BETWEEN HISTORY AND NATURAL LANGUAGE PROCESSING : STUDY, ENRICHMENT AND ONLINE PUBLICATION OF FRENCH PARLIAMENTARY DEBATES OF THE EARLY THIRD REPUBLIC (1881–1899)

The AGODA project

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

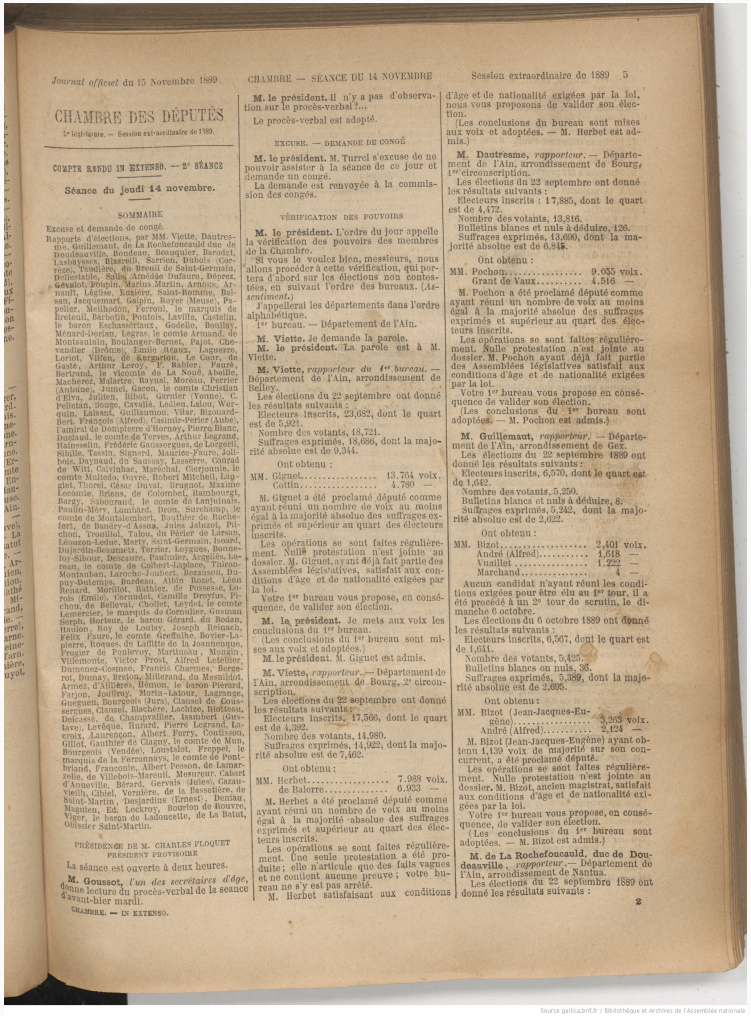
20 June 2022

ParlaCLARIN III at LREC2022 - Workshop on Creating, Enriching and Using Parliamentary Corpora, Marseille

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

Figure – Workflow

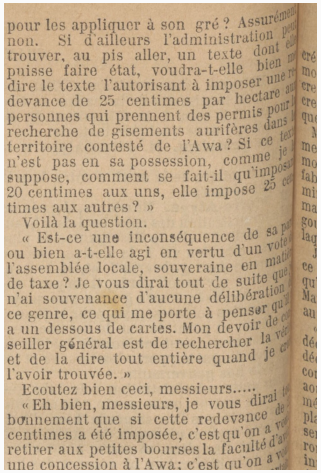
OCERISE THE DEBATES

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



pour les appliquer à son gré ? Assure.

non. Si d'ailleurs l'administration fj, trouver, au pis aller, un texte (1011t eu puisse faire état, voudra-t-elle dire le texte l'autorisant à imposer u devance de 25 centimes par hectar , personnes qui prennent des permis Pour1 recherche de gisements aurifères a teSI territoire contesté de l'Awa ? Si ce tes

n'est pas en sa possession, cornas le suppose, comment se fait-il qulIWjs3j 20 centimes aux uns, elle impose 20 cet times aux autres ? » , Voilà la question. é

« Est-ce une inconséquence de sa te ou bien a-t-elle agi en vertu d'un v l'assemblée locale, souveraine en OlqallC de taxe? Je vous dirai tout de suite Il n'ai souvenance d'aucune délibératl°ll'V ce genre, ce qui me porte à penser a un dessous de cartes. Mon devoir seiller général est de rechercher la cJ'!" et de la dire tout entière quand Je l'avoir trouvée. »

Ecoutez bien ceci, messieurs. l «Eh bien, messieurs, je vous dIra de bonnement que si cette redevance J centimes a été imposée, c'est qu'on a, retirer aux petites bourses la faculté" une concession à l'Awa; c'est qu'on a r¡;:i\

Figure – 14 May 1890 (p.786)

Figure – OCR result (Gallica)

We took decision to re-ocerise the corpus.

CONSEQUENCES ON THE PROCESSING AND PUBLISHING CHAIN

1. Incorrect segmentation
2. Poor OCR performance
3. Very curved and even cut text : not very readable by a human being, even less readable by a machine
4. Automatic annotation in XML-TEI more difficult because linked to segmentation
5. Text too faulty to publish online
6. Negative influence of OCR errors on the performance of NLP analyses (NER, topic modeling, word embedding) (cf. Assessing the Impact of OCR Quality on Downstream NLP Tasks)

WHAT SOLUTIONS ?

