



ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

REPORT

**Quantifying views about the future
during the Second Industrial Revolution
(1840-1920)**

BACHELOR PROJECT

Lina Bousbina and Tamara Antoun

9th June 2023

Abstract

Quantifying views about the future poses a unique challenge as it encompasses subjective and intangible concepts. In this study, we employ innovative methods, leveraging Natural Language Processing (NLP) techniques, to address this challenge. Inspired by the research conducted by Joris van Eijnatten and Pim Huijnen on reconstructing temporalities in Dutch parliamentary debates (1814–2018), we adapt their methods to analyze our corpus. Our focus lies in investigating the perceptions and views of the future during the transformative Second Industrial Revolution, spanning from 1840 to 1920. Specifically, we select three distinct regions—France, Spain, and the United States—that played significant roles in this revolution. Through our study, we aim to understand the contextual framework within which references to the future were made in newspapers of this period and explore the impact of the revolution on people's visions. Our analysis reveals a shift from an innocent perspective to a more nationalist and progressive view of the future.

“ ..or in our modern age, as it is shaped by science, technology, and industry, the future in fact implies different and new things, which cannot be entirely derived from previous experience. Precisely the impossibility of foreseeing technical inventions has become a principle derived from experience, and permanently keeps open the difference between past and future[1]. ”

Reinhart Koselleck's - The Practice of Conceptual History (2002)

CHAPTER 1

INTRODUCTION

1.1 CONTEXT

This semester project is embedded into the Marie Curie Post-Doctoral grant lead by Dr. Elena Fernández Fernández (Principal Investigator): Time, Technology, and Globalization. A study of the role of technology in processes of modernization and globalization using the Press, Big Data, and Computational Research Methodologies (GLOTECH), grant agreement No 101024996.

The article "Something Happened to the Future" [2], which explores the transformation of the future in Dutch parliamentary speech. [2], during the final quarter of the twentieth century, served as a significant inspiration for our project. Based on this study, we aim to investigate the impact of the revolution on people's visions about the future.

1.2 MOTIVATION

Newspapers have always served as vital sources of reliable information, offering insights into the prevailing issues and historical context of their time. They enable researchers to delve into the development of events and understand the dynamics of specific periods. In this study, we explore the rich potential of analyzing newspapers from the Second Industrial Revolution, a transformative era spanning from 1840 to 1920, which makes this time period historically significant and a time of interest. It indeed witnessed rapid industrialization, technological advancements, and significant socioeconomic changes.

Our focus encompasses three influential regions during this period—France, Spain, and the United States. By closely examining the newspapers, including *Le Figaro*, *El Imparcial*, and *The New York Herald*, we seek to unravel the perceptions and views of the future held within these publications.

However, we should not forget that newspapers were all written on actual paper back in the time. So, while retrieving back the given data, the Optical Character Recognition (OCR) that were used are quite noisy which makes it more difficult and challenging to analyse. 2

1.3 GOALS

The primary objective of our study is to quantify and assess the perspectives on the future that emerged during the transformative period of the Second Industrial Revolution. To achieve our research objectives, we analyze data from three prominent multilingual newspapers. Each of these newspapers holds a significant position in the historical context of the Second Industrial Revolution. In the corpus section

of our research paper, we provide comprehensive descriptions of these newspapers, emphasizing their relevance, influence, and contributions during the period we are interested in.

Our goals are to explore how references to the future were made in these newspapers and how views about the future evolved over time. By identifying common themes and analogies across the newspapers, we aim to uncover shared perspectives and concerns regarding the future, shedding light on prevailing societal sentiments and ideologies of the time. We aim to identify shifts in perspectives and explore the underlying factors that influenced these changes.

Through our analysis, we seek to contribute to a deeper understanding of the societal impact of the Second Industrial Revolution and to provide valuable insights into how technological advancements and societal changes during this period shaped people's views about the future.

CHAPTER 2

CORPUS

Newspaper	Number of Articles
Le Figaro	28,503
The New York Herald	147,709
El Imparcial	23,008

TABLE 2.1
Number of Articles per Newspaper

2.1 LE FIGARO - FRANCE



FIGURE 2.1
Figaro Logo - 1966

Established in 1826 during the reign of Charles X, *Le Figaro* [3] holds the distinction of being one of the longest-running newspapers in France. Taking its name from the beloved character Figaro in Beaumarchais' plays, the publication has carved a niche as a right-wing and center-right newspaper, attracting diverse ideological perspectives and serving as a platform for their convergence.

Studying *Le Figaro* during the Second Industrial Revolution in France, from 1840 to 1920, comprising a total of **28,503** articles(Figure 2.1), provides valuable insights into media, politics, and societal changes while quantifying anxiety about the future. As one of the oldest titles in the French press, *Le Figaro*'s coverage and influence allow us to understand the impact of rapid industrialization on public sentiment. By analyzing its portrayal of technological advancements, social disruptions, and economic transformations, we can comprehensively assess the public's perception and response to the profound changes of the era.

During our study of *Le Figaro* from 1840 to 1920, we encountered a gap in the data between 1840 and 1860. This gap prevented us from analyzing and studying the articles from that specific period.(Figure 4.4)

2.2 EL IMPARCIAL - SPAIN

Established in 1867, *El Imparcial* [4] holds the distinction of being one of the oldest and most prominent newspapers in Spain. Throughout its history, the publication has played a significant role in shaping public



FIGURE 2.2
El Imparcial Logo - 1867

opinion and providing insights into the country's political, social, and cultural landscape. Named after its commitment to impartiality, *El Imparcial* has served as a platform for diverse ideological perspectives, attracting readers from various backgrounds. Over the years, the newspaper has covered a wide range of topics, including politics, economy, arts, and sports, offering a comprehensive view of Spanish society.

Studying *El Imparcial* during the late 19th and early 20th centuries, encompassing a total of **11,776** articles, provides valuable insights into the period known as the Restoration in Spain (1874-1931). This era witnessed significant political, economic, and social changes as Spain transitioned from a monarchy to a parliamentary system. By analyzing *El Imparcial*'s coverage of these developments, we can gain a deeper understanding of the challenges, aspirations, and debates that characterized Spanish society during this transformative period. The Restoration period was marked by political tensions, economic fluctuations, cultural shifts, and societal changes, all of which are reflected in the pages of *El Imparcial*.

2.3 THE NEW YORK HERALD - UNITED STATES

The New York Herald [5], established in 1835 by James Gordon Bennett Sr., was a highly influential and widely-read newspaper in the United States. Throughout its history, it played a pivotal role in shaping American journalism and setting new standards for reporting. Covering a wide range of topics including politics, business, culture, and sports, the newspaper attracted a diverse readership and presented news from various perspectives.

