

## Rapport TP4 :

### **Exploration des politiques de représentation et d'appropriation culturelles de l'État indépendant du Congo sous le règne de Léopold II de 1885 à 1908.**

#### **fichiers tac/tp4/corpus**

Avant d'être une colonie du royaume de la Belgique jusque 1960, le Congo a été dans un premier temps une propriété privée du roi Léopold II de 1885 jusque sa mort en 1908. Durant cette période, Léopold II a exploité et importé en Belgique de nombreuses ressources naturelles du Congo, notamment le caoutchouc mais aussi le sucre, les bois précieux et l'ivoire. Leopold II, aussi appelé « roi bâtisseur », a également marqué le paysage belge de constructions architecturales coloniales dans un but de promouvoir l'intérêt commercial et culturel du Congo.

Dans ce rapport, nous allons tenter d'explorer, à l'aide de techniques de traitement automatique, quelles ont été les politiques coloniales mises en place à travers le prisme de la culture entre 1885 et 1908.

L'utilisation de techniques de traitement automatique permettent notamment de traiter du contenu textuel en données afin d'en extraire des informations analytiques. Pour cela, nous allons nous baser sur des documents textuels provenant de la librairie numérique CAMille. Celle ci regroupe des archives numériques de la presse et de l'histoire du journalisme en Belgique. Nous pourrions de cette manière interroger quelles ont été les positions des journalistes face à ces outils de propagande.

Dans le cadre de cette analyse, les techniques de traitement automatique de corpus sont utilisés comme un medium. Elles proposent une lecture de l'histoire de manière analytique. Nous verrons ensemble les avantages et les limites de ces procédures.

Pour commencer notre recherche, nous avons effectué différentes requêtes sur la plateforme CAMille afin d'affiner la sélection de nos documents sources selon notre problématique. La première chose a été de délimiter la date de publication de nos documents sur la période choisi (1885-1908). Nous avons ensuite précisé quelques mots clefs pour être plus précis : « Léopold », « Congo », « politique », « coloniale », « construction », « architecturale ».

Après avoir téléchargé nos documents nous avons fait quelques statistiques qui nous informent sur certains points.

Nos fichiers sont biens francophones, la plupart ont été écrits par le journal de L'Indépendance belge, et toutes les années entre 1885 et 1908 sont comprises dans notre corpus.

Nous avons ensuite créé un corpus par année regroupant tous les fichiers de l'année en question. Ceci va nous permettre d'analyser notre corpus pour une date précise.

Pour pouvoir effectuer une analyse textuelle de nos documents, nous devons procéder à un nettoyage de nos fichiers. Effectivement, il se peut que lors de l'océrisation des journaux numérisés il y ai eu des erreurs de reconnaissance optique par le logiciel utilisé. C'est la raison pour laquelle il est important de nettoyer son corpus avant tout. Pour ce faire nous avons procédé à une *tokénisation* de nos documents. Il s'agit là de *tokéniser*, ou autrement dit segmenter, le texte de chacun de nos fichiers en *tokens* afin de mieux pourvoir cibler les potentielles erreurs d'océrisation. Un *token* correspond à une chaîne de caractères entre deux espaces et traite les signes de ponctuations comme des items séparés. Nous avons pu ainsi nettoyer notre corpus entier et nos corpus par année à l'aide de différentes fonctions qui ont permit également de supprimer des *stopwords* (mots n'ayant peu de signification dans une phrase), ou encore les ponctuations et les chiffres.

À l'aide de nos corpus nettoyés, nous avons pu constater la fréquence des mots pour chaque année et pour tout notre corpus. Voici ce que nous avons comme fréquence de mots pour notre corpus entier :

```
[('congo', 9302), ('bruxelles', 6701), ('belgique', 6018),  
('gouvernement', 5872)]
```

Les résultats n'indiquent pas d'information particulière.

