



Grasping the Anti-Modern Discourse on Europe in the Swiss Digitised Press, or can Text Mining Generate a Research Corpus from an Article Collection?

ESTELLE BUNOUT 

SPECIAL COLLECTION:
COMPUTATIONAL
HUMANITIES
RESEARCH DATA

RESEARCH PAPER

ubiquity press

ABSTRACT

In this paper, we discuss how different types of automatic annotation of digitised newspaper articles can be integrated into the iterative questioning of the source material and the creation of research corpora out of a collection of unstructured texts (kept in a structured collection). We annotate a sizeable collection of Swiss press articles (183,270), extracted via the *impresso* interface¹ using topic modelling (MALLET)² as well as a naïve Bayes classifier (script by Milan van Lange).

The methodological discussion we propose is to explore how text mining can help identify historical discourses that are difficult to query with keywords because of their inherent ambiguity and how to grasp them in a large corpus. We argue that the automated annotations can provide a body of corroborating evidence of the searched discourse, to be used as an intermediary and heuristic analysis step.

CORRESPONDING AUTHOR:

Estelle Bunout

Department for Media
history, Leibniz Centre for
Contemporary History
Potsdam, Potsdam, Germany

Estelle.bunout@zzf-potsdam.de

KEYWORDS:

digitised newspapers; history
of ideas; text mining

TO CITE THIS ARTICLE:

Bunout, E. (2021). Grasping the Anti-Modern Discourse on Europe in the Swiss Digitised Press, or can Text Mining Generate a Research Corpus from an Article Collection? *Journal of Open Humanities Data*, 7: 21, pp. 1–11. DOI: <https://doi.org/10.5334/johd.37>

In the introduction to his reference book on the European idea, Gollwitzer (1951) shared his feeling of unfinished business. Indeed, working on the European idea, we are confronted with a rich diversity of thought, practices and projects formulated over the past centuries, written by philosophers and politicians who serve as the pantheon of the European idea, as collected by Böttcher (2014). Europe is increasingly considered in the historiography as being a utopian project, and, as Stråth (2016) explains, that means that it is a mental projection based on rational, artificial conceptions, aiming at materialising an ideal (Baczko, 1984). The Europe-Utopia is considered to have emerged with the Enlightenment. Europe-Utopia is treated in that context as one facet of the modern utopia that emerged in the 18th century and has since then been continuously debated as part of it. In other words as Stråth (2016, p. 44) stated: “modernity has meant a continuous and contentious search for the definition of Europe.” Modernity itself has to be understood as a process of constant efforts to translate utopian conception into a social order, and, as Strath & Wagner (2017, p. 16) explain, the modernity is the “work of interpreting the imaginary signification of modernity with a view to sedimenting them in institutions, and to re-interpret modernity in the light of experiences made with modern institutions and of the critique of their falling short of expectations.” The opposition to this Utopia, as part of a more general opposition to modernity, has also risen in the historiographic discussion, with reflections produced on the chimaera Europe (Burgdorf, 1999), anti-liberal Europe (Gosewinkel, 2015) and in a broader perspective, the particular connotation associated with the “West”, in the German intellectual tradition (Pöpping, 2001). Outside of the historiography of European ideas and movements, the study of the antagonist views on the modernisation of European societies and culture has flourished, especially in the field of literature (Compagnon, 2016; Jessup, 2001) and for the intellectual history of Central and Eastern Europe (Antohi & Trencsényi, 2018).

Anti-modernism, as its name indicates opposes modernism. This oppositional dimension is central, and, as Compagnon (2016) underlines, it is essential to anti-modernism that its actors position themselves, through their publications and attitude, against a phenomenon that is diverse and complex but more essentially at the core of their societies. Anti-modernism is an ambiguous posture: it is not reaction, not longing to return to previous times while still being part of modernity, but advocating for another definition of what modernity should be (Antohi and Trencsényi 2018). We could summarise the anti-modern worldview as the emphasis on aesthetic perception of reality. It is both very fuzzy to predefine and very recognisable when reading individual texts. Antoine Compagnon proposes six elements to capture anti-modernism:

- Figure of style: Vituperation
- Moral figure: Pessimism
- Aesthetic figure: Sublime
- Historical/political figure: Counter-revolution
- Philosophical figure: Anti-Enlightenment
- Religious/Theological Figure: Original Sin

Those are cumulative elements, and it is in their accumulation that the anti-modern style becomes apparent and communicable. Some aspects refer to style (vituperation, pessimism) while others could be captured potentially by names or specific keywords. It remains however difficult to predict what particular reference will be used in a particular historical collection. In that context, how could this typology be operationalized for a collection of unstructured texts, to extract a corpus of articles containing anti-modern discourse, and how could it be translated into measurable quantities (Nguyen et al., 2020) which can be used to feed the historical analysis?

We describe in this paper how the data-driven and supervised annotations can by cumulation help build clusters of relevant articles, containing the searched discourse but also contribute to

1 impresso. (2019). *Media Monitoring of the Past*. Supported by the Swiss National Science Foundation under grant CR- SII5_173719. <https://impresso-project.ch>.

