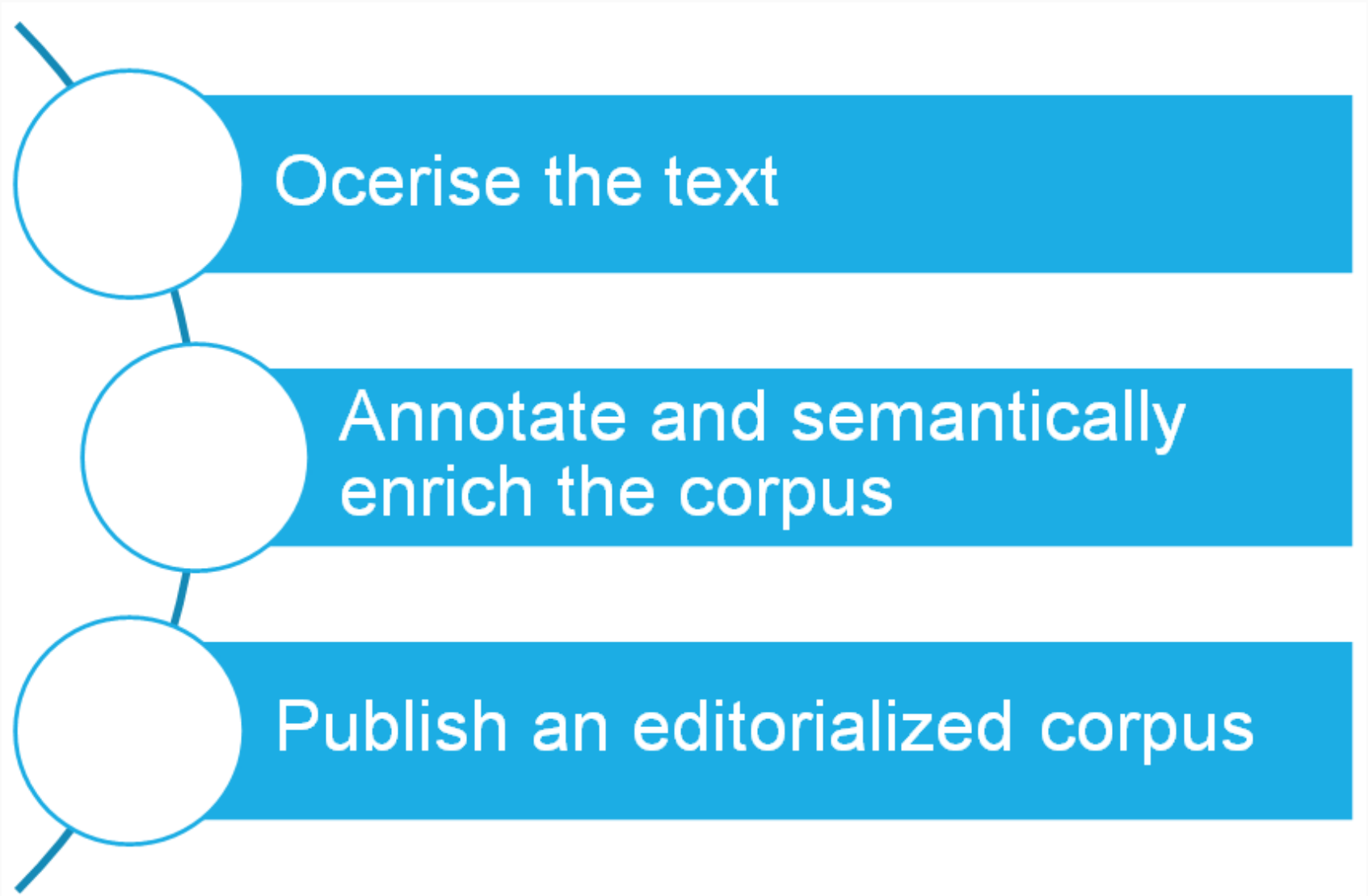
- Debates in the lower house of French Parliament ("Chambre des députés") carefully recorded in the Journal officiel de la République française. Débats parlementaires (1881-1940)
- Available online on the digital library of the National Library of France (Gallica)
- Still difficult to work on this corpus : better if you already know what you're looking for !