Studying The New York Herald during the late 19th and early 20th centuries provides valuable insights into the transformative era of American history, including the Gilded Age, the Progressive Era, and the advent of industrialization. With a vast collection of **147,709** articles, researchers can gain a deeper understanding of societal trends, political debates, economic transformations, and cultural shifts that characterized this period. The newspaper's coverage of significant events such as the Spanish-American War, the Industrial Revolution, and the suffrage movement offers primary source material to examine public sentiment and reactions. Additionally, the newspaper's reporting on technological advancements, social inequalities, and emerging cultural trends provides insights into the changing landscape of American society.

2.4 OCR QUALITY OF THE CORPUS

To assess the quality of OCR, we computed the misspelled word count by comparing the words in the corpus to dictionaries. Specifically, we checked each word in the text against a reference dictionary (French [6], English [7], Spanish [8]) and identified any words that were not found in the dictionary as potential misspelled words.

Additionally, it's important to note that for the New York Herald dataset, we took a small subset of the data to perform the OCR assessment. Specifically, we selected the first 30,000 articles from the dataset using the `data.head(30000)` function. This subset was used to provide an overview of the OCR quality, and the results can be considered representative of the larger dataset.

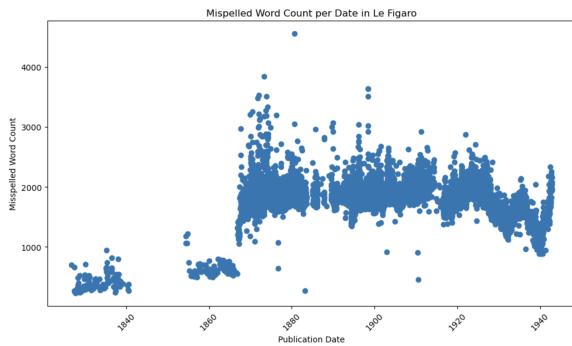


FIGURE 2.3
Misspelled Word Count in Le Figaro

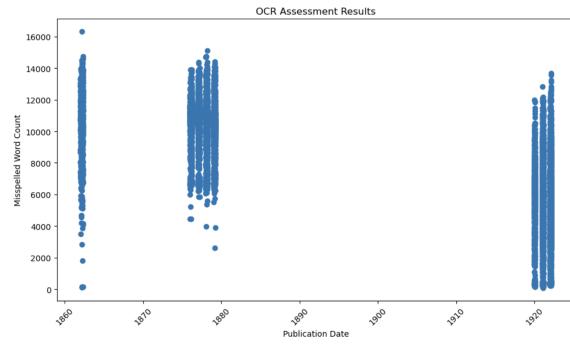


FIGURE 2.4
Misspelled Word Count in The New York Herald

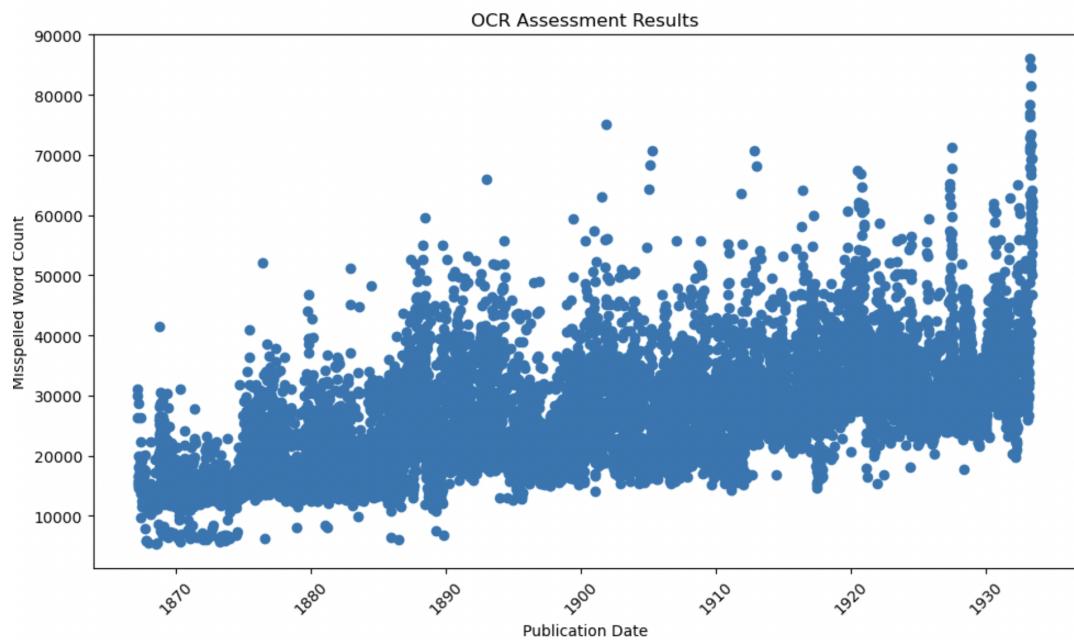


FIGURE 2.5
OCR assesments results for El Imparcial

After conducting the OCR assessment on the New York Herald dataset, we observed a significantly higher count of misspelled words compared to the other newspaper, Le Figaro. The misspelled word count in the New York Herald dataset reached a maximum of 16,000 misspelled words in a single article, whereas the maximum count in Le Figaro was only 4,000. This stark difference in misspelled word counts suggests a higher level of OCR errors or poor text quality in the New York Herald dataset. Consequently, our analysis focused more on Le Figaro due to the limitations posed by the OCR quality in the New York Herald dataset. It is important to note that meaningful results could not be obtained for the New York Herald dataset due to these limitations.

The count of misspelled words in Spanish articles is significantly higher compared to French and American ones, highlighting a noticeable difference. In fact, the count reaches 90,000, which is quite substantial. This disparity can be due to a couple of factors. Firstly, the Spanish dictionary used might not include accents, which are crucial in the Spanish language. Secondly, the variation in language over time could also contribute to this contrast.

CHAPTER 3

METHODS

For each of the three corpus we used the same methods trying to replicate "Reconstructing Dutch Parliamentary debates" study conducted by Joris van Eijnatten and Pim Huijnen.[2]

In our project, we conducted our analysis using Python programming language and Jupyter Notebook as the primary environment. We leveraged various Natural Language Processing (NLP) methods to delve into the study of the selected corpus. To support our computations, we relied on several essential libraries, including:

- NumPy for numerical computation and array operations
- Panda for data manipulation, preprocessing and analysis
- NLTK (Natural Language Toolkit) for tokenization, stopword removal ...
- Matplotlib for data visualization
- Python-tesseract for Optical Character Recognition.