- Dewarping solution : unsatisfactory as it can bend the flat parts and even cut the image
  - Consider developing a solution that removes the curvature only for the parts of the image concerned (column at the binding edge)
- Use a language model : when the text is not human readable and therefore not machine readable, use a language model to predict the missing word(s)
  - Language model for French : trained on (too much ?) contemporary French - perhaps need to re-train on late 19th century French
- Words cut because of the curvature + words ocerised but appearing in disorder (consequence of the bad segmentation due to the curvature of the page) : use an AI to restore the text, as Ithaca of Deep Mind does (cf. Restoring and attributing ancient texts using deep neural networks)

IMAGE CLEANING TOOL DEVELOPED BY ANR SODUCO

Année 1881 DÉBATS ET DOCUMENTS PARLEMENTAIRES 25 Février [CHAMBRE] 345

CHAMBRE DES DÉPUTÉS

Session ordinaire de 1881

COMPTE RENDU IN EXTENSO. — 23e SÉANCE

Séance du jeudi 24 février 1881

SOMMAIRE

Procès-verbal : M. Devès.

Excuse.

Dépôt, par M. Marquiset, du rapport de la commission chargée d'examiner le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction d'un canal de navigation dans la vallée de la Chiers, entre Longwy et la Meuse, à Mouzon.

Dépôt, par M. Morel, du rapport de la 2e commission d'intérêt local, sur le projet de loi tendant à autoriser le département de la Haute-Marne à contracter un emprunt pour l'amélioration de l'asile des aliénés de Saint-Dizier.

Dépôt, par M. Sarrien, du rapport de la commission chargée d'examiner le projet de loi ayant pour objet de faire déclarer d'utilité publique l'amélioration du canal du centre.

Suite de la 1re délibération sur le projet de loi, adopté par le Sénat, concernant l'administration de l'armée : M. le comte de Roys, rapporteur. — Ajournement à samedi.

Adoption en 2e délibération du projet de loi ayant pour objet l'établissement d'un canal de Dombasle à Saint-Dié.

Discussion des conclusions du rapport de la 2e commission des pétitions sur la pétition du sieur Jean Dupuis, citoyen français, demeurant à Han-Kow (Chine) : M. Emile Bouchet, rapporteur. Adoption.

1re délibération sur les propositions de loi : 1° de M. Germain Casse et de plusieurs de ses collègues, tendant à régler les rapports entre les compagnies de chemins de fer et leurs agents commissionnés ; 2° de M. le baron de Janzé, relative à la réglementation des compagnies de chemins de fer : MM. René Brice, le baron de Janzé.

Règlement de l'ordre du jour : MM. Latrade, Sarrette, de Gasté, Bardoux.

Reprise de la discussion des propositions tendant à régler les rapports entre les compagnies de chemins de fer et leurs agents commissionnés : M. Trarieux.

Présentation, par M. le ministre de l'intérieur et des cultes, d'un projet de loi tendant à autoriser le département des Côtes-du-Nord à contracter deux emprunts pour les travaux des chemins vicinaux ordinaires.

Demande, par M. le ministre de la marine, de discussion immédiate du projet de loi, adopté par la Chambre des députés, adopté avec modifications par le Sénat, ayant pour objet d'ouvrir au ministre de la marine et des colonies un crédit de 8,352,751 fr., sur l'exercice 1881, pour les chemins de fer du Haut-Sénégal. — Adoption. = Article 1er : M. Des Rotours. Adoption. = Adoption de l'article 2. = Adoption, au scrutin, de l'ensemble du projet de loi.

Règlement de l'ordre du jour : MM. le président, de Gasté.

Présentation par M. le sous-secrétaire d'Etat des travaux publics :

1° D'un projet de loi ayant pour objet : 1° L'incorporation dans le réseau d'intérêt général du chemin de fer d'intérêt local de Mézidon à Dives ; 2° l'approbation d'une convention passée avec la compagnie des chemins de fer de la vallée d'Auge pour le rachat de ladite ligne ;

2° D'un projet de loi ayant pour objet : 1° De déclarer d'utilité publique l'établissement, dans le département des Landes, de plusieurs chemins de fer d'intérêt local et d'en autoriser l'exécution ; 2° d'approuver un traité passé entre la compagnie du Midi et les concessionnaires de ces chemins.

Dépôt par M. Cavalié d'un rapport de la 25e commission d'intérêt local sur le projet de loi tendant à autoriser le département de la Haute-Saône à contracter un emprunt pour les travaux des chemins vicinaux.

PRÉSIDENCE DE M. GAMBETTA

La séance est ouverte à deux heures.

M. Frémiet, *l'un des secrétaires*, donne lecture du procès-verbal de la séance de mardi dernier.

M. le président. La parole est à M. Devès sur le procès-verbal.

M. Paul Devès. Messieurs, MM. Tassin, de Sonnier et Deniau auraient voté contre la proposition d'enquête faite par M. Lorgé s'ils eussent assisté à la séance de mardi. Je le déclare en leur nom.

M. le président. Il n'y a pas d'autre observation ?...

Le procès-verbal est adopté.

EXCUSE

M. le président. M. Dremel s'excuse de ne pouvoir assister à la séance de ce jour.

DÉPÔT DE RAPPORTS

M. le président. Je donne la parole à ceux de MM. les députés qui ont des rapports à déposer.

M. Sarrien. Au nom de la commission des voies navigables, j'ai l'honneur de déposer sur le bureau de la Chambre un rapport sur le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction du canal du Centre.