2 McCallum, Andrew Kachites. (2002). *MALLET: A Machine Learning for Language Toolkit*. <http://mallet.cs.umass.edu>.

the reflection on the issue while modelling and analysing the output of these annotations. The annotation of each article produced by common text mining tools results in a cumulation of clues that can support the interpretation of historical source material.

We chose to search for the expression of anti-modern discourses in the Swiss press, as Switzerland was the home of European movements (Jilek, 1990) and international institutions such as the League of Nations, and it was surrounded by countries where fascism took root (Italy and Germany), which was echoed in the local political and intellectual life (Butikofer, 1996; Mattioli, 1997). Newspapers also contain a broader range of discourses than the traditional sources for the history of the European idea, as we can find, in addition to opinion articles, reports on major and minor daily political and cultural events (Greiner, 2014). We will now present the steps from data acquisition and the text mining tools applied to the corpus before presenting and discussing the results.

(2) OPERATIONALISING DISCOURSE ANALYSIS WITH TEXT MINING

Another evident characteristic of the press as historical source material is its size. We will now present how we use topic modelling to manage the diversity of thematic content and a naïve Bayes classifier to expand a seed collection of anti-modern articles. As shown below in the sketch, we first extracted a starting corpus of articles containing the word “Europe” or “European”, and second, we applied topic modelling to each title so that we could have a first “grid” applied to this corpus (*Figure 1*). Along the lines of this grid made of topics, we could read and select relevant articles that were then used to train a classifier. The superposition of the classifier and topic modelling output results in clusters which form the basis for the discourse analysis.

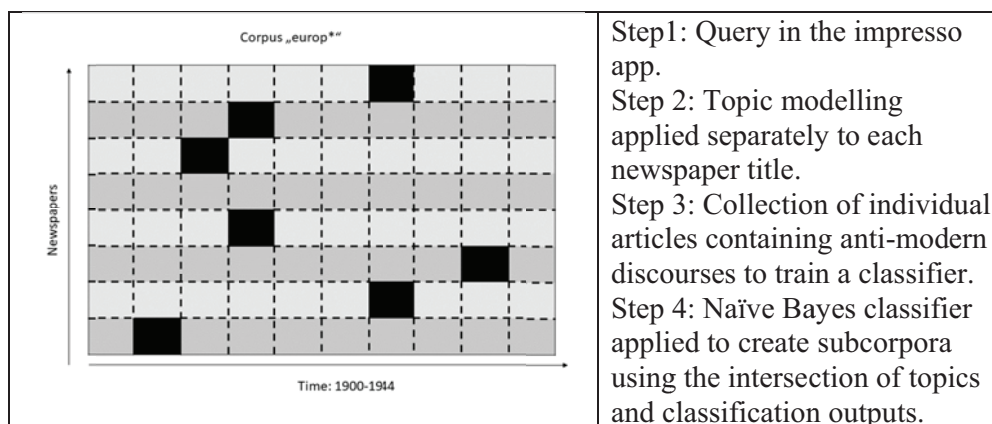


Figure 1 Corpus annotation steps.

(2.1) TRAPPED IN THE STRING QUERY? COLLECTING A CORPUS OF ARTICLES CONTAINING “EUROPE” FROM THE IMPRESSO APP

The recent evolution towards the digitisation of historical sources, especially the press, makes it possible to deal with the discussion of Europe in a broader sense rather than focusing on dedicated organisations or in dedicated publications. Using the digitised newspapers opens the possibility to study the broader social uses of Europe beyond the formalised ideas of Europe, developed by organised entities (Abel, 2013; Ackerley, 2019; Koenen, 2018; Wijfjes, 2017).

Contrary to many other cases, we benefited from easy access to the digitised newspapers via the *impresso* interface, which enables us to explore digitised newspapers from Switzerland and Luxembourg, mainly in German and French. This app was developed by an interdisciplinary team of computational linguists, digital humanists, designers, and historians, and it includes natural language processing (NLP) tools for historical print media and an interface for the active exploration and critical analysis of newspaper corpora.

We focused on six French language Swiss newspapers: *Journal de Genève* (JDG), *Gazette de Lausanne* (GDL), *Express* (EXP), *l’Impartial* (IMP), *Le Confédéré* (LCE) and *l’Essor* (LES). These titles cover a relative diversity of newspaper types: the *Journal de Genève* (JDG) and *Gazette de Lausanne* are considered liberal-conservative, and, although not formally affiliated to any

political party, are perceived to be opinion newspapers, whereas *Express* and *Impartial* were originally created to disseminate commercial news and advertisements. *Le Confédéré* had close ties to the liberal party and *l'Essor*, to the Christian socialism (Clavien et al., 2015).

Using the keyword “Europe” can of course not be equated to the study of the idea of Europe (Grunewald & Bock, 1997), let alone, the antimodern discourse on Europe. However, as both dimensions (antimodernism and europeanism) are described with many polysemic words and are difficult to query as such. Also, in order to increase the discovery, we decided to collect a wide range of press articles and narrow down with the help of automatic annotation.