One of the challenging aspects of our project was the preprocessing of the data before we could effectively utilize it. This involved performing specific tasks for each newspaper dataset:

Le Figaro:

We had to deal with JSON files, thus, we implemented a conversion process to transform the JSON files into CSV format. This conversion allowed us to efficiently handle and analyze the data within the CSV files, enabling easier manipulation and extraction of relevant information.

El Imparcial:

In order to handle the data for the Spanish newspaper, we had txt files initially. However, we found that the CSV format was more convenient for our work, so we proceeded to convert the txt files into CSV. This conversion has once again simplified the data manipulation process for us.

New York Herald:

The data was packaged in a compressed tar file. To access and work with the data set, we needed to unzip the tar file. This process involved extracting the files from the compressed format, allowing us to access the individual files and perform further processing and analysis.

To streamline our analysis process, we adopted a straightforward approach of creating a separate CSV file for each newspaper. This enabled us to leverage the CSV files as data frames, facilitating various computations and analysis tasks. By starting with CSV files, we were able to conveniently filter the data

based on specific criteria, tokenize the text, and apply a range of methods to gain insights from the data set.

3.1 FILTER THE CORPUS

We started by filtering our corpus according to the key-words:

- futur
- nation
- president
- Country's name : United States - France - España

After applying the filtering process to the corpus, we were able to gain insights into the number of articles that contain specific words. This filtering step was of utmost importance, indeed, considering that we are studying newspapers spanning a period of 80 years, the dataset is inherently vast. Therefore, it was crucial to narrow down the articles to those that are most relevant for our further analysis.

Additionally, the filtering process allowed us to compare the frequency of articles containing each word and ascertain the relative importance of the words under examination. This comparison gave us a valuable perspective on the prominence and relevance of the specific words in the context of the corpus as it facilitated assessments of the significance of the studied words.

3.2 COUNTING RAW FREQUENCY OF THE KEY-WORDS

In the subsequent stage of our project, we focused on calculating the raw frequency of each word in the respective newspapers. This involved analyzing the occurrence of each word throughout the dataset and tracking its chronological appearance. By employing linear regression, we were able to plot the raw frequency of each word as a function of the date, allowing us to observe the temporal trends and fluctuations in word mentions.

After loading the data into a pandas DataFrame, we sorted it chronologically based on the publication date. Then, we converted the publication date column from string format to a date format. By applying a custom function, we created a new column in the DataFrame to store the raw frequency of the word in each article. The linear regression allowed us to gain insights into the temporal patterns of word occurrences throughout the studied period.

3.3 TOKENIZATION

To begin, the data for each newspaper was divided into multiple blocks to ensure manageable processing. For Le Figaro and The New York Herald, the data was split into four blocks:

- 1840 - 1860
- 1860 - 1880
- 1880 - 1900
- 1900 - 1920

While for El Imparcial, it was divided into two blocks:

- 1867 - 1900

- 1900 - 1933

Each block represented a specific time period of interest within the overall study period. To address the uneven distribution of articles across the study period due to multiple sources and archives, we divided the data into distinct time periods based on data availability in our corpus. This approach ensured a representative sample within each period.

Once the data was divided into blocks, tokenization was applied to each block using the

```
nltk.word_tokenize
```

method.

This step aimed to convert the text into a sequence of individual tokens.

Following tokenization, we extracted the content from each block, filtered it based on the defined date ranges, and tokenized it. The resulting tokens were stored in a new column named "tokens" within each respective block of the DataFrame.

3.4 COUNTING THE TWO-GRAMS BEFORE AND AFTER

In this part of the project, we focused on analyzing the occurrences of two-grams (consecutive word pairs) before and after a target word. This analysis was performed individually for the three newspapers: *Le Figaro*, *The New York Herald*, and *El Imparcial*. Using the tokenized text from each block, we initialized empty lists to store the two-grams before and after the target word. Next, we processed the text by removing punctuation marks and digits, and then identified the position of the target word within the document. When the target word was found, we extracted the adjacent words to create a two-gram, which was added to the corresponding list. The frequency distribution of these two-grams was computed and saved in a file. Lastly, we printed the occurrences and the most common two-grams before and after the target word.

3.5 CONCORDANCE ANALYSIS

To gain a better understanding of how the word "future" is used within the text corpus, we employed concordance analysis. This method allowed us to examine the context in which the word appears and provided us with a key-word in context view. By creating lists of token strings from the text blocks and generating Text objects, we were able to iterate over the concordance lists for the words we study in each block. This process enabled us to extract and display the lines containing the word, shedding light on its usage and providing valuable insights into its context within the studied newspapers.

3.6 TOPIC MODELING

We used Elisa Michelet script to perform Topic Modeling method.[9] The code implements topic modeling using the tomotopy library in Python. It focuses on applying topic modeling to *Le Figaro* newspaper articles from 1840 to 1920. The code preprocesses the text, creates a corpus, and trains multiple topic models with different parameters. It selects the best model based on the coherence score and generates a heatmap to visualize the scores. Finally, it saves the topics generated by the best model in a text file. Overall, the code identifies and analyzes topics related to the keyword "futur" in the newspaper articles using topic modeling.

To present the results and findings for the french dataset, we will analyze the topics generated through the topic modeling process and provide an explanation of the discovered themes. Our objective is to

quantify the views and perspectives about the future by examining the topics and associated keywords. By understanding the main themes and subjects covered in each newspaper, we can gain insights into how the Second Industrial Revolution influenced societal perceptions and views about the future.

3.7 OTHER METHODS

During our analysis of the french dataset, we implemented a method to count the occurrence of 5-grams that included the word "futur" and a method that counts the occurrence of the tri-gram "le future est". However, these approaches did not yield meaningful or relevant results in our study. We also examined the raw frequency of word combinations, specifically "futur" with "president" or "futur" with "nation." Unfortunately, these combinations did not provide significant insights for our research objectives. Consequently, we decided not to pursue further analysis using these methods on our corpus.

CHAPTER 4

RESULTS

In this section, we will present the results of our analysis, focusing on quantifying the views of the future depicted in the selected newspapers during the second industrial revolution. Through comparative analysis, we have identified recurring themes and insights into people's perspectives and expectations regarding the future. Our objective is to gauge the level of optimism, concerns, and anxieties prevalent during this transformative era by analyzing the frequency and prominence of specific topics and keywords related to the future. By presenting these findings, we aim to enhance our understanding of the societal impact of the second industrial revolution and gain valuable insights into the cultural and intellectual contexts of that time.