M. Marquiset. J'ai l'honneur de déposer sur le bureau de la Chambre, au nom de la commission des voies navigables, un rapport sur le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction d'un canal de navigation dans la vallée de la Chiers, entre Longwy et la Meuse, à Mouzon.

M. Morel. Au nom de la 2e commission d'intérêt local, j'ai l'honneur de déposer sur le bureau de la Chambre un rapport sur le projet de loi tendant à autoriser le département de la Haute-Marne à contracter un emprunt pour l'amélioration de l'asile des aliénés de Saint-Dizier.

M. le président. Ces rapports seront imprimés et distribués.

AJOURNEMENT DE LA SUITE DE LA DISCUSSION DU PROJET DE LOI SUR L'ADMINISTRATION DE L'ARMÉE.

M. le président. L'ordre du jour appelle la suite de la première délibération sur le projet de loi, adopté par le Sénat, concernant l'administration de l'armée.

M. le rapporteur a la parole.

M. le comte de Roys, *rapporteur*. Messieurs, M. le ministre de la guerre étant appelé aujourd'hui au Sénat par une interpellation, la commission vous prie de vouloir bien ajourner à la séance de samedi la suite de la discussion sur l'administration de l'armée, en lui maintenant son rang en tête de l'ordre du jour de cette séance.

M. le président. Il n'y a pas d'opposition ?...

L'ajournement à samedi est prononcé.

DISCUSSION DU PROJET DE LOI CONCERNANT LE CANAL DE DOMBASLE A SAINT DIÉ

M. le président. L'ordre du jour appelle la deuxième délibération sur le projet de loi ayant pour objet l'établissement d'un canal de Dombasle à Saint-Dié.

Art. 1er. — Sont déclarés d'utilité publique les travaux à faire pour la construction du canal latéral à la Meurthe, de Dombasle à Saint-Dié, et son raccordement par voie ferrée avec la gare du chemin de fer de l'Est à Saint-Dié, conformément aux dispositions de l'avant-projet adopté par le conseil général des ponts et chaussées dans ses délibérations des 7 novembre 1878 et 29 janvier 1880. »

(L'article 1er est mis aux voix et adopté.)

« Art. 2. — Un péage dont le produit est garanti, comme il est dit à l'article 3, par les départements des Vosges et de Meurthe-et-Moselle, sera établi sur la voie navigable, au fur et à mesure de la mise en exploitation de ses diverses sections.

« Ce péage se composera, par tonne, d'une constante de 0,30 et d'un élément variable de 0,0005 par kilomètre.

« Il prendra fin vingt ans après l'ouverture du canal sur toute sa longueur. » — (Adopté.)

« Art. 3. — Le ministre des travaux publics est autorisé à accepter, au nom de l'Etat :

« 1° Les subventions souscrites par les industriels intéressés à titre de libération du péage ;

« 2° L'offre faite par les deux départements de Meurthe-et-Moselle et des Vosges, ainsi qu'il résulte des délibérations des conseils généraux en date des 22 et 23 août 1879, de garantir solidairement et à parts égales les subventions précitées ainsi qu'un produit annuel de 140,000 francs pour le péage à partir de la mise en exploitation du canal sur toute sa longueur, sauf déduction, au taux de 4 1/2 p. 100, de l'annuité d'amortissement en vingt ans de ces subventions ;

« 3° Les offres de concours des villes de Lunéville, Baccarat, Raon l'Etape, Saint-Dié et Fraize montant ensemble à la somme de 510,000 francs. » — (Adopté.)

« Art. 4. — La somme de 20 millions de francs à laquelle sont évalués les travaux sera imputée sur les ressources extraordinaires inscrites au budget de chaque année.

« Viendra en déduction de la dépense à la charge du Trésor le montant des subventions mentionnées à l'article précédent. » — (Adopté.)

(L'ensemble du projet de loi est mis aux voix et adopté.)

DISCUSSION D'UN RAPPORT DE LA 2e COMMISSION DES PÉTITIONS

M. le président. L'ordre du jour appelle la discussion des conclusions du rapport fait au nom de la 2e commission des pétitions sur la pétition du sieur Jean Dupuis, citoyen français, demeurant à Han-Kow (Chine).

La parole est à M. le rapporteur.

M. Emile Bouchet, *rapporteur*. Messieurs, sur cette question qui aurait comporté de longs développements, je n'aurai que très-peu de mots à prononcer : je demande simplement à la Chambre de vouloir bien accueillir les conclusions de la commission, tendant au renvoi de la pétition au ministre de la marine, conclusions acceptées par M. le ministre de la marine et par le Gouvernement.

M. de Mahy. Très-bien ! très-bien !

M. La Vieille. Quelles sont ces conclusions ?

M. le président. La commission conclut au renvoi de la pétition au ministre de la marine ; le renvoi est accepté par le Gouvernement.

M. le rapporteur. Les conclusions sont celles de toutes les commissions de pétitions qui, après avoir examiné les faits, concluent, quand elles les trouvent dignes d'une attention particulière, au renvoi de la pétition au ministre compétent. C'est ce qu'a fait la deuxième commission des pétitions.