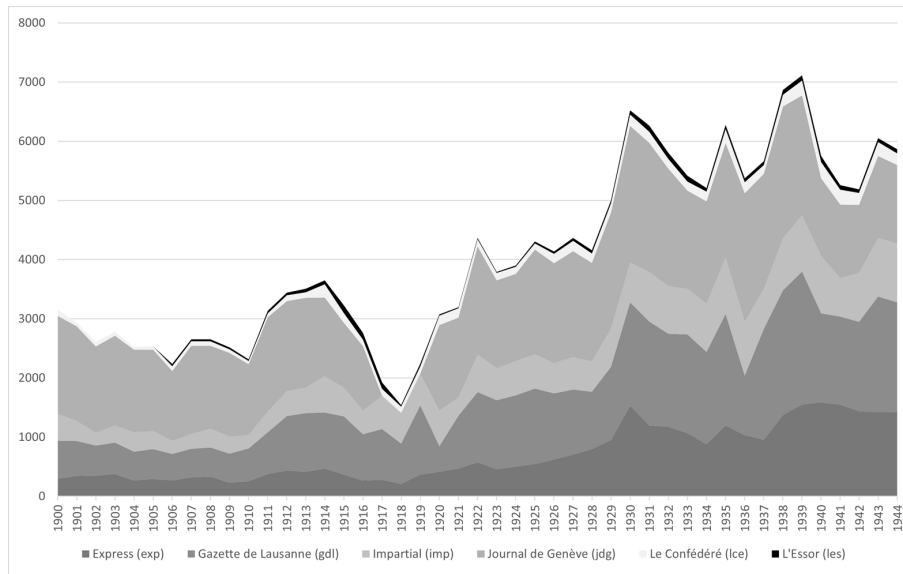


Figure 2 cumulated frequencies of articles containing “europ*” in *l'Express*, *Gazette de Lausanne*, *Journal de Genève*, *l'Impartial*, *le Confédéré* and *l'Essor*, 1900–1944, extracted from the *impresso* app.

The other outcome of the query for the words “Europe”, “European”, brought together with the query for “europ*” (including the variations “Europe”, “européen”, “européenne” etc. in French) for these 6 newspapers is 183 270 articles, for the period ranging from 1900–1944, with some significant difference in size, across the different newspapers titles (Figure 2). Looking at the simple frequency measure, one can already notice the increase over time of the results, which may not necessarily translate into a more intensive use of the word Europe in the press but may be connected to a change in the quantity of pages printed, to a changing quality in the OCR recognition and the subsequent higher number of hits. The later may especially explain the variations around 1917–1920, when the paper quality decreased in general all across Europe, and therefore also the quality of the digitised newspapers (Gooding, 2016). Beyond these formal biases, we observe that the frequency meets our expectations, with a peak in the early 1930s, which may reflect the debate that took place around Aristide Briand’s proposition (Fleury et al., 2001) or the 1938 peak, which may echo the political tensions in Europe, resulting in many mentions of the word “Europe”. These observations are a first and superficial check but cannot lead to conclusions. The *impresso* app lets us extract the articles’ bibliographic metadata, content, and detected named entities. We lose, however, the visual dimension, although it still available via the app. For instance, the caricature shown below will only appear as a short text in our corpus (Illustration 1).



Illustration 1 Screenshot of the *impresso* app, displaying a caricature published in *L'Impartial*, 13 April 1944.

This first step involving the collection of the textual source material leaves us with a large collection of individual articles but with little information on their content. We will now discuss how to add a new layer of automated annotation to navigate this collection of digital primary source material, and doing so, creating another kind of context for this historical source material.

(2.2) PAST THE POLYSEMY: FINDING “EUROPE” IN THE DIGITISED PRESS

Looking closer at each of the European projects reveals many gradations of richness: those of contradictions, unexpected ideological affiliations, and ambiguous formulations serving a particular political goal, which have since then been disconnected from the texts that has entered the pantheon (Schirmann, 2007, p. 108). Using a query with the word “Europe” returns many different types of uses of that word, and very few of them will be relevant to our research question. To sort out this richness, we resorted to topic modelling, which “models a comprehensive representation of the corpus by inferring latent content variables, called *topics*” (Maier et al., 2018, p. 94). Topic modelling produces an estimation of word co-occurrences in a collection of documents and then uses the co-occurrences to cluster the documents, and it has been often discussed in the context of exploring large digital textual collections covering a period of a few years (Galen & Nicholson, 2018; Jacobi et al., 2016; Mimno, 2012). For each cluster, a list of words describes the most frequent co-occurring words. We have used a popular tool, MALLET, which offers probabilistic topic modelling. This means that the clusters created by the topic models are not stable, and so, if run several times, the soft clusters which are created will not overlap. However, after a few trials, it appears that some trends in the topics (but not the exact article distribution) are repeated. One strength of this tool is that it is less sensitive to faulty text, which is a common feature of digitised historical materials.

As formulated by Baumer et al (2017, p. 1399), “[d]ocuments and words are then mapped to a low-dimensional latent space wherein geometric proximity matches human notions of semantic similarity”, capturing both the allure of this tool and the repeated calls for caution surrounding their use in digital humanities. We use it here to get a sense of the topic diversity and select articles that are relevant to our research. After a few times experimenting with the number of topics or the chronological segmentation of the corpus, we trained the topic modelling for each newspaper title separately and set the number of topics to be detected to 100 in order to have a fine-grained structuration for each article collection. We then selected the articles that contained at least 30% of a given topic. The computed topics link very divergent numbers of articles: some link thousands of articles and some just a few. We chose this threshold for a first exploration, to have enough articles for each topic in order to estimate the intra-topic semantic coherence but not too many, so that manual inspection remains doable. For the manual selection of articles tagged with the most relevant topics, the threshold was lowered to 20%.