A PROOF OF CONCEPT

Working on a sub-part of the corpus : legislature 1889-1893, i.e. **10418 images to be processed**

- Give easier access to ancient transcriptions of parliamentary debates
- Facilitate research in this corpus
- Offer new ways of visualizing the documents

WORKFLOW


Ocerise the text
Annotate and semantically enrich the corpus
Publish an editorialized corpus

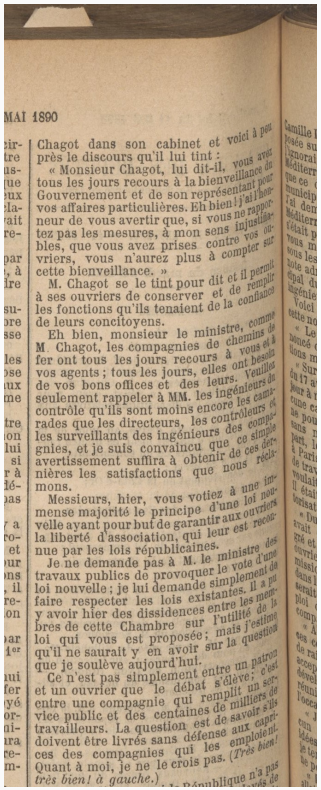
OCERISE THE DEBATES

THE CASE OF PARLIAMENTARY DEBATES

- Retrieval of ocerised texts via Gallica's API Document => uneven quality of the OCR
- Errors due to :
  - Document quality : stains and overprinting
  - **Curvature of the page** at the level of the binding

Figure – Quality assessment of the OCR provided by Gallica

EFFECT OF CURVATURE ON OCR RESULTS



MAI 1890

Chagot dans son cabinet et voici à peu près le discours qu'il lui tint :

« Monsieur Chagot, lui dit-il, vous avez tous les jours recours à la bienveillance du Gouvernement et de son représentant pour vos affaires particulières. Eh bien ! j'ai l'honneur de vous avertir que, si vous ne rapportez pas les mesures, à mon sens injustifiables, que vous avez prises contre vos ouvriers, vous n'aurez plus à compter sur cette bienveillance. »

M. Chagot se le tint pour dit et il permit à ses ouvriers de conserver et de remplir les fonctions qu'ils tenaient de la confiance de leurs concitoyens.

Eh bien, monsieur le ministre, comme M. Chagot, les compagnies de chemins de fer ont tous les jours recours à vous et à vos agents ; tous les jours, elles ont besoin de vos bons offices et des leurs. Veuillez seulement rappeler à MM. les ingénieurs de contrôle qu'ils sont moins encore les camarades que les directeurs, les contrôleurs et les surveillants des ingénieurs des compagnies, et je suis convaincu que ce simple avertissement suffira à obtenir de ces dernières les satisfactions que nous réclamons.

Messieurs, hier, vous votiez à une immense majorité le principe d'une loi nouvelle ayant pour but de garantir aux ouvriers la liberté d'association, qui leur est reconnue par les lois républicaines.

Je ne demande pas à M. le ministre des travaux publics de provoquer le vote d'une loi nouvelle ; je lui demande simplement de faire respecter les lois existantes. Il a pu y avoir hier des dissidences entre les membres de cette Chambre sur l'utilité de la loi qui vous est proposée ; mais j'estime qu'il ne saurait y en avoir sur la question que je soulève aujourd'hui.

Ce n'est pas simplement entre un patron et un ouvrier que le débat s'élève ; c'est entre une compagnie qui remplit un service public et des centaines de milliers de travailleurs. La question est de savoir s'ils doivent être livrés sans défense aux caprices des compagnies qui les emploient. Quant à moi, je ne le crois pas. (*Très bien ! très bien ! à gauche.*)

Figure – 14
May 1890
(p.786)

Chagot dans son cabinet et voi il r près le discours qu'il lui tint : a'JrlZ « Monsieur Chagot, lui dit-il, uS avtf tous les jours recours à labieaveLu Gouvernement et de son représeu c t vos affaires particulières. EhbieflJ rappof' neur de vous avertir que, si vous tez pas les mesures, à mon sens i ^oS o«' bles, que vous avez prises contre Pter sllr vriers, vous n'aurez plus à coropter $Ut cette bienveillance. » '1 per¡¡lt M. Chagot se le tint pour dit et il etelit à ses ouvriers de conserver et de rpoolplir les fonctions qu'ils tenaient de a de leurs concitoyens. , cow1 Eh bien, monsieur le m~~i~,"!