M. de Larochefoucauld, duc de Bisaccia. Quelle est cette pétition ?

M. le président. Le rapport a été imprimé et distribué, et la discussion est inscrite au feuilleton depuis deux mois.

M. de Larochefoucauld, duc de Bisaccia. Avant de voter, je demande des explications.

M. le président. La commission conclut au renvoi à M. le ministre de la marine.

(a) Original image

Année 1881 DÉBATS ET DOCUMENTS PARLEMENTAIRES 25 Février [CHAMBRE] 345

CHAMBRE DES DÉPUTÉS

Session ordinaire de 1881

COMPTE RENDU IN EXTENSO. — 23e SÉANCE

Séance du jeudi 24 février 1881

SOMMAIRE

Procès-verbal : M. Devès.

Excuse.

Dépôt, par M. Marquiset, du rapport de la commission chargée d'examiner le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction d'un canal de navigation dans la vallée de la Chiers, entre Longwy et la Meuse, à Mouzon.

Dépôt, par M. Morel, du rapport de la 2e commission d'intérêt local, sur le projet de loi tendant à autoriser le département de la Haute-Marne à contracter un emprunt pour l'amélioration de l'asile des aliénés de Saint-Dizier.

Dépôt, par M. Sarrien, du rapport de la commission chargée d'examiner le projet de loi ayant pour objet de faire déclarer d'utilité publique l'amélioration du canal du centre.

Suite de la 1re délibération sur le projet de loi, adopté par le Sénat, concernant l'administration de l'armée : M. le comte de Roys, rapporteur. — Ajournement à samedi.

Adoption en 2e délibération du projet de loi ayant pour objet l'établissement d'un canal de Dombasle à Saint-Dié.

Discussion des conclusions du rapport de la 2e commission des pétitions sur la pétition du sieur Jean Dupuis, citoyen français, demeurant à Han-Kow (Chine) : M. Emile Bouchet, rapporteur. Adoption.

1re délibération sur les propositions de loi : 1° de M. Germain Casse et de plusieurs de ses collègues, tendant à régler les rapports entre les compagnies de chemins de fer et leurs agents commissionnés ; 2° de M. le baron de Janzé, relative à la réglementation des compagnies de chemins de fer : MM. René Brice, le baron de Janzé.

Règlement de l'ordre du jour : MM. Latrade, Sarrette, de Gasté, Bardoux.

Reprise de la discussion des propositions tendant à régler les rapports entre les compagnies de chemins de fer et leurs agents commissionnés : M. Trarieux.

Présentation, par M. le ministre de l'intérieur et des cultes, d'un projet de loi tendant à autoriser le département des Côtes-du-Nord à contracter deux emprunts pour les travaux des chemins vicinaux ordinaires.

Demande, par M. le ministre de la marine, de discussion immédiate du projet de loi, adopté par la Chambre des députés, adopté avec modifications par le Sénat, ayant pour objet d'ouvrir au ministre de la marine et des colonies un crédit de 8,352,751 fr., sur l'exercice 1881, pour les chemins de fer du Haut-Sénégal. — Adoption. = Article 1er : M. Des Rotours. Adoption. = Adoption de l'article 2. = Adoption, au scrutin, de l'ensemble du projet de loi.

Règlement de l'ordre du jour : MM. le président, de Gasté.

Présentation par M. le sous-secrétaire d'Etat des travaux publics :

1° D'un projet de loi ayant pour objet : 1° L'incorporation dans le réseau d'intérêt général du chemin de fer d'intérêt local de Mézidon à Dives ; 2° l'approbation d'une convention passée avec la compagnie des chemins de fer de la vallée d'Auge pour le rachat de ladite ligne ;

2° D'un projet de loi ayant pour objet : 1° De déclarer d'utilité publique l'établissement, dans le département des Landes, de plusieurs chemins de fer d'intérêt local et d'en autoriser l'exécution ; 2° d'approuver un traité passé entre la compagnie du Midi et les concessionnaires de ces chemins.

Dépôt par M. Cavalié d'un rapport de la 25e commission d'intérêt local sur le projet de loi tendant à autoriser le département de la Haute-Saône à contracter un emprunt pour les travaux des chemins vicinaux.

PRÉSIDENCE DE M. GAMBETTA

La séance est ouverte à deux heures.

M. Frémiet, *l'un des secrétaires*, donne lecture du procès-verbal de la séance de mardi dernier.

M. le président. La parole est à M. Devès sur le procès-verbal.

M. Paul Devès. Messieurs, MM. Tassin, de Sonnier et Deniau auraient voté contre la proposition d'enquête faite par M. Lorgé s'ils eussent assisté à la séance de mardi. Je le déclare en leur nom.

M. le président. Il n'y a pas d'autre observation ?...

Le procès-verbal est adopté.

EXCUSE

M. le président. M. Dremel s'excuse de ne pouvoir assister à la séance de ce jour.

DÉPÔT DE RAPPORTS

M. le président. Je donne la parole à ceux de MM. les députés qui ont des rapports à déposer.

M. Sarrien. Au nom de la commission des voies navigables, j'ai l'honneur de déposer sur le bureau de la Chambre un rapport sur le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction du canal du Centre.

M. Marquiset. J'ai l'honneur de déposer sur le bureau de la Chambre, au nom de la commission des voies navigables, un rapport sur le projet de loi ayant pour objet de faire déclarer d'utilité publique la construction d'un canal de navigation dans la vallée de la Chiers, entre Longwy et la Meuse, à Mouzon.