Nous avons ensuite extrait des mots clefs pour chaque année de notre corpus :

1885.txt mentions these keywords: Société anonyme, roi Léopold, rue Léopold, politique coloniale, dit qu'il, qu'il faut, Société royale...  
1886.txt mentions these keywords: Dessin ombré, prix partagé, arrêté royal...  
1887.txt mentions these keywords: ler prix, grand concours, lieu dans...  
1888.txt mentions these keywords: prix partagé, Grand concours, service militaire, Léon XIII, nouvelle loi...  
1889.txt mentions these keywords: grand nombre, van der, politique coloniale, comte van...  
1890.txt mentions these keywords: roi Léopold, politique coloniale, gouvernement belge, l'Etat Belge, Conseil provincial, Conseil communal...  
1891.txt mentions these keywords: L'impératrice Frédéric, Chambre française, prix partagé, année d'études...  
1892.txt mentions these keywords: prix partagé, l'Etat indépendant, Cercle d'escrime, VAN KERCKHOVEN, Dessin ombré...  
1893.txt mentions these keywords: Congo n'est, politique coloniale...  
1894.txt mentions these keywords: grand concours, CAPITAIN JACQUES, Hier soir, grand nombre...  
1895.txt mentions these keywords: politique coloniale, Congo belge, PARTI OUVRIER, Congo n'est, l'Etat indépendant, roi Léopold, gouvernement bel  
1896.txt mentions these keywords: nous nous, nous avons, Mais nous, Nous sommes, Congo nous, mais c'est, qu'il nous, politique coloniale, conseil  
1897.txt mentions these keywords: nous avons, nous nous, Nous sommes, nous faire, conseil communal, grand nombre...  
1898.txt mentions these keywords: Conseil communal, conseil général, année d'études, rue Royale...  
1899.txt mentions these keywords: Banque Nationale, Villes Belgique, Banque ottomane, Lots Congo, Banque Paris, Banque Coloniale, Banque France,  
1900.txt mentions these keywords: Nord Espagne, Trust Colonial, Société Générale, Banque Coloniale, Villes Belgique, Banque France, Banque Natio  
1901.txt mentions these keywords: Bruxelles dans, VILLES Belgique, Grand Prix, politique coloniale, ler prix, place dans, Congo belge, Société Be  
1902.txt mentions these keywords: ler prix, société anonyme, GRAND HOTEL, BRUXELLES Prix, villes Belgique, Société Générale, Bruxelles Banque, BF  
1903.txt mentions these keywords: Idem div, Idem ord, Société générale, Idem fond, Nord France, Société Belge, FER Congo, Bruxelles Société, Cong  
1904.txt mentions these keywords: roi Léopold, conseil communal...  
1905.txt mentions these keywords: l'Etat Indépendant, Congo belge, Petit Bleu, roi Léopold, Congo Français, d'un grand, grand nombre...  
1906.txt mentions these keywords: l'Etat indépendant, Congo belge, Chambre belge, roi Léopold, politique coloniale, Congo français, gouvernement  
1907.txt mentions these keywords: roi Léopold, Parlement belge, Congo belge, Chambre belge, gouvernement belge, qu'il faut, loi coloniale, l'Etat  
1908.txt mentions these keywords: Congo belge, gouvernement belge, politique coloniale, Chambre belge, roi Léopold, loi coloniale, PARLEMENT BELC

Nous constatons que le mot clefs « politique coloniale » revient souvent. Les mots clefs semblent correspondre aux mots clefs utilisés initialement sur la plateforme CAMille. Ils nous indiquent pas plus d'informations.

Nous avons effectué une reconnaissance d'entité nommée dans nos différents corpus. Nous allons montrer ci-dessous seulement les résultats qui nous ont intéressé.

Corpus entier :

#### Personnalités :

beernaert apparait 20 fois dans le corpus

Nous pouvons observer que Beernaert apparaît plusieurs fois. Il s'agit du premier ministre de Leopold II entre 1884 et 1894. Il a été une forte personnalité à s'opposer au roi à cause de la politique brutale menée au Congo.

