

“Simultaneous and Systematic Abolition”?: Automated Content Analysis and the 1789 *Cahiers de Doléances*

Jeff Jacobs
Columbia University
jbj2122@columbia.edu

Abstract

Upon reading a selection of the grievance books (“*cahiers de doléances*”) proffered to King Louis XVI in 1789 by the various regions and estates of France, Alexis de Tocqueville reflected with horror that “what they were asking for was the simultaneous and systematic abolition of all the laws and customs in force throughout the country” (Tocqueville 1856, 131–132). Given that he couldn’t read all 60,000 books, however, to what extent is Tocqueville’s conclusion simply the result of a selection bias?

Shapiro and Markoff (1998), the culmination of decades of work manually tagging a subset of these grievance books, was the first step towards an answer to this question. After centuries of interpretive studies, Shapiro and Markoff were able to subject hypotheses concerning the contents, geographic distribution, and socioeconomic covariates of the *cahiers* to rigorous statistical testing. Not only did their study overturn accepted knowledge from the outset (*e.g.*, Tocqueville’s assertion that “all of the *cahiers* demand[ed] that *corvees* be definitively abolished” when in fact only 13% did so), it also provided an extraordinarily rich dataset of tags with sufficient generality to allow future researchers to test a range of intuitions regarding the French Revolution.

In this paper, I take quantitative studies of the French Revolution and the *cahiers de doléances* one step further, by developing a *fully*-automated (“unsupervised”) content analysis algorithm based on computational topic modeling and sentiment analysis. By eliminating the need for manual labeling, the paper serves as a case study demonstrating the feasibility of obtaining insights in mere minutes which previously required decades of work by a team of research assistants. I show that it is possible to obtain substantive findings regarding the relative concerns of the Nobility, Clergy, and Third Estate using this algorithm, by illustrating the striking degree to which the results of the model comport with both Tocqueville’s and Shapiro and Markoff’s comparative analyses of the *cahiers*. I then use peasant revolt data from Markoff (1996) to explore the relationship between emphasis on the automatically-discovered “seigneurial rights” grievance topic and variation in “anti-seigneurial events” among the peasantry. I conclude with a proposal for a “supervised” learning approach which allows researchers full control over the learned grievance categories via a small set of manually-produced “seed” tags.

1 Introduction

With the summoning of the Estates-General in 1789 by King Louis XVI, requesting “a concourse of our faithful subjects, to assist us in surmounting all the difficulties we find relative to the state of our finances”, the drafting of the *cahiers de doléances* (“grievance books”) began, inaugurating a self-reflective prelude to the revolution that would change the course of history perhaps more than any event before or since. In preparation for the May 1789 meeting, the Clergy, Nobility, and Third Estate of each *bailliage* of France was asked to assemble and to compile a list of their grievances regarding the social, political, and economic practices of the *ancien régime*. Because of this remarkable process, the French Revolution of 1789 stands out as “the only major revolution at the beginning of which so much of the nation gathered in public assemblies and recorded its grievances, aspirations, and demands for change” (Markoff 1996, 21).

Jean Jules Jusserand, French historian and ambassador to the US, once besought a young student of the French Revolution: “Read the *cahiers*. It’s all in the *cahiers*” (Andrews 1935, 212). Upon reading a selection of the *cahiers*, Alexis de Tocqueville was stunned:

When I came to the end of this vast collection of texts and put all these specific requests together, I was horrified to realize that what they were asking for was the simultaneous and systematic abolition of all the laws and customs in force throughout the country. I saw at once that this would lead to one of the most extensive and dangerous revolutions that had ever occurred anywhere in the world. (de Tocqueville 1856, 131–132)

Indeed, it is difficult to avoid reading the *cahiers* dispassionately, knowing that they represent the hopes and fears of a population standing on the precipice of a revolution which would, in Mallet du Pan’s unforgettable phrase, “devour its children”. Tocqueville described it as “a phenomenon so new and so different from anything that had ever happened before, yet so monstrous and incomprehensible, that the human mind could not grasp it” (de Tocqueville 1856, 13). With an event so laden with meaning for so many, and so central to the identity of France and “the West” to this day¹ interpreters of the Revolution like Tocqueville are bound to bring their own biases to the table when studying its annals.

It is with this inevitable affective bias in mind that I develop and conduct the automated content analysis herein. Although human biases are embedded in each step of any given computer algorithm, I contend that the logic behind the trajectory of internal states in an algorithm is significantly more transparent and amenable to error analysis than the internal states of the human mind. Hence, as is emphasized throughout Markoff and Shapiro (1998), automation of a subtask in a content analysis allows the parameters of that subtask to be adjusted rapidly and continuously (in contrast to manual content analyses, where small changes can entail months of recoding), thus allowing the researcher to *encode*, *observe*, and *explore* the effects of a particular bias and “fine-tune” these parameters with an aim towards counterbalancing their own biases. To this end, I have made all code, data, and other replication materials available on GitHub² to enable future researchers to explore different parameter settings and encodings of bias from those constituting this work.

In the next section I review prior work on both the *cahiers* and computational-linguistically-informed automated content analysis, before moving into a description of the data and pre-processing procedures in Section 3.1. Section 3 describes and presents findings from my

¹As Timothy Tackett puts it in his introduction to Lefebvre’s *Coming of the French Revolution*, “an understanding of the Revolution and its origins was linked to the very concept of national identity” in France, compelling French statesmen to return to the Revolution time and time again for lessons bearing on contemporary issues (Tackett 2004, *viii*).

²<https://github.com/jjacobs3>

first study of the contents of *cahiers*, an unsupervised topic model revealing differences between the estates with respect to automatically-discovered topics latent within the *cahiers*. In Section 4 I describe a second “supervised” study of the *cahiers* in two stages. First, after explaining the data processing and manual tagging required as inputs, I describe how to measure how “well-calibrated” the results of the model are to the results of a manual content analysis (described here with in Shapiro and Markoff 1998 as the example). I then explain how to generate a plot illustrating how these results change as the number of training samples decreases, thus providing what is essentially a power analysis, indicating how few training samples a researcher would need to obtain to get within a desired margin of error α of the “true” results obtained via manual content analysis. Section 5 concludes, outlining a series of datasets from manual content analyses which would allow for additional calibration as well as text corpora which are