M. Morel. Au nom de la 2e commission d'intérêt local, j'ai l'honneur de déposer sur le bureau de la Chambre un rapport sur le projet de loi tendant à autoriser le département de la Haute-Marne à contracter un emprunt pour l'amélioration de l'asile des aliénés de Saint-Dizier.

M. le président. Ces rapports seront imprimés et distribués.

AJOURNEMENT DE LA SUITE DE LA DISCUSSION DU PROJET DE LOI SUR L'ADMINISTRATION DE L'ARMÉE.

M. le président. L'ordre du jour appelle la suite de la première délibération sur le projet de loi, adopté par le Sénat, concernant l'administration de l'armée.

M. le rapporteur a la parole.

M. le comte de Roys, *rapporteur*. Messieurs, M. le ministre de la guerre étant appelé aujourd'hui au Sénat par une interpellation, la commission vous prie de vouloir bien ajourner à la séance de samedi la suite de la discussion sur l'administration de l'armée, en lui maintenant son rang en tête de l'ordre du jour de cette séance.

M. le président. Il n'y a pas d'opposition ?...

L'ajournement à samedi est prononcé.

DISCUSSION DU PROJET DE LOI CONCERNANT LE CANAL DE DOMBASLE A SAINT DIÉ

M. le président. L'ordre du jour appelle la deuxième délibération sur le projet de loi ayant pour objet l'établissement d'un canal de Dombasle à Saint-Dié.

Art. 1er. — Sont déclarés d'utilité publique les travaux à faire pour la construction du canal latéral à la Meurthe, de Dombasle à Saint-Dié, et son raccordement par voie ferrée avec la gare du chemin de fer de l'Est à Saint-Dié, conformément aux dispositions de l'avant-projet adopté par le conseil général des ponts et chaussées dans ses délibérations des 7 novembre 1878 et 29 janvier 1880. »

(L'article 1er est mis aux voix et adopté.)

« Art. 2. — Un péage dont le produit est garanti, comme il est dit à l'article 3, par les départements des Vosges et de Meurthe-et-Moselle, sera établi sur la voie navigable, au fur et à mesure de la mise en exploitation de ses diverses sections.

« Ce péage se composera, par tonne, d'une constante de 0,30 et d'un élément variable de 0,0005 par kilomètre.

« Il prendra fin vingt ans après l'ouverture du canal sur toute sa longueur. » — (Adopté.)

« Art. 3. — Le ministre des travaux publics est autorisé à accepter, au nom de l'Etat :

« 1° Les subventions souscrites par les industriels intéressés à titre de libération du péage ;

« 2° L'offre faite par les deux départements de Meurthe-et-Moselle et des Vosges, ainsi qu'il résulte des délibérations des conseils généraux en date des 22 et 23 août 1879, de garantir solidairement et à parts égales les subventions précitées ainsi qu'un produit annuel de 140,000 francs pour le péage à partir de la mise en exploitation du canal sur toute sa longueur, sauf déduction, au taux de 4 1/2 p. 100, de l'annuité d'amortissement en vingt ans de ces subventions ;

« 3° Les offres de concours des villes de Lunéville, Baccarat, Raon l'Etape, Saint-Dié et Fraize montant ensemble à la somme de 510,000 francs. » — (Adopté.)

« Art. 4. — La somme de 20 millions de francs à laquelle sont évalués les travaux sera imputée sur les ressources extraordinaires inscrites au budget de chaque année.

« Viendra en déduction de la dépense à la charge du Trésor le montant des subventions mentionnées à l'article précédent. » — (Adopté.)

(L'ensemble du projet de loi est mis aux voix et adopté.)

DISCUSSION D'UN RAPPORT DE LA 2e COMMISSION DES PÉTITIONS

M. le président. L'ordre du jour appelle la discussion des conclusions du rapport fait au nom de la 2e commission des pétitions sur la pétition du sieur Jean Dupuis, citoyen français, demeurant à Han-Kow (Chine).

La parole est à M. le rapporteur.

M. Emile Bouchet, *rapporteur*. Messieurs, sur cette question qui aurait comporté de longs développements, je n'aurai que très-peu de mots à prononcer : je demande simplement à la Chambre de vouloir bien accueillir les conclusions de la commission, tendant au renvoi de la pétition au ministre de la marine, conclusions acceptées par M. le ministre de la marine et par le Gouvernement.

M. de Mahy. Très-bien ! très-bien !

M. La Vieille. Quelles sont ces conclusions ?

M. le président. La commission conclut au renvoi de la pétition au ministre de la marine ; le renvoi est accepté par le Gouvernement.

M. le rapporteur. Les conclusions sont celles de toutes les commissions de pétitions qui, après avoir examiné les faits, concluent, quand elles les trouvent dignes d'une attention particulière, au renvoi de la pétition au ministre compétent. C'est ce qu'a fait la deuxième commission des pétitions.