4.1 FILTERING AND COUNTING RAW FREQUENCY

	Le Figaro	The NYH	El Imparcial
futur	17,102	60,444	11,776
avenir	23,034	-	-
nation	27,946	125,006	22,678
president	855	27,759	22,049
country	15,113	88,150	22,685

TABLE 4.1
Number of Articles per Word for Each Newspaper

	Le Figaro	The NYH	El Imparcial
futur	60%	41%	51%
avenir	81%	-	-
nation	98%	84.6%	98.5%
president	3%	17.5%	95.8%
country	53%	60%	98.5%

TABLE 4.2
Proportion of Articles per Word in Each Newspaper

Following our computational analysis, we successfully generated plots illustrating the raw frequency and occurrences of the words we analyzed across the three different newspapers.

Figure 4.2 reveals a noticeable gap in the data between 1880 and 1920, indicating missing or unavailable

Raw frequency of the word "nation" according to Linear regression

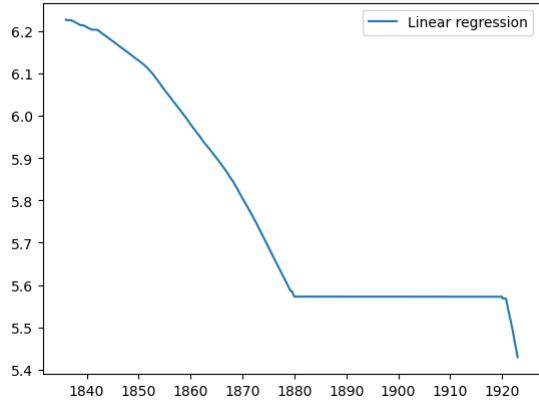


FIGURE 4.1
Raw Frequency of the Word "Nation" in the New York Herald

Number of occurrences of the word "nation"

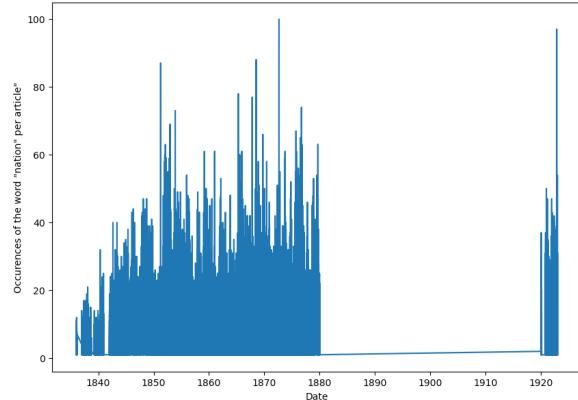


FIGURE 4.2
Occurrences of the Word "Nation" in the New York Herald

information regarding the occurrences of the word "Nation" within the New York Herald during that period. This data gap is consistent across all analyzed words within the newspaper.

In Figure 4.1, we observe a distinct decrease in the frequency of the word "Nation" in the articles during the analyzed period.

Raw frequency of the word "avenir"

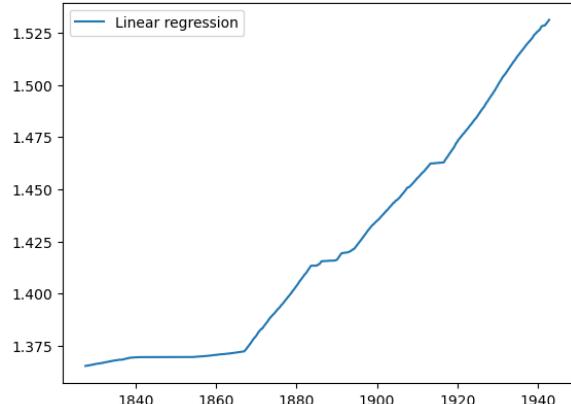


FIGURE 4.3
Raw Frequency of the word "Avenir" in Le Figaro

Number of occurrences of the word 'Avenir'

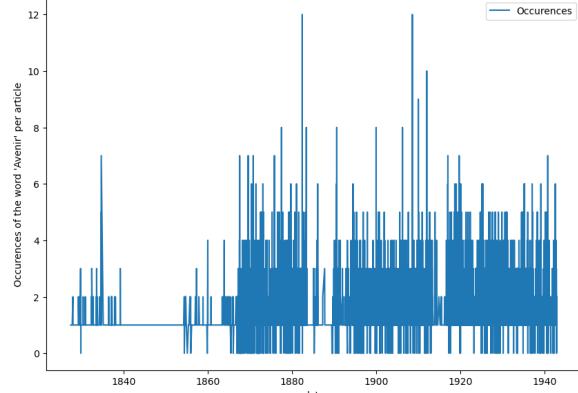


FIGURE 4.4
Occurrences of the word "Avenir" in Le Figaro

In figure 4.3 we can see the clear growth of the frequency of the word "Avenir" as time evolves

In figure 4.4 we can observe a gap from 1840 to 1860 which means there is missing data during this time span. It is important to note that this gap is present in this period in the whole Figaro corpus.

4.2 COUNTING THE TWO-GRAMS BEFORE AND AFTER

Most common 2-grams before:

```
[(("qu", "à"), 53), (("musique", "de"), 27), (("et", "de"), 20), (("et", "dans"), 16), (("ce", "que"), 15), (("confiance", "dans"), 13), (("présent", "et"), 12), (("l'homme", "de"), 12), (("passé", "et"), 11), (("que", "de"), 9), (("et", "à"), 9), (("et", "pour"), 9), (("et", "que"), 8), (("plus", "à"), 8), (("lire", "dans"), 6), (("de", "prévoir"), 6), (("éventualités", "de"), 6), (("pas", "à"), 5), (("foi", "dans"), 5), (("nous", "réserve"), 5), (("rassuré", "sur"), 5), (("et", "sur"), 5), (("", "A"), 5), (("espérance", "pour"), 5), (("théâtre", "de"), 5), (("douter", "de"), 4), (("répondre", "de"), 4), (("quant", "à"), 4), (("que", "soit"), 4), (("musicien", "de"), 4), (("vue", "de"), 4), (("que", "dans"), 4), (("être", "à"), 4), ("a", "de"), 4), (("pas", "de"), 4), (("sans", "engager"), 4), (("pour", "assurer"), 4), (("donc", "à"), 3), (("même", "dans"), 3), (("rien", "à"), 3), (("soucieux", "de"), 3), (("certitudes", "de"), 3), (("qu", "a"), 3), (("politique", "de"), 3), (("songe", "à"), 3), (("désespérer", "de"), 3), (("vision", "de"), 3), (("projets", "pour"), 3), (("préoccuper", "de"), 3), (("le", "présent"), 3), (("garantis", "pour"), 3), (("moins", "pour"), 3), (("que", "à"), 3), (("sent", "et"), 3), (("artistes", "de"), 3), (("confiance", "en"), 3), (("mettre", "à"), 3), (("hommes", "de"), 3), (("peur", "de"), 3), (("assurer", "dans"), 3), (("intérêt", "de"), 3), (("portes", "de"), 3), (("à", "qui"), 3), (("espérances", "de"), 3), (("L'homme", "de"), 3), (("sique", "de"), 3), (("d'éviter", "à"), 3), (("sûr", "de"), 3), (("sécurité", "pour"), 3), (("que", "pour"), 3), (("historiens", "de"), 3), (("Musique", "de"), 3), (("serait", "à"), 2), (("rassure", "sur"), 2), (("académicien", "d'e"), 2), (("député", "de"), 2), (("songeant", "à"), 2), (("que", "si"), 2), (("l'espérance", "de"), 2), (("désespér
```