We proceeded to label each topic. By doing so, one can also manually create connections between the topics computed on each title. This exercise indicated that the dominant topics, in quantity, are topics that consist mainly of stop words that were not removed in the pre-processing, followed by topics labelled as financial news, sports, and cultural programs. This indicates a significant number of articles containing mere mentions of the word “Europe”, without discourses on Europe. The topic modelling was intended to collect articles that correspond to expected themes. It can also be used to explore aspects that were not anticipated. In each title, we could identify topics that may raise interesting questions on the use of the term “Europe”. For instance, several topics seem to gather articles that contain reports on ceremonial speeches, or political, cultural or sporting events. This is a way to cluster articles to search for a particular use of the word “Europe” in public speeches, providing emphasis and references to relevant elements of collective imagination.

Some topics seem to indicate more than an association of themes, and even give some indication of the tone of the articles (Hulden, 2019) and contain some traces of anti-modern references (names or keywords), made possible with the setting of bigrams for the training of the topic modelling (shown below with the underscore linking two words). Following Jacobs & Tschötschel (2019), we use topic modelling to also feed directly our discourse analysis on Europe. We will now present some examples of topics produced, in their original French.

The topics read as follows: abbreviation of the newspaper, T + the topic number, and the top 20 words or bigrams identified for the topic. What is interesting about these topics is the association of terms: war with peace associated with civilisation, nation and moral, offers a notable connotation, in contrast to an association with international institutions or front and army.

- GDL_T26_l'homme civilisation libéralisme l'individu doctrine humaine problème moderne valeurs reynold philosophie morale pensée dieu christianisme notion conception nietzsche mystique spirituelle
- IMP_T85_guerre peuple paix peuples nos monde nations liberté pays nation avons droit patrie lutte l'humanité vie hommes sommes civilisation l'histoire
- LES_T63_maître christianisme sois disciple fidèle_jusqu'à jusqu'à_mort shaw sois_fidèle non-résistance mort_disciple maître_sois opportuniste morales_spirituelles l'absolu transition période_transition l'eglise_demain notions_morales moude vos_ennemis

The labelling and the manual inspection of the clusters created by the topic modelling helped us to have a first global understanding of the large article collection but also detect some niches of potential anti-modern discourses, thanks to the chosen level of granularity. We could start to collect individual articles to feed a naïve Bayes classifier that would help us find additional similar texts.

(2.3) TRANSLATING THE ANTI-MODERN FIGURES INTO WORDLISTS WITH A NAÏVE BAYES CLASSIFIER

The idea of using a naïve Bayes classifier as a way to transpose onto the digital material, the traditional historians' practice of "finding and collecting similar texts", came thanks to a fruitful exchange with Milan van Lange, during his fellowship at the C²DH (van Lange, 2019). Following extensive discussion on the research question, Milan van Lange proposed to use such a classifier and prepared a script in R for this. This tool is often best explained through its popular use in training a spam filter but has also been implemented in the detection of literary genres (Long & So, 2016). To train the naïve Bayes classifier, we collected about one hundred articles labelled as "anti-modern" and one hundred that were either labelled neutral or "modernist". The manual collection was mostly validated by another historian, who was given the criteria of Antoine Compagnon, however this validation was conducted only on a small portion of articles, due to the limited availability of other domain experts.

The naïve Bayes classifier produces two main outcomes: a list of words that have been identified to discriminate one category from another, and an estimation of the probability of each subsequently submitted document to belong to one or the other category. This wordlist is a useful tool to reflect on the selection criteria we applied. It makes it possible to become aware of distinctive words, as determined by our own collection, not an indisputable anti-modern vocabulary list. For instance, as shown in the table below, the output of probabilities that some words are discriminatory for a given category, for the training on anti-modernism, as shown in the table below.

For example, the word "décadence" appears in the training set, i.e., the list of articles classified manually as anti-modern vs. non anti-modern, and occurs most frequently in the articles that have been tagged as anti-modern. The consequence is that the unknown, i.e., the not yet tagged articles that are submitted to the naïve Bayes classifier, containing the word "décadence" will have a higher probability to be tagged as "anti-modern" by the classifier.

Following the settings defined by Milan van Lange in his script, numbers were not excluded from the text in the pre-processing steps. Keeping numbers in the text can prove to be useful if a date is a marker of a particular discourse. Here for anti-modernism, the French Revolution being a target, we hope to catch some reference to it in our NBC net. Indeed, the bigram "révolution_1789" has the probability of 66% to indicate that a text containing it is of anti-modern discourse on Europe and of 52% to indicate an anti-modern discourse with a simple mention of Europe, based on the training set we prepared.

We can see also in the [Table 1](#) that words like "marxisme" or "phénomène" are also used to tag texts as "anti-modern", which reflects our selection but does not inform about the context in which they are used in the untagged articles. The fact that "Marxism" describes "anti-modern"

NONAM	AM	WORD
4%	96%	reynold
4%	96%	décadence
4%	96%	danielrops
5%	95%	marxisme
5%	95%	phénomène

Table 1 Top 5 words describing the category “antimodern”, produced by the naïve Bayes classifier.