Corpus 1885 :

#### Personnalités :

clemenceau apparait 16 fois dans le corpus

jules ferry apparait 9 fois dans le corpus

Georges Benjamin Clemenceau est un journaliste et un homme politique français. Il s'oppose à la colonisation, faisant tomber le gouvernement Jules Ferry sur cette question. Son combat contre le ministère Jules Ferry aboutit le 30

mars 1885 à la démission de ce dernier après le résultat négatif des votes d'un budget pour les troupes françaises au Tonkin attaqué par l'armée chinoise.

#### Organisations :

afrique association internationale congo apparait 2 fois dans le corpus

L'Association internationale du Congo est une association créée en 1879 par Léopold II. L'association est dissoute en 1885 et ses structures ont été reprises par l'État indépendant du Congo.

#### Corpus 1908 :

##### Personnalités :

vandervelde apparait 52 fois dans le corpus

schollaert apparait 14 fois dans le corpus

Émile Vandervelde est un homme politique socialiste belge, docteur en droit, en sciences sociales et en économie politique. Il fut un fervent opposant à Léopold II et au pouvoir absolu dont il jouissait au Congo durant les années 1890. Il amena à l'annexion du Congo par la Belgique en 1908.

Schollaert est le gouvernement fédéral qui dirigea la Belgique à partir de 1908.

#### Observations :

Nous pouvons constater que dans les personnalités les plus évoqués de nos corpus il existent un grand nombre de personnes ayant lutté contre la colonisation. Cependant, dans les organismes nous n'avons pas plus d'information quant à notre problématique. Il aurait été intéressant d'expérimenter cette technique avec une reconnaissance de constructions faites.

Nous allons maintenant entraîner notre corpus avec un modèle word2vec (word embeddings). Word2vec a été développé par une équipe de recherche chez Google sous la direction de Tomas Mikolov dans les années 2010. Il s'agit d'un ensemble de méthode d'apprentissage visant à représenter les mots d'un texte par des vecteurs de nombres réels. Il repose sur des réseaux de neurones artificiels à deux couches entraînés pour reconstruire le contexte linguistique des mots. De cette manière, les mots qui partagent des contextes similaires sont représentés par des vecteurs numériques proches et occupent des positions spatiales proches. Avec ceci, nous allons pouvoir connaître les mots se rapprochant de nos mots clefs initiaux.

```
model.wv.most_similar("politique", topn=10)
[36] ✓ 0.0s

... [('pensee', 0.8452361822128296),
      ('pratique', 0.8444719314575195),
      ('nation', 0.8364242315292358),
      ('politique_coloniale', 0.8321669101715088),
      ('conclusion', 0.8309684991836548),
      ('la_colonisation', 0.819184422492981),
      ('personnelle', 0.8182772994041443),
      ('volonte', 0.8175395727157593),
      ('liberte', 0.8110671639442444),
      ('conscience', 0.8079628348350525)]

model.wv.most_similar("politique_coloniale", topn=10)
[42] ✓ 0.1s

... [('solution', 0.9076110124588013),
      ('situation', 0.8912819027900696),
      ('verite', 0.8867180943489075),
      ('preoccupation', 0.8863690495491028),
      ('neutralite', 0.8842679858207703),
      ('honte', 0.8841599822044373),
      ('nation', 0.8784273862838745),
      ('these', 0.8780742883682251),
      ('affirmation', 0.8769941329956055),
      ('consequence', 0.8764239549636841)]
```

Les mots « honte » et « conséquence » ressortent de cette liste comme indications d'effets d'une politique coloniale.

Avec les résultats présents nous ne pouvons pas démontrer quoi que ce soit vis à vis de notre sujet de base. Nous pouvons seulement supposer que les journalistes de cette période mettaient en évidence certaines personnalités critiques de la politique coloniale de Léopold II.