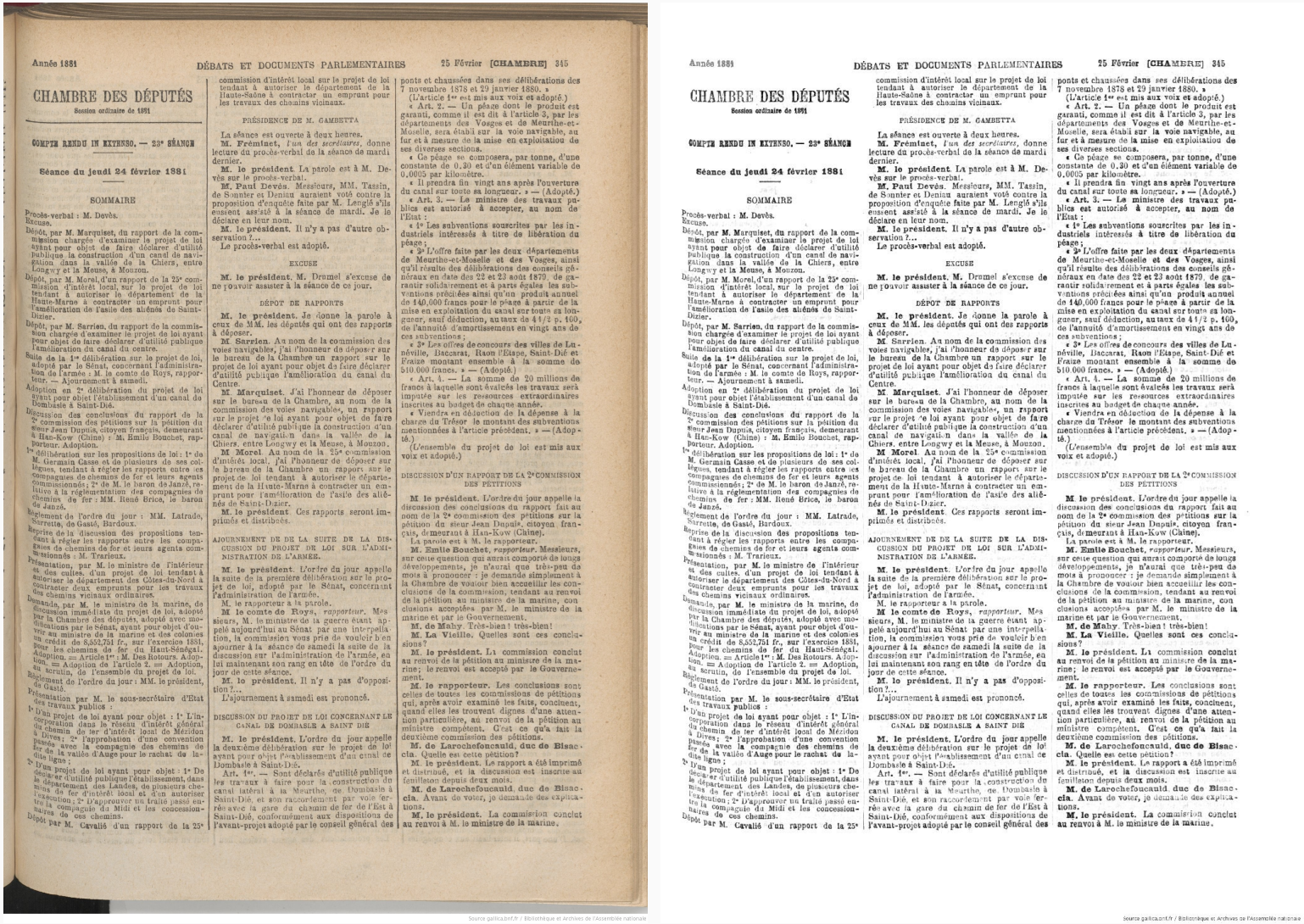
M. Chagot, les compagnies de r fer ont tous les jours recours a vou90m vos agents ; tous les jours, elles onV'euiUde de vos bons offices et des leurso.nvellieez seulement rappeler à MM. les îûe c~ contrôle qu'ils sont moins encorrôleurs e rades que les directeurs, les con coI11P' les surveillants des ingénieurs dee coH1P' les surveillants des ingénieurs c0 der, gnies, et je suis convaincu qe ces def avertissement suffira à obtenu nières les satisfactions que no r mons. une Messieurs, hier, vous vouez à loi ou j0j no mense majorité le principe ou velle ayant pour but de garantir x oUvr» la liberté d association, qui leur est recoe, nue par les lois républicaines. inistr8, de Je ne demande pas à M. ie III iigtre, travaux publics de provoquer * j ^eiii loi nouvelle ; je lui demande Siniple gleut P faire respecter les lois exis^ leS ro% y avoir hier des dissidences en bres de cette Chambre sur ais j'es l'Oll loi qui vous est propose; IIlla questl qu'il ne saurait y en avoir sur 011 que je soulève aujourd'hui. un t Ce n'est pas simplement en 'élève; C ef' et un ouvrier que le débat s lit ull S de entre une compagnie q lui ren'Pgtser.