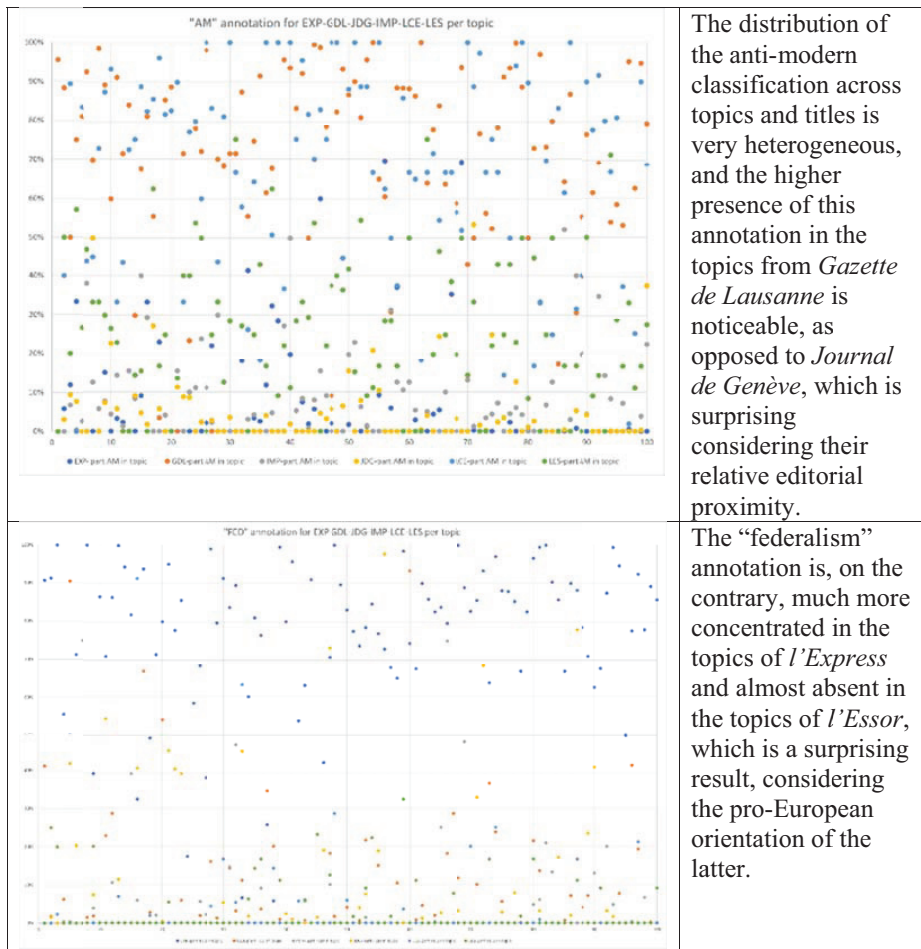
texts reflects the importance of anti-communist texts in the manual selection, however, it is the whole list of words that is considered to calculate the probability of belonging to the “anti-modern” category. We opted for another cumulation of annotations, from this same tool, to create a more nuanced picture. This accumulation of signs of similarity (or their absence) is a way to compensate for the inherent fuzziness when using this tool for this purpose. The categories were paired with the category “neutral” for the article selection. In this context, “neutral” means a diversity of articles simply mentioning Europe, without any identifiable discourse on it. The categories were also paired with “opposite” and “similar” categories, to highlight the words that would surface for each category.

A first round of naïve Bayes classifier was used to test the reliability of this tool, which was validated after a manual inspection of samples of the results. A second round added another category: the presence of anti-modernism was separated from the expression of a European anti-modern discourse. The opposing categories to “anti-modern” were also expanded: to contrast with the target discourse, we collected training sets for reports on European diplomatic events, or the simple mention of Europe, without any discourses on it. Several categories were tested during this round, and some, like “studies on Europe” or “paneurope”, were either not yielding relevant results or the training set was not large enough to be used reasonably. A third round was conducted with a set larger than 100 articles, which resulted in a more diverse but less relevant results. A fourth, and so far, final, round focused on contrasting the selected categories to learn more about how they differed in terms of words and distribution. Instead of looking at opposing categories (anti-modern vs. non anti-modern), we contrasted the categories that were also close to each other (anti-modern vs. anti-communist), to see how it would impact the wordlists. The juxtaposition of the words produced by the pairing of the anti-modern conceptual category with “similar” categories such as the fascist conception of Europe (FE) or anti-communism (AC), and with “neutral” or “opposite” categories such as the crisis in Europe (CE) or utopian conception of Europe (UE) help us highlight the terms, persons, and slogans that we have indirectly selected with the collection of articles.