M. de Larochefoucauld, duc de Bisaccia. Quelle est cette pétition ?

M. le président. Le rapport a été imprimé et distribué, et la discussion est inscrite au feuilleton depuis deux mois.

M. de Larochefoucauld, duc de Bisaccia. Avant de voter, je demande des explications.

M. le président. La commission conclut au renvoi à M. le ministre de la marine.

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

(b) Cleaned image

Figure – Demonstration of the SODUCO tool on a 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

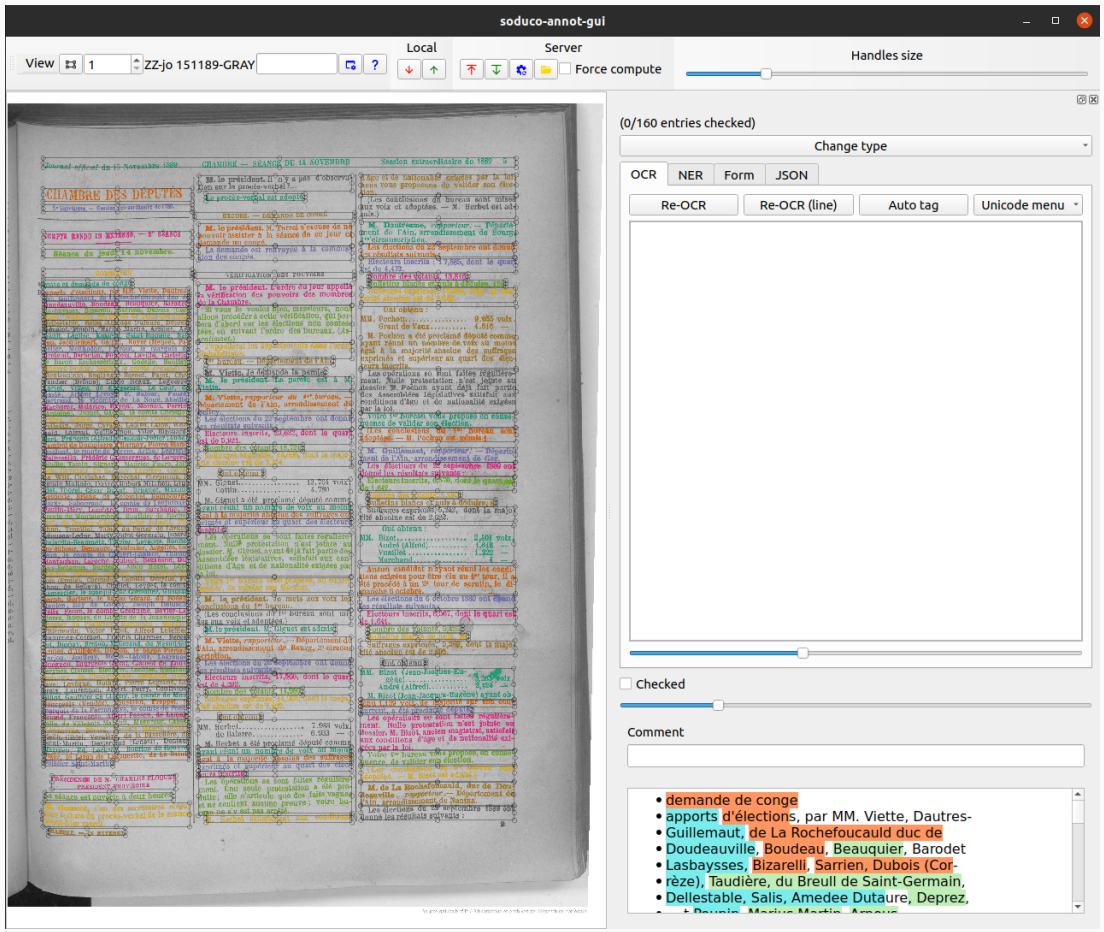
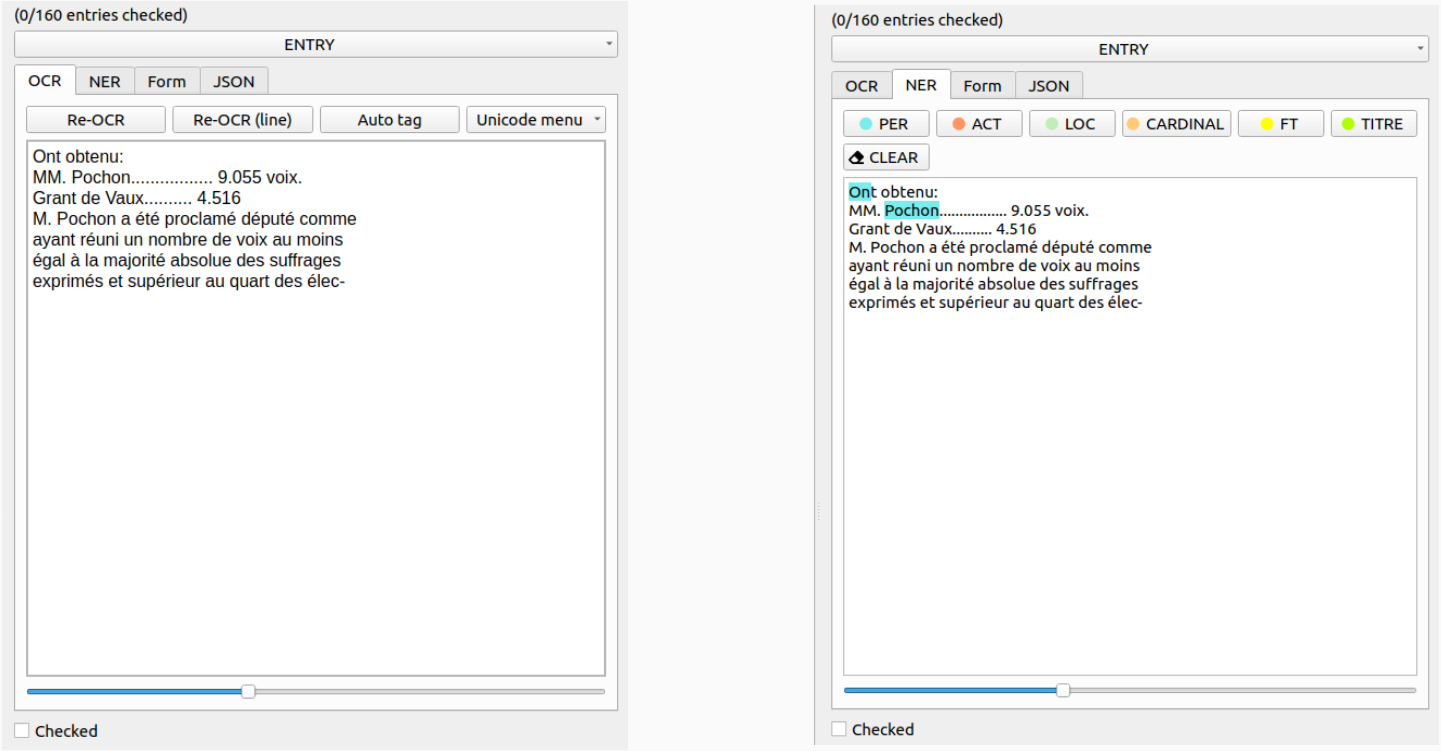