vice public et des centaines de jugi gaVOir j, travailleurs. La question est de auJc Jlt.

doivent être livrés sans défensee au* JieV ces des compagnies qui les ifres 01'0 et, Quant à moi, je ne le crois p •

Figure – OCR result

IMAGE CLEANING TOOL DEVELOPED BY ANR SODUCO



(a) Original image

(b) Cleaned image

Figure – Demonstration of the SODUCO tool on a parliamentary debate page

OCR TOOL DEVELOPED BY SODUCO

Tool based on the OCR engine PERO OCR : very efficient on historical texts

Developed within the framework of the ANR SODUCO

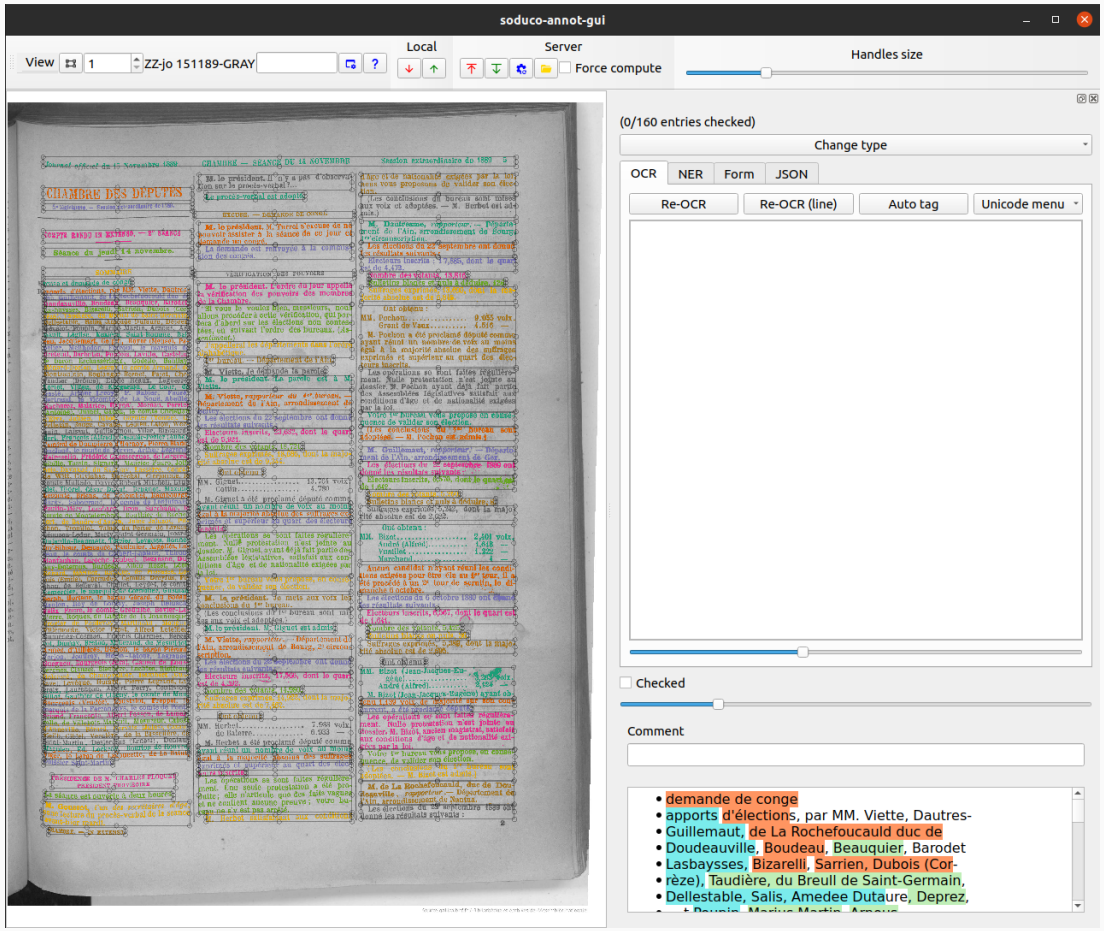
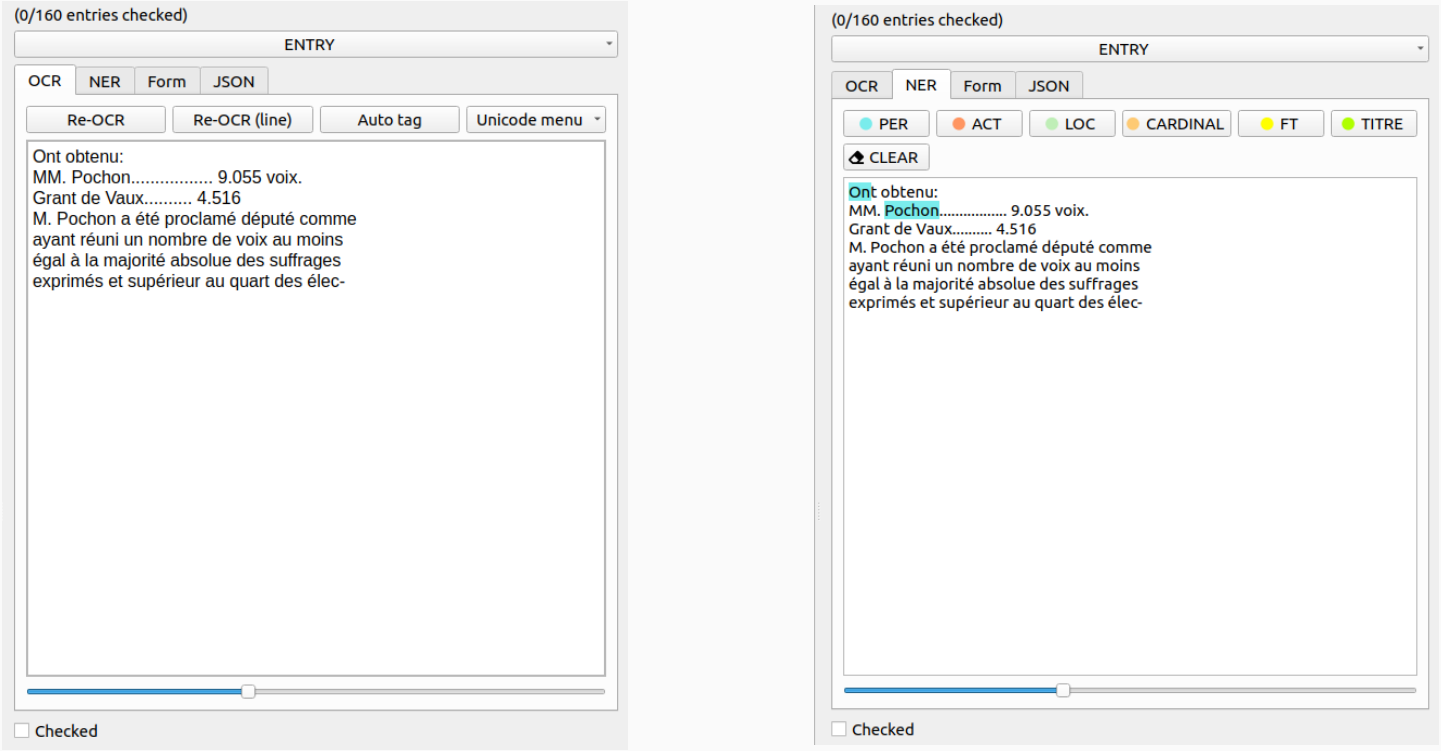