(3) CLUSTERS OF USES OF EUROPE AND ANTI-MODERN DISCOURSES IN THE SWISS PRESS

With these tools, we could produce automatic annotations for each article collected: one being data-driven (topic modelling), and the other supervised (naïve Bayes classifier). These helped us not only to create clusters but also to reflect on the categories we envisaged at the start of the research and to become aware of the unexpected proximity of certain categories (e.g. utopian and anti-modern) in part of the tagged articles and the relative scarcity of other categories (e.g. federalist ideas) in the corpus. The result also highlights the differences and proximity of each newspaper title. We can now measure various levels of intersection, including, for instance, the proportion of articles annotated as “anti-modern” in each topic, like on the graphs below in *Figure 3*. The topics that were previously not inspected because of an indication of seemingly irrelevant content might get highlighted by a high proportion of naïve Bayes classifier annotation.

The article clusters, constituted in this way, help to raise questions regarding the content of the articles. What brings these articles together? How relevant is the categorisation? The initial categories we used for the manual annotation were taken from the historiography on Europe, mainly developed by Greiner (2014) and reflecting the uses expected to be found in the press, and were used to map out the uses of Europe:



The distribution of the anti-modern classification across topics and titles is very heterogeneous, and the higher presence of this annotation in the topics from *Gazette de Lausanne* is noticeable, as opposed to *Journal de Genève*, which is surprising considering their relative editorial proximity.

The “federalism” annotation is, on the contrary, much more concentrated in the topics of *l'Express* and almost absent in the topics of *l'Essor*, which is a surprising result, considering the pro-European orientation of the latter.

Figure 3 Proportion of naive Bayes classifier categories per topic, per title.

- Experienced Europe (l'Europe vécue) as defined by (Girault, 1994): awareness of belonging to a common cultural, historical, or political entity in a broad sense.
- Representation of Europe (Europabild) as defined by (Gollwitzer, 1951): the representation of European history and culture, and specificities of Europe compared to the rest of the world.
- Idea of Europe (Europagedanke) as defined by Gollwitzer (1951): expression of a political project at European scale, and cooperation between states and people.
- Scarecrow Europe: Europe as a source of danger for national society.
- Hope Europe: Europe as a saviour which can address the dangers a society faces, without any specific project formulated.
- Diplomatic Europe: reports on meeting of European politicians.
- Geography: use of the word Europe as a geographic entity.
- Crisis Europe: Europe as being the scale of the political or economic crisis, as defined by Florian Greiner. Used to contrast with the national situation.
- Europe as echo chamber: use of the word to emphasize the importance of a phenomena, as defined by Oschema (2013).
- Decadent Europe: Europe as the source of society's decadence.

After several iterations via topic modelling and the naïve Bayes classifier, we retained the following categories, based on the significance of their presence and the relevance to the discourse analysis:

- Anti-communism
- Anti-modern articles simply mentioning Europe
- Crisis in Europe
- Reports on European diplomatic affairs
- Anti-modern conception of Europe
- Europe as a common space
- Fascist conception of Europe

- Federalist projects for Europe
- Neutral (non anti-modern, no discourse on Europe)
- Studies on Europe
- Utopian ideas for Europe

To sum up, the building of the clusters relied on the interplay between data-driven annotation and manual inspection, as well as revision of studied categories and careful interpretation of the annotation output, as described by Grimmer & Stewart (2013).

(4) CORPUS ANNOTATION AS OPERATIONALISATION OF DISCOURSE ANALYSIS VIA TEXT MINING

This clustering process creates a new framework to read the analysed material: an article might not be read anymore in the context of leafing through the pages of a newspaper or by manually surveying a collection of press clippings, or even in the result list of an interface, but rather in the context of a topic and the output of a classifier. This raises the question of the connection between the qualitative and quantitative dimensions: what are the similarities detected by the text mining and how can we integrate them in our historical analysis? We would argue that the detected textual overlap and co-occurrences are a stimulating preselection and serve as a grid to manage a large collection of historical source material. Alongside the use of word embeddings to operationalise the conceptual history (Marjanen et al., 2020), the text mining tools can offer a constructive support for the study of discourses. Following-up this initial iteration of creating and adapting research categories to the output of a classifier, and given the difficulty to manually validate the whole corpus, it would be interesting to apply the process to other source materials, for instance, in the context of the *impresso* collection, to further Swiss newspapers, in order to also include socialist titles and Luxembourgish newspapers of the time.

ADDITIONAL FILES

The additional files for this article can be found as follows:

- **Clusters of articles.** Details of the topic modeling and naïve Bayes classifier values, normalised for the 6 selected press titles: *l'Express*, *la Gazette de Lausanne*, *l'Impartial*, *Journal de Genève*, *l'Essor*, *le Confédéré*. DOI: [10.5281/zenodo.4891841](https://doi.org/10.5281/zenodo.4891841)
- **Contrasting classifications.** Wordlists produced by a Naive Bayes Classifier applied to digitised newspapers, following a contrastive approach to discourses on Europe: 1) AM: antimodern discourses vs. non-antimodern discourses; 2) EA-AC: antimodern discourses on Europe vs. anticommunist discourses; 3) DE-FED: reports on diplomatic events and federalist discourses on Europe. DOI: [10.5281/zenodo.4891843](https://doi.org/10.5281/zenodo.4891843)
- **European Topics in Swiss digitised newspapers.** Top 20 words describing the topics, produced by topic modelling applied to a selection of articles from six Swiss newspapers, containing the word “europ*”. DOI: [10.5281/zenodo.4895200](https://doi.org/10.5281/zenodo.4895200)
- **Training materials to capture discourses on Europe in digitised newspapers.** List of articles used to training a naïve Bayes classifier, selected from French language Swiss newspapers, extracted from the *impresso* app. DOI: [10.5281/zenodo.5094936](https://doi.org/10.5281/zenodo.5094936)
- **Annotated corpus of Swiss press articles on Europe (1900–1944).** For each article extracted from the *impresso* app, containing the word “europ*” for 10 French language Swiss newspapers, all the values generated by a topic modelling and naïve Bayes classifier, trained to detect specific uses of Europe in this corpus. DOI: <https://doi.org/10.5334/johd.37.s1>