FIGURE 4.5

Most common 2-grams before "l'avenir" in 1860-1880

One of the most interesting aspects revealed by this analysis is the recurring presence of the 2-gram "musique de" and the combinations like "théâtre de" and "artistes de". This suggests that the text explores various aspects of music, and a focus on theater and artists respectively. These indicate a broader exploration of artistic expressions, creative endeavors, and the cultural impact of art forms during that time span.

Most common 2-grams before:

```
[(("qu", "à"), 31), (("et", "de"), 20), (("confiance", "dans"), 15), (("ce", "que"), 13), (("et", "dans"), 12), (("foi", "dans"), 12), (("présent", "et"), 11), (("passé", "et"), 10), (("et", "pour"), 9), (("historiens", "de"), 8), (("celle", "de"), 7), (("que", "dans"), 7), (("soucieux", "de"), 6), (("musique", "de"), 6), (("secret", "de"), 6), (("vue", "de"), 6), (("Français", "de"), 6), (("quant", "à"), 5), (("et", "dont"), 5), (("et", "que"), 5), (("confidence", "en"), 5), (("augure", "pour"), 4), (("croit", "à"), 4), (("pas", "de"), 4), (("et", "sur"), 4), (("tion", "de"), 4), (("désespérer", "de"), 4), (("leçon", "pour"), 4), (("plus", "à"), 4), (("maîtres", "de"), 4), (("songer", "à"), 4), (("réserve", "pour"), 4), (("rassuré", "pour"), 4), (("décider", "de"), 3), (("garantis", "pour"), 3), (("sécurité", "pour"), 3), (("voir", "dans"), 3), (("solutions", "de"), 3), (("croire", "à"), 3), (("point", "de"), 3), (("pour", "assurer"), 3), (("souci", "de"), 3), (("l'armée", "de"), 3), (("tranquille", "sur"), 3), (("chances", "de"), 3), (("réparations", "de"), 3), (("éventualités", "de"), 3), (("sécurité", "et"), 3), (("projets", "pour"), 3), (("s'agit", "de"), 3), (("seront", "à"), 3), (("pas", "dans"), 3), (("sera", "dans"), 3), (("crois", "que"), 3), (("foi", "en"), 3), (("yeux", "vers"), 3), (("avantage", "dans"), 3), (("teurs", "de"), 3), (("intentions", "pour"), 3), (("Prévoyants", "de"), 3), (("peur", "de"), 2), (("que", "pour"), 2), (("espérance", "pour"), 2), (("craines", "sur"), 2), (("à", "cœur"), 2), (("ni", "dans"), 2), (("cest", "que"), 2), (("désespérant", "de"), 2), (("rester", "dans"), 2), (("rassuré", "sur"), 2), (("fiance", "en"), 2), (("vision", "de"), 2), (("qui", "représent"), 2), (("fon", "der"), 2), (("qui", "sont"), 2), (("que", "soit"), 2), (("confiant", "en"), 2), (("avoir", "pour"), 2), (("empêcher", "à"), 2), (("la", "France"), 2), (("menaces", "pour"), 2), ((.., "Pour"), 2), (("marcher",
```

FIGURE 4.6

Most common 2-grams before "l'avenir" in 1880-1900

The presence of "confiance dans" and "foi dans" suggests a discussion about trust and faith.

We can observe again the 2-gram "musique de", emphasizing the relevance of music in the text which indicates a focus on musical aspects.

CHAPTER 4 – RESULTS

Most common 2-grams after:

```
[(("la", "France"), 14), (("notre", "pays"), 11), (("dans", "le"), 9), (("de", "la"), 9), (("la", "race"), 9), ((pays, '.'), 8), (("la", "civilisation"), 8), ((de, "notre"), 7), (("la", "Russie"), 7), (("notre", "race"), 6), ((il, "ne"), 6), (("", "a"), 6), ((son, "pays"), 5), ((le, "passé"), 5), ((la, "République"), 5), ((il, "est"), 5), ((la, "patrie"), 4), ((il, "faut"), 4), ((il, "ne"), 4), ((confiance, " "), 4), ((la, "nation"), 4), ((pays, " "), 4), ((a, "la"), 4), ((il, " "), 4), ((et, "qui"), 4), ((le, "présent"), 4), ((Il, "est"), 4), ((", "et"), 4), ((l'humanité, ". "), 4), ((la, "Roumanie"), 4), ((mais, " "), 4), ((-", "et"), 3), ((mais, "il"), 3), ((et, "c'est"), 3), ((et, "le"), 3), ((et, "si"), 3), ((et, "la"), 3), ((non, "seullement"), 3), ((En, "effet"), 3), ((nos, "enfants"), 3), ((*, "*"), 3), ((qui, "est"), 3), ((notre, "patrie"), 3), ((aux, "plus"), 3), ((l, " "), 3), ((la, "religion"), 3), ((., "Et"), 3), ((que, "les"), 3), ((on, "ne"), 3), ((et, "de"), 3), ((plein, "de"), 3), ((a, "Dieu"), 3), ((passé, " "), 3), ((Il, "y"), 3), ((Et, "le"), 3), ((tous, "les"), 3), ((une, "confiance"), 3), ((toutes, "les"), 2), ((vers, "le"), 2), ((et, "ils"), 2), ((n'en, "sera"), 2), ((la, "dy"), 2), ((., "Les"), 2), ((;, "elle"), 2), ((Mais, " "), 2), ((., " "), 2), ((sous, "les"), 2), ((l'Eglise, "et"), 2), ((., "il"), 2), ((Un, "jour"), 2), ((toujours, "une"), 2), ((Ces, "garçons"), 2), ((cours, "de"), 2), ((je, "vous"), 2), ((L'e), 2), ((l'hu, "manité"), 2), ((Elle, "est"), 2), ((de, "l'Europe"), 2), ((a, "personne"), 2), ((les, "deux"), 2), ((de, "se"), 2), ((un, "programme"), 2), ((lieu, "de"), 2), ((qui, "se"), 2), ((., " "), 2), ((., "Nous"), 2), ((., " "), 2), ((plus, "lointain"), 2), ((la, "force"), 2), ((tant, "que"), 2), ((a, "
```

FIGURE 4.7

Most common 2-grams after "l'avenir" in 1900-1920

The pair "notre pays" (our country), implies a sense of ownership to a specific nation and suggests a discussion about the a homeland or a shared national identity. Overall, the analysis shows discussions about France, national identity, and topics related to race and ethnicity.