Figure – OCR tool developed by SODUCO

OCR AND NER



(a) OCR

(b) NER

Figure – OCR and NER zones

OUTPUT IN JSON


```
        50
    ],
    "id": 302,
    "ner_xml": "<PER>Suirrages</PER> exprim\u00e9s, 9,90<CARDINAL>8</CARDINAL>, dont la majo-\u2029rit\u00e9 absolue est de <CARDINAL>4,455</CARDINAL>.",
    "origin": "computer",
    "parent": 263,
    "persons": ["Suirrages"],
    "text_ocr": "Suirrages exprim\u00e9s, 9,908, dont la majo-\nrit\u00e9 absolue est de 4,455.",
    "type": "ENTRY",
    "activities": [],
    "comment": "",
    "checked": false
}
```


Figure – Output in JSON (extract)

THE PRINCIPLES OF THE ENCODING IN XML-TEI

ENCODING IN XML-TEI

Encoding is designed according to 4 principles :

- The different uses of texts
- The particularities of the source
- Similar projects : ParlaClarin and ParlaMint
- The automatic tagging process

For more information on this aspect :
https ://github.com/mpuren/agoda/tree/ODD

EXPERIMENTING WITH DIFFERENT SOLUTIONS

```
<lb/><u who="#pers_ID" xml:id="CR_1889-11-26_u5" ana="#speaker">
   <seg xml:id="CR_1889-11-26_u5.1">
      <persName ref="#pers_ID">M. Paul Déroulède</persName>. Mes amis et moi
      <lb/>avons applaudi à ce que nous avons cru, à
      <lb/>ce que nous croyons encore être la forma-
      <lb/>tion spontanée d'une majorité de con-
      <lb/>science, de justice, de tolérance et d'union.
      <lb/><incident><desc>(Applaudissements sur quelques bancs à l'ex-
      <!-- Pas de lb possible dans incident --> trémité gauche de La salle. – Exclamations
      <!-- Pas de lb possible dans incident --> au centre et à gauche.)</desc></incident>
   </seg>
</u>

<lb/><u who="pers_ID" xml:id="CR_1889-11-26_u6" ana="#speaker">
   <seg xml:id="CR_1889-11-26_u6.1">
      <persName ref="#pers_ID">Gustave Rivet</persName>. La tolérance pour
      <lb/>toutes Les ignominies qui ont été commises
      <lb/>Pendant la période électorale.
   </seg>
</u>
```

(a) Encoding model 1 : semantics and layout

```
        <u who="#pers_ID" xml:id="CR_1889-11-26_u5" ana="#speaker">
           <seg xml:id="CR_1889-11-26_u5.1"><persName ref="#pers_ID">M. Paul Déroulède</persName>. Mes amis et moi
avons applaudi à ce que nous avons cru, à ce que nous croyons encore être la formation spontanée d'une majorité de
conscience, de justice, de tolérance et d'union. <incident><desc>(Applaudissements sur quelques bancs à l'extrémité
gauche de la salle. –Exclamations au centre et à gauche.)</desc></incident></seg>
        </u>

        <u who="#pers_ID" xml:id="CR_1889-11-26_u6" ana="#speaker">
           <seg xml:id="CR_1889-11-26_u6.1"><persName ref="#pers_ID">Gustave Rivet</persName>. La tolérance pour
toutes les ignominies qui ont été commises pendant la période électorale.</seg>
        </u>
```