ACKNOWLEDGEMENTS

This research was made possible by the combined efforts of the team from the *impresso* project, to which I would like to express my gratitude. The help of Milan van Lange from the NIOD Institute for War, Holocaust and Genocide Studies (NL) was essential to design this workflow, and I would like to express my gratitude here to him.

The ‘impresso – Media Monitoring of the Past’ project is funded by the Swiss National Science Foundation (SNSF) under grant number [CRSII5_173719] (Sinergia program). The project aims to develop tools to process and explore large-scale collections of historical newspapers, and to study the impact of these new tools on historical research practices. More information at <https://impresso-project.ch>.

COMPETING INTERESTS

The author has no competing interests to declare.

AUTHOR AFFILIATIONS

Estelle Bunout  orcid.org/0000-0003-1009-3426

Department for Media history, Leibniz Centre for Contemporary History Potsdam, Potsdam, Germany

REFERENCES

- Abel, R.** (2013). The Pleasures and Perils of Big Data in Digitized Newspapers. *Film History*, 25(1–2), 1–10. DOI: <https://doi.org/10.2979/filmhistory.25.1-2.1>
- Ackerley, A.** (2019). Historic Newspapers in the Digital Age: ‘Search All About It!’ *Media History*, 25(2), 261–264. DOI: <https://doi.org/10.1080/13688804.2019.1592901>
- Antohi, S., & Trencsényi, B.** (2018). Introduction. Approaching Anti-modernism. In D. Mishkova, M. Turda, & B. Trencsenyi (Eds.), *Anti-modernism: Radical revisions of Collective Identity* (pp. 1–43). Central European University Press. <http://books.openedition.org/ceup/2969>
- Baczko, B.** (1984). *Les imaginaires sociaux: Mémoires et espoirs collectifs*. Payot.
- Baumer, E., Mimno, D., Guha, S., Quan, E., & Gay, G.** (2017). Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence? *Journal of the Association for Information Science and Technology*, 68, 1397–1410. https://epublications.marquette.edu/mscs_fac/560. DOI: <https://doi.org/10.1002/asi.23786>
- Böttcher, W.** (Ed.). (2014). *Klassiker des europäischen Denkens: Friedens- und Europavorstellungen aus 700 Jahren europäischer Kulturgeschichte* (1. Auflage). Nomos. DOI: <https://doi.org/10.5771/9783845241371>
- Burgdorf, W.** (1999). *Chimäre Europa: Antieuropäische Diskurse in Deutschland (1648–1999)*. Verlag Dr. Dieter Winkler.
- Butikofer, R.** (1996). *Le refus de la modernité: La Ligue vaudoise, une extrême droite et la Suisse, 1919–1945*. Payot Lausanne – Nadir.
- Clavien, A., Scherrer, A., Mena, F., & Collenberg, A.** (2015). Presse. In *Dictionnaire historique de la Suisse (DHS)*. <https://hls-dhs-dss.ch/articles/010464/2015-04-10/>
- Compagnon, A.** (2016). *Les Antimodernes: De Joseph de Maistre à Roland Barthes*. Folio.
- Fleury, A., Jilek, L., & Badel, L.** (2001). Le plan Briand d’union fédérale européenne. Perspectives nationales et transnationales, avec documents. *Journal of European integration history*, 7(1), 133–134.
- Galen, Q. V., & Nicholson, B.** (2018). In Search of America. *Digital Journalism*, 6(9), 1165–1185. DOI: <https://doi.org/10.1080/21670811.2018.1512879>
- Girault, R.** (1994). *Identité et conscience européennes au XXe siècle*. Hachette.
- Gollwitzer, H.** (1951). *Europabild und Europagedanke: Beiträge zur deutschen Geistesgeschichte des 18. u. 19. Jahrhunderts*. Beck.
- Gooding, P.** (2016). *Historic Newspapers in the Digital Age: ‘Search All About It!’* Routledge. DOI: <https://doi.org/10.4324/9781315586830>
- Gosewinkel, D.** (2015). *Anti-liberal Europe: A Neglected Story of Europeanization*. Berghahn Books.
- Greiner, F.** (2014). *Wege nach Europa: Deutungen eines imaginierten Kontinents in deutschen, britischen und amerikanischen Printmedien, 1914–1945*. Wallstein Verlag. DOI: <https://doi.org/10.5771/9783835326385>
- Grimmer, J., & Stewart, B. M.** (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis*, 21(3), 267–297. DOI: <https://doi.org/10.1093/pan/mps028>
- Grunewald, M., & Bock, H. M.** (1997). *Le discours européen dans les revues allemandes = Der Europadiskurs in den deutschen Zeitschriften: (1918–1933)*. Bern, Switzerland: Peter Lang CH.
- Hulden, V.** (2019). A Computational Analysis of Topics and Tone in American Labor and Mainstream Newspapers, 1909–191. Presented at *The Labor and Working Class History Association*. https://www.colorado.edu/faculty/hulden-vilja/sites/default/files/attached-files/hulden_computationalanalysis_lawcha.pdf