Most common 2-grams after:

```
[(("la", "France"), 15), (("dans", "le"), 13), (("notre", "pays"), 11), (("la", "race"), 10), (("la", "Russie"), 8), ((la, "civilisation"), 8), ((le, "passé"), 7), ((de, "notre"), 7), ((n, "y"), 6), ((notre, "race"), 6), ((la, "République"), 6), ((a, "la"), 6), ((de, "la"), 6), ((son, "pays"), 5), ((la, "nation"), 5), ((le, "présent"), 5), ((la, "patrie"), 4), ((il, "ne"), 4), ((nos, "enfants"), 4), ((ne, "faut"), 4), ((la, "paix"), 4), ((que, "le"), 4), ((une, "confiance"), 4), ((qu'il, "faut"), 3), ((de, "se"), 3), ((lieu, "de"), 3), ((n, "a"), 3), ((cet, "avenir"), 3), ((notre, "patrie"), 3), ((ce, "qui"), 3), ((aux, "plus"), 3), ((la, "religion"), 3), ((ceux, "qui"), 3), ((la, "loi"), 3), ((la, "victoire"), 3), ((a, "Dieu"), 3), ((y, "a"), 3), ((la, "Roumanie"), 3), ((monde, "et"), 3), ((vers, "le"), 2), ((ils, "se"), 2), ((n'en, "sera"), 2), ((la, "dy"), 2), ((faut, "que"), 2), ((sociétés, "humaines"), 2), ((a, "son"), 2), ((n'est, "pas"), 2), ((sous, "les"), 2), ((l'Eglise, "et"), 2), ((toujours, "une"), 2), ((la, "plus"), 2), ((garçons, "en"), 2), ((cours, "de"), 2), ((où, "le"), 2), ((l'hu, "manité"), 2), ((de, "l'Europe"), 2), ((a, "personne"), 2), ((la, "domination"), 2), ((est, "plain"), 2), ((quand, "les"), 2), ((plus, "lointain"), 2), ((ne, "le"), 2), ((est, "un"), 2), ((de, "nous"), 2), ((de, "ces"), 2), ((notre, "marine"), 2), ((et, "qui"), 2), ((dépit, "de"), 2), ((de, "nos"), 2), ((Et, "nous"), 2), ((il, "était"), 2), ((la, "question"), 2), ((présent, "n'est"), 2), ((que, "les"), 2), ((n'est, "pourtant"), 2), ((la, "locomotion"), 2), ((a, "nous"), 2), ((jour, "vien"), 2), ((tous, "ceux"), 2), ((plain, "de"), 2), ((de, "ne"), 2), ((ce, "pays"), 2), ((char, "gera"), 2), ((ce, "que"), 2), ((Les, "deux"), 2), ((dans, "son"), 2), (
```

FIGURE 4.8

Most common 2-grams after "futuro" in 1867-1900

This analysis suggests a focus on political leadership and presidency in the context. It also implies a focus on legislative bodies or specific instances involving Congress.

Most common 2-grams after:

```
[(("de", "su"), 64), (", ", "y"), 39), ((de, "la"), 31), ((presidente, "de"), 28), ((presidente, "del"), 25), ((., "El"), 21), ((u, "n"), 15), ((P, "a"), 15), ((jefe, "del"), 14), ((Congreso, "de"), 14), ((., "la"), 13), ((", "que"), 13), ((., "l"), 12), ((Congreso, ". "), 12), ((jefe, "de"), 11), ((., "Y"), 11), ((d, "e"), 11), ((., "No"), 10), ((., "no"), 10), ((., "al"), 10), ((., "pero"), 9), ((ministro, "de"), 9), ((., "en"), 9), ((Gobierno, ". "), 9), ((p, "a"), 8), ((suegro, ". "), 8), ((régimen, "de"), 8), ((desarrollo, "de"), 8), ((proyecto, "de"), 8), ((., " "), 8), ((., "En"), 8), ((., "la"), 8), ((de, "l"), 8), ((., "es"), 7), ((., "A"), 7), ((., " "), 7), ((., "Se"), 7), ((hijo, "político"), 7), ((., "De"), 7), ((yerno, " "), 7), ((., "como"), 7), ((Congreso, ". "), 7), ((próximo, ". "), 7), ((., "el"), 7), ((., "Si"), 6), ((m, "á"), 6), ((l, "a"), 6), ((., "Los"), 6), ((., "la"), 6), ((m, "a"), 6), ((., "a"), 6), ((más, "o"), 6), ((yerno, ". "), 5), ((esposo, "de"), 5), ((presupuesto, ". "), 5), ((., "p"), 5), ((., "pues"), 5), ((do, "la"), 5), ((y, "que"), 5), ((., " "), 5), ((autor, "de"), 5), ((Parlamento, ". "), 5), ((h, "a"), 5), ((esposo, ". "), 5), ((., "con"), 5), ((., "porque"), 5), ((de, "las"), 5), ((., "P"), 5), ((., "se"), 5), ((Madrid, ". "), 5), ((a, "r"), 4), ((., "cuando"), 4), ((., "por"), 4), ((cuñado, ". "), 4), ((rey, "de"), 4), ((ministerio, ". "), 4), ((., "E"), 4), ((Cónclave, ". "), 4), ((., "si"), 4), ((desarrollo, ". "), 4), ((., "pero"), 4), ((., " "), 4), ((suegro, ". "), 4), ((rey, ". "), 4), ((., "Por"), 4), ((gobierno, ". "), 4), ((presidente, "d"), 4), ((a, "l"), 4), ((presidente, ". "), 4), ((Tratado, "con"), 4), ((Parlamento, ". "), 4), ((Tratado, "de"), 4), ((., "Hay"), 4), ((., "del"), 4)
```

FIGURE 4.9

Most common 2-grams after "futuro" in 1900-1933

The repeating occurrences of "president" shows a focus on politics too. It highlights the importance of political leadership in the text, suggesting a focus on the actions, impact, and significance of presidents in the context.