Figure – OCR tool developed by SODUCO

OCR AND NER



(a) OCR

(b) NER

Figure – OCR and NER zones

OUTPUT IN JSON


```
      "box": [
           827.8018264183805,
           2374.55023568486,
           589.3638270234404,
           207.5897599646105
       ],
       "checked": true,
       "comment": "u-par",
       "id": 352,
       "ner_xml": "<PER>M. Francis Laur</PER>. Je n'ai pas l'intention\u2029d'apporter \u00e0 la Chambre d'autr\u2029S de",
       "origin": "computer",
       "parent": 274,
       "persons": ["M. Francis Laur"],
       "text_ocr": "M. Francis Laur. Je n'ai pas l'intention\nd'apporter \u00e0 la Chambre d'autres documents\nque ceux qui me sont fournis par la lettre\nde
d\u00e9mission de M. Lev\u00eaque, adress\u00e9e \u00e0\nM. le ministre des finances. Je demanderai\n\u00e0 la Chambre si elle conna\u00eet cette lettre ou\nsi
elle veut que j'en donne lecture.",
       "type": "ENTRY"
   },
   {
       "activities": [],
       "addresses": [],
```


Figure – Output in JSON (extract)

POST-CORRECTION

- Post-correction dictionary
- Use of regular expressions : manage multiple spaces, line breaks, “-”
- Endogenous corrections
- CamemBERT language model to correct faulty words (see above)
- Automatic translation ? “Translate” wrong words into the correct form

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

GENERAL STRUCTURE OF TEI FILES

```xml
<teiCorpus xmlns="http://www.tei-c.org/ns/1.0" xml:id="FR_3R_5L" xml:lang="fr">

   <!-- Métadonnées du corpus (non développées) -->
   <teiHeader>
      <fileDesc>
         <titleStmt>
            <title></title>
         </titleStmt>
         <publicationStmt>
            <p></p>
         </publicationStmt>
         <sourceDesc>
            <p></p>
         </sourceDesc>
      </fileDesc>
   </teiHeader>

   <!-- Données liées et informations contextuelles -->
   <standOff>
      <listPerson>
         <person></person>
      </listPerson>

      <listOrg>
         <org></org>
      </listOrg>

      <listPlace>
         <place></place>
      </listPlace>
   </standOff>

   <!-- Stockage du composant correspondant à la séance du 26 novembre 1889 -->
   <xi:include xmlns:xi="http://www.w3.org/2001/XInclude" href="FR_3R_5L_1889-11-26.xml"/>


   <!-- Stockage des autres composants du corpus de façon identique -->
   <!-- ... -->

</teiCorpus>
```

(a) General structure of a corpus file

```xml
<TEI xmlns="http://www.tei-c.org/ns/1.0" xml:id="FR_3R_5L_1889-11-26" xml:lang="fr">

   <!-- Métadonnées du composant (non développées) -->
   <teiHeader>
      <fileDesc>
         <titleStmt>
            <title></title>
         </titleStmt>
         <publicationStmt>
            <p></p>
         </publicationStmt>
         <sourceDesc>
            <p></p>
         </sourceDesc>
      </fileDesc>
   </teiHeader>

   <text>

      <!-- Transcription du compte rendu -->
      <body>
         <div></div>
      </body>

      <!-- Annexes du compte rendu -->
      <back></back>

   </text>

</TEI>
```

(b) General structure of a component file

Figure – Encodings - General structure of corpus and component files

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)

APPLYING ENCODING


```
    "box": [
        827.8018264183805,
        2374.55023568486,
        589.3638270234404,
        207.5897599646105
    ],
    "checked": true,
    "comment": "u-par",
    "id": 352,
    "ner_xml": "<PER>M. Francis Laur</PER>. Je n'ai pas l'intention\u2029d'apporter \u00e0 la Chambre d'autr\u2029s de",
    "origin": "computer",
    "parent": 274,
    "persons": ["M. Francis Laur"],
    "text_ocr": "M. Francis Laur. Je n'ai pas l'intention\nd'apporter \u00e0 la Chambre d'autres documents\nque ceux qui me sont fournis par la lettre\nde d\u00e9mission de M. Lev\u00eaque, adress\u00e9e \u00e0\nM. le ministre des finances. Je demanderai\n\u00e0 la Chambre si elle conna\u00eet cette lettre ou\nsi elle veut que j'en donne lecture.",
    "type": "ENTRY"
},
{
    "activities": [],
    "addresses": [],
```