(b) Encoding model 2 : semantics only

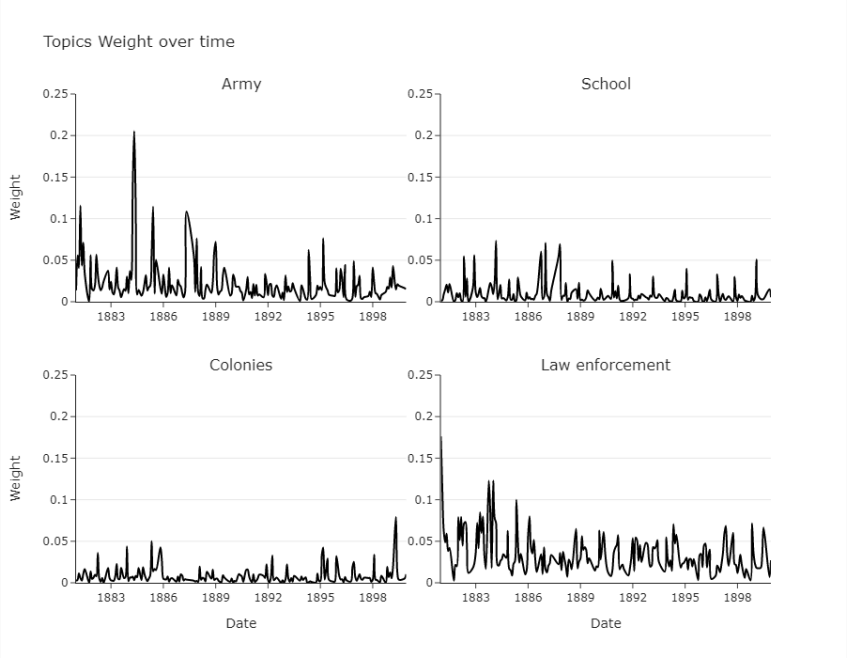
Figure – Parliamentary session of 26 November 1889 (extract)

Topic Modeling and Word Embedding

EXPLORING DEBATES WITH TOPIC MODELING

Topic 8	Topic 11	Topic 15
salaire	général	pari
question	commission	télégraphe
gouvernement	régiment	faire
jour	troupe	ingénieur
patron	monsieur	train
chambre	année	ligne
droit	jeune	chambre
syndicat	temps	personnel
délégué	faire	etat
monsieur	corps	administration
travail	soldat	employé
travaux	ministre	poste
ministre	homme	public
grève	loi	travaux
faire	an	service
mineur	guerre	agent
mine	service	ministre
loi	militaire	fer
compagnie	officier	chemin
ouvrier	armée	compagnie

(a) 3 topics among 40 : working class (8), army (11) and infrastructures (15)



(b) Topic's evolution over time

Figure – Topic modeling with LDA

WORD EMBEDDING : WORD2VEC AND TOP2VEC

(a) Word vectors projected with t-SNE (word2vec : CBOW,W=5)

Cluster 55	Cluster 68	Cluster 70
victimes	divorce	enveloppes
inondations	epoux	timbres
secourir	mariage	poste
eprouvees	conjugal	postale
orages	divorces	timbre
sinistres	adultere	recepisses
grele	conjugale	postes
secours	remarier	postaux
venir	separation	telegraphes
infortunes	indissolubilite	colis
ravages	conjoints	fixe
miseres	mutuel	recouvrements
catastrophe	separations	graphes
evenements	mari	postales
repartition	mariages	taxe
incendies	femme	decide
soulager	conjoint	soit

(b) 3 clusters among the 113 topics found by top2vec : storm (55), divorce (68) and poste (70)

Figure – Results of word2vec and top2vec

FOR MORE INFORMATION

- Nicolas Bourgeois, Aurélien Pellet, Marie Puren. ”Using Topic Generation Model to explore the French Parliamentary Debates during the early Third Republic (1881-1899)”. (hal-03526254v2)
- Marie Puren, Nicolas Bourgeois, Aurélien Pellet, Pierre Vernus, Fanny Lebreton. ”Between History and Natural Language Processing : Study, Enrichment and Online Publication of French Parliamentary Debates of the Early Third Republic (1881-1899)”. ParlaCLARIN III at LREC2022 - Workshop on Creating, Enriching and Using Parliamentary Corpora, Jun 2022, Marseille, France. (hal-03623351)

THANK YOU FOR YOUR ATTENTION!

Marie Puren : marie.puren@epitech.eu

Pierre Vernus : pierre.vernus@msh-lse.fr

Aurélien Pellet : aurelien.pellet@epitech.eu

Nicolas Bourgeois : nicolas.bourgeois@epitech.eu

Fanny Lebreton : fanny.lebreton@chartes.psl.eu