4.3 CONCORDANCE ANALYSIS (KEY-WORD-IN-CONTEXT)

The concordance analysis provides snippets of text containing the word "futur" in each block of the Industrial Revolution period (1840-1860, 1860-1880, 1880-1900, and 1900-1920). The purpose of this analysis is to understand how the concept of the future was discussed and viewed during this period. Here's a snippet of the first period and the 4th period:

rompu . Le prince de Paradoxoff , le futur beau-père , a déclaré qu'il ne donne direj des farces . Il entra chez son futur beau-père par-dessus les mûrs , pour à celle qui n ' a élé comtesse qu'au futur passé . Une proche parente de Mlle H lus tard Napoléon * ne permit pas au futur au teur de la Comédie humaine de ren oisifs d'épuiser leur malignité . Le futur est riche , très-riche- , - il recon t une petite bourse de 20,000 fr.-Le futur n ' a rien , comme il con vient aux dans lesquels la dame a remercié son futur bien faiteur , '' lorsqu'il lui anno au fond de l ' I sère pour prêter au futur orateur le secours de son expé rienc le plan d'éducation an térieure d'un futur homme d'État . ♦ # * J'ai connu , il ux- , et avant toute discussion , le futur acquéreur avait com- * 4 '' ? Année nvirons de Beauvais . Un jour que le futur artiste extrayait de la terre en cha vec l'amour qu'elle éprouve pour son futur . Vous voyez d'ici la lutte qui s'en sa tante la marquise de Savenay , le futur , M. et madame d & Guibert , le beau o va la servir à souhait . Un ami du futur de madame de Chatenay , Léonce Champ rès difficile sur le chapitre de son futur gendre . Qu'il soit jeune , beau , é e prise de tabac significa tive ; le futur sort la tête basse en étouffant un s lontaire.-Montluc et le bracelet.-Un futur in- -complet.-Ecrivain et soldat . -

FIGURE 4.10
Key-word-in-context of the word 'futur' between 1840 - 1860 in Le Figaro

ur dans ses t lancs : Maximilien , le futur empereur au Mexique , , qui entra da de la Compagnie des Omnibus et leur futur président . Et nous sommes certain q Maiit Le baron Rapineau écrit à son futur gendre et commence ainsi : « Cher mo tte et Laffitte , avec entrée sur le futur boule vard Haussmann . - Mais le tem ffrage universel . ; A son programme futur : la réforme du Code pénal , - natur . Il y a. des curieux jusque dans le futur bassin où l'eau tombera en nappes im e , car elle ne fait qu'appliquer au futur ' ce qui s'est passé depuis l'appa r .-Le Saint-Siège et la France . - Le futur Pape . De retour de Rome , je vois q ssayait de se représenter le tableau futur que je venais d'imaginer , et il dit , fatigué , ne res tera pas dans le futur cabinet . Il faudra d'ici là lui che occupés à achever des levés pour le futur chemin de fer devant aller d'Hanoï à le portrait du prince de Galles , le futur George III . Ce divin portrait de fe * Telle est la composition de notre futur corps expéditionnaire . Ces officier ent d'un certain nombre de salles du futur musée'des Arts décoratifs , salles e nos représen tants , M. Canibon , le futur gouverneur générât de l'Algérie , qu ement une suite . Mais cela c'est le futur . Ne parlons que du pré- sent. . _ , -on remplacer ce conditionnel par un futur , et dire que les employés s'échappe qu'il visait , , en rhétorique , Son futur Charivari , - était , par . excellen , mises en cir culation au sujet du futur mouvement épiscopal . .Un de nos con

FIGURE 4.11
Key-word-in-context of the word 'futur' between 1900 - 1920 in le Figaro

Block 1 (1840-1860):

The snippets touch upon the concept of a "futur beau-père" (future father-in-law) in a story, highlighting the anticipation of familial relationships. They also mention someone's future elevation to the status of a countess, hinting at social aspirations. Additionally, the discussions revolve around the future success and wealth of individuals, showcasing the importance placed on personal achievements and prosperity.

Block 4 (1900-1920):

The snippets highlight discussions on the future impact of actions, plans for infrastructure projects like railways and museums, and aspirations for future roles. They reflect a forward-thinking mindset during the Industrial Revolution, where consequences, progress, and personal ambitions were central to shaping the path ahead.

These distinct blocks offer a glimpse into the evolving perspectives on the future during significant periods of history, one focusing on personal and social aspects, and the other on progress, infrastructure, and ambitious aspirations

4.4 TOPIC MODELING

```
Topic #0
    → front ennemi allemand artillerie prisonnier attaque secteur combat troupe ennemis
Topic #1
    → cardinal pape eglise catholique abbé évêque prêtre religieux xiii rome
Topic #2
    → thier monsieur rossini mademoiselle hugo votant girardin buffet Monsieur victor
Topic #3
    → allemand allemagne allié traité paix lloyd conférence poincarer georg wilson
Topic #4
    → arbre jardin soleil paysage montagne forêt automne parc fleur monument
Topic #5
    → comte mlle comtesse princesse marquis baron carnot prince baronne duchesse
Topic #6
    → gambetta grévy républicain thier conservateur ferry dufaure maréchal gauche freycinet
Topic #7
    → hitler blum page daladier laval stavisky mussolini colonne eden reich
Topic #8
    → comédie théâtre écrivain talent auteur roman dramatique volume pièce comédien
Topic #9
    → mariage marquis princesse prince empereur duchesse comte comtesse fille marier
Topic #10
    → pont voyageur bord port anglais navire catastrophe bâtiment août mètre
Topic #11
    → budget mars projet intérêt principe nécessaire janvier accorder soumettre crise
Topic #12
    → million impôt dépense franc chiffre produit recette industrie payer revenu
Topic #13
    → avocat juge procès magistrat crime justice police prison enquête cour
Topic #14
    → turc pacha sultan turquie grec grèce serbe egypte orient constantinople
Topic #15
    → sang terrible crier mourir vieillard douleur frère silence chien regarder
```

FIGURE 4.12
Topic Modeling of the word 'futur' in le Figaro

The topic modeling analysis of the french dataset resulted in a diverse range of topics. These topics cover various subjects such as military conflicts, religion, politics, nature, social status, arts and culture, economy, law, and international relations. While there is no specific topic explicitly focused on the future, discussions related to that are likely embedded within different topics, including technological

advancements, social change, and economic forecasts. The generated topics provide valuable insights into the views and interests discussed in the articles, showcasing the multidimensional nature of the dataset.

The results obtained from the *Figaro* dataset for this method did not meet our expectations or align with our research objectives, leading us to exclude the application of topic modeling on the other datasets.