- Jacobi, C., van Atteveldt, W., & Welbers, K.** (2016). Quantitative analysis of large amounts of journalistic texts using topic modelling. *Digital Journalism*, 4(1), 89–106. DOI: <https://doi.org/10.1080/21670811.2015.1093271>
- Jacobs, T., & Tschötschel, R.** (2019). Topic models meet discourse analysis: A quantitative tool for a qualitative approach. *International Journal of Social Research Methodology*, 22(5), 469–485. DOI: <https://doi.org/10.1080/13645579.2019.1576317>
- JAMAIS CONTENT.** (1944, April 13). L'Impartial, 6. Retrieved via *impresso* app, <https://impresso-project.ch/app/issue/IMP-1944-04-13-a/view?sq=&articleId=i0062&text=1&p=1>
- Jessup, L. L.** (2001). *Antimodernism and Artistic Experience: Policing the Boundaries of Modernity*. University of Toronto Press. <http://ebookcentral.proquest.com/lib/unilu-ebooks/detail.action?docID=4669961>. DOI: <https://doi.org/10.3138/9781442623101>
- Jílek, L.** (1990). *L'esprit européen en Suisse de 1860 à 1940*. Cahiers d'histoire contemporaine.
- Koenen, E.** (2018). Digitale Perspektiven in der Kommunikations- und Mediengeschichte. *Publizistik*, 63(4), 535–556. DOI: <https://doi.org/10.1007/s11616-018-0459-4>
- Long, H., & So, R. J.** (2016). Literary Pattern Recognition: Modernism between Close Reading and Machine Learning. *Critical Inquiry*, 42(2), 235–267. DOI: <https://doi.org/10.1086/684353>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., Häussler, T., Schmid-Petri, H., & Adam, S.** (2018). Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. *Communication Methods and Measures*, 12(2–3), 93–118. DOI: <https://doi.org/10.1080/19312458.2018.1430754>
- Marjanen, J., Kurunmäki, J., Pivovarov, L., & Zosa, E.** (2020). The expansion of isms, 1820–1917: Data-driven analysis of political language in digitized newspaper collections. *Journal of Data Mining and Digital Humanities*. <https://hal.inria.fr/hal-02491304v5>. DOI: <https://doi.org/10.46298/jdmdh.6159>
- Mattioli, A.** (1997). *Gonzague de Reynold: Idéologue d'une Suisse autoritaire*. Editions Universitaires.
- Mimno, D.** (2012). Computational historiography: Data mining in a century of classics journals. *Journal on Computing and Cultural Heritage (JOCCH)*, 5(1), 3. DOI: <https://doi.org/10.1145/2160165.2160168>
- Nguyen, D., Liakata, M., DeDeo, S., Eisenstein, J., Mimno, D., Tromble, R., & Winters, J.** (2020). How We Do Things With Words: Analyzing Text as Social and Cultural Data. *Frontiers in Artificial Intelligence*, 3. DOI: <https://doi.org/10.3389/frai.2020.00062>
- Oschema, K.** (2013). *Bilder von Europa im Mittelalter*. Jan Thorbecke Verlag.
- Pöpping, D.** (2001). *Abendland: Christliche Akademiker und die Utopie der Antimoderne 1900–1945*. Metropolis Verlag.
- Schirmann, S.** (2007). *Penser et construire l'Europe (1919–1992): États et opinions nationales face à la construction européenne*. ÉdSedes.
- Stráth, B.** (2016). *Europe's Utopias of Peace: 1815, 1919, 1951*. Bloomsbury Publishing.
- Strath, B., & Wagner, P.** (2017). *European modernity: A global approach*. Bloomsbury Academic. <http://cadmus.eui.eu/handle/1814/49224>
- van Lange, M.** (2019, April 17). *Impresso Talk #5: Beyond the tantrum: Questioning shared intuitions in historiography with emotion mining*. Impresso Talks Series, Belval. <https://www.c2dh.uni.lu/events/beyond-tantrum-questioning-shared-intuitions-historiography-emotion-mining>
- Wijffes, H.** (2017). Digital Humanities and Media History. A Challenge for Historical Newspaper Research. *Tijdschrift Voor Mediageschiedenis*, 20(1), 4–24. DOI: <https://doi.org/10.18146/2213-7653.2017.277>

TO CITE THIS ARTICLE:

Bunout, E. (2021). Grasping the Anti-Modern Discourse on Europe in the Swiss Digitised Press, or can Text Mining Generate a Research Corpus from an Article Collection? *Journal of Open Humanities Data*, 7: 21, pp. 1–11. DOI: <https://doi.org/10.5334/johd.37>

Published: 17 September 2021

COPYRIGHT:

© 2021 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

Journal of Open Humanities Data is a peer-reviewed open access journal published by Ubiquity Press.