Textual data in JSON

↓

Python scripts

↓

Textual data in XML-TEI

PUBLISH AN EDITORIALIZED CORPUS

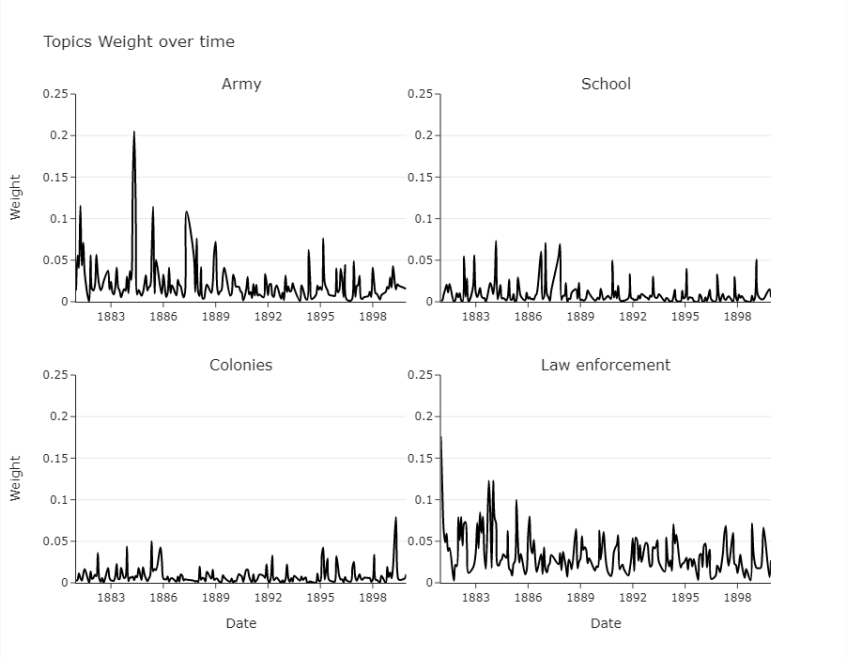
- Using TEI Publisher : a tool for publishing corpora in TEI
- Based on Web Components technology : allows the development of new functionalities
- Develop new modes of visualisation : for example, explore debates according to the subjects they evoke, and not only by key words

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

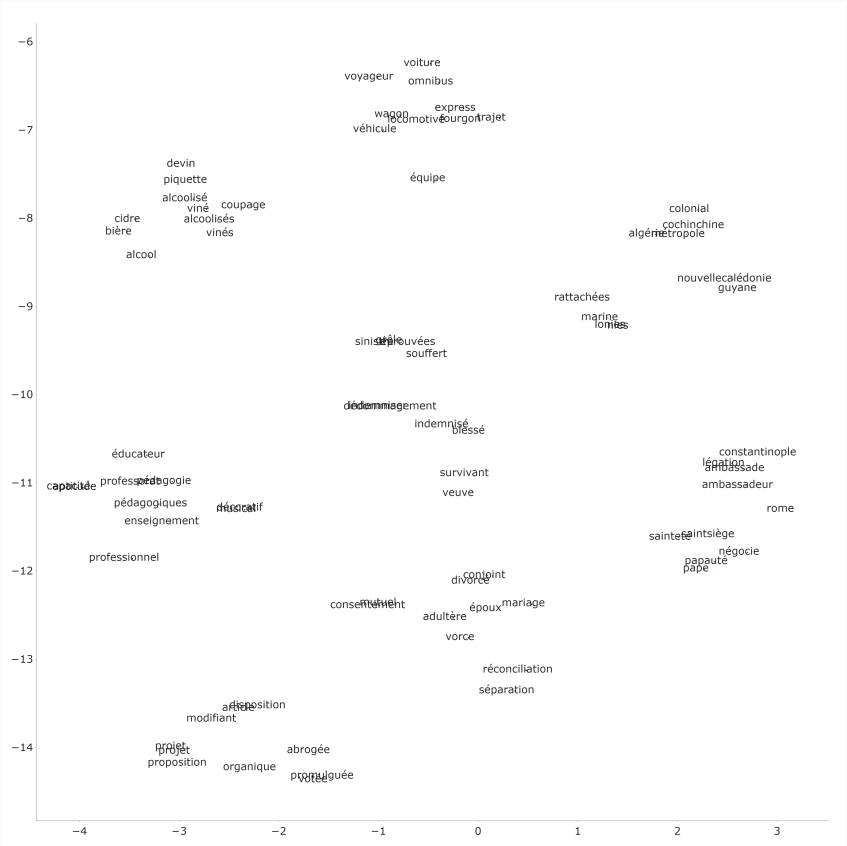
(a) 3 topics among 40 : working class (8), army (11) and infrastructures (15)



(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

THANK YOU FOR YOUR ATTENTION!

Nicolas Bourgeois : nicolas.bourgeois@epitech.eu

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

Aurélien Pellet : aurelien.pellet@epitech.eu

Marie Puren : marie.puren@epitech.eu

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

Github repository : https://github.com/mpuren/agoda