CHAPTER 5

DISCUSSION

5.1 RAW FREQUENCY

The figure 4.3 shows a clear and important rise of the frequency of the word "avenir". The increasing frequency of this word suggests that Le Figaro may have placed a greater emphasis and interest on discussions related to the future during this period. The growth of "avenir" in Le Figaro could also be linked to the emergence of social and political movements during that time. The period saw the rise of ideologies such as liberalism, socialism, and nationalism, which often included discussions about the future direction of society.

The analysis revealed a similar growth pattern for the words "futur" , "nation" , and "président" (president) in the French newspaper dataset. Likewise, a comparable increase was observed for the words "futuro" , "nación" , and "presidente" in the Spanish newspaper dataset. However, in the case of The New York Herald, while the word "futur" showed an increase, there was a decrease in the frequency of the word "nation."

5.2 TWO-GRAMS BEFORE AND AFTER

The analysis of the frequency of the word "avenir" in the French newspaper, Le Figaro, from 1840 to 1940, provides valuable insights into the evolving themes and interests associated with the term.

The selection of these three specific blocks was based on their relevance in showcasing the significant changes in the contextual usage of the keyword "avenir." By examining these blocks, we can clearly observe a notable shift in the main topics and discussions in which the word appears.

Initially, in the second block, 4.5, the predominant context surrounding the word "avenir" revolves around musical and artistic themes. This suggests that during this period, Le Figaro explored the future prospects of the music and arts scene.

However, as we move to the third block, 4.6, a discernible change in the main topic becomes evident. The contextual usage of "avenir" shifts towards a more political dimension. The discussions now revolve around political events, social movements, and the future of the nation. This transition indicates a broader shift in societal interests and concerns, with the readership becoming more engaged in political matters and expressing a growing interest in the future direction of the country.

The most significant transformation occurs in the fourth block, 4.7, where the frequency of "avenir" in a political and nationalistic environment is particularly pronounced. This block marks a clear turning point in the newspaper's coverage.Indeed, the readers' attention shifts from the artistic and cultural dimension

to matters of national importance, emphasizing a growing interest in political developments and the long-term path of the country.

5.3 CONCORDANCE ANALYSIS

The concordance analysis of the word "futur" provides fascinating insights into how the concept of the future was discussed and perceived during different periods of the Industrial Revolution. In Block 1 (1840-1860), discussions revolved around the anticipation of familial relationships, social aspirations, and personal success and wealth. These findings indicate a strong emphasis on individual progress and societal advancement during that time.

In contrast, Block 4 (1900-1920) showcases a notable shift in perspectives towards a more forward-thinking mindset. Discussions in this period focused on the future impact of actions, plans for infrastructure projects, and ambitious aspirations. This suggests a broader nationalization of the concept of the future, where the collective progress, consequences, and ambitious goals took center stage.

The concordance analysis, along with the presented snippets Figure 4.10 and Figure 4.11 and offers a rich dataset that captures the essence of discussions about the future and its role in driving progress and national development.

5.4 TOPIC MODELING

The absence of a specific topic dedicated to the future highlights the need for further analysis to explore views about the future. By combining the topic modeling results with sentiment analysis and considering the context of the articles, we can gain a deeper understanding of sentiments and perspectives regarding the future. This approach allows us to examine how the french dataset captures views on the future across different subjects.

CHAPTER 6

CONCLUSION

In conclusion, our analysis of *Le Figaro*, as well as *El Imparcial* and *The New York Herald* datasets, has revealed interesting trends in the usage of words related to the future. Our research project was motivated by the concept of social acceleration of time, as discussed in the paper "Measuring the Acceleration of the Social Construction of Time" [13]. We aimed to explore and quantify the societal anxiety about the future over time. However, our analysis revealed a different trend. We observed a process of nationalization associated with the word "future", during the second industrial revolution. Further research and analysis of sentiments and perspectives related to the future can provide a deeper understanding of how societies envision and shape their future trajectories.

CHAPTER 7

REFERENCES

1. Reinhart Koselleck's, (2002). *The Practice of Conceptual History* , Stanford University, Chapter 6. Time and History: <http://files.eshkolot.ru/koselleck.pdf>
2. an Eijnatten, J., Huijnen, P. (2021). *Something Happened to the Future: Reconstructing Temporalities in Dutch Parliamentary Debate, 1814–2018*. Contributions to the History of Concepts, 16(2), 78-97. <https://www.berghahnjournals.com/view/journals/contributions/16/2/choc160204.xml>
3. "Le Figaro." Wikipedia, The Free Encyclopedia. Wikimedia Foundation, (2023)https://en.wikipedia.org/wiki/Le_Figaro
4. "The New York Herald." Wikipedia, The Free Encyclopedia. Wikimedia Foundation, (2023)https://en.wikipedia.org/wiki/New_York_Herald
5. "El Imparcial." Wikipedia, The Free Encyclopedia. Wikimedia Foundation(2023) https://en.wikipedia.org/wiki/El_Imparcial
6. Hussem Ben Belgacem, (2022). *hbenbel/French-Dictionary*. GitHub repository.<https://github.com/hbenbel/French-Dictionary/blob/master/dictionary/dictionary.csv>
7. OPTED Project Gutenberg. (2021). *The Online Plain Text English Dictionary (OPTED): Dictionary in CSV Format Based on the Webster's Dictionary 1913 Edition*. Retrieved from <https://www.kaggle.com/datasets/dfydata/the-online-plain-text-english-dictionary-opt-resource=download>
8. Jeremy Druin, (2018) . *Webpwnized/byepass/dictionaries/all-spanish-words.txt* . GitHub repository.<https://github.com/webpwnized/byepass/blob/master/dictionaries/all-spanish-words.txt>
9. Elisa Michelet, (2023). *industrial-west/notebooks/a_PAM.ipynb*. GitHub repository. <https://github.com/arobaselisa/industrial-west>
10. Le Figaro: Texte des documents de presse du projet Europeana Newspapers (XIXe-XXe siècles) <https://api.bnf.fr/fr/texte-des-documents-de-presse-du-projet-europeana-newspapers>
11. El imparcial: Hemeroteca digital de la Biblioteca Nacional de España: texto completo - Conjunto de datos <https://datos.gob.es/es/catalogo/ea0019768-hemeroteca-digital-de-la-biblioteca-nacional-de-espana>

12. New York Herald: Chronicling America « Library of Congress <https://chroniclingamerica.loc.gov>
13. Elena Fernández Fernández, Mirco Schoenfeld and Juergen Pfeffer (2020). *Measuring the Acceleration of the Social Construction of Time*. <https://ceur-ws.org/Vol-2723/short17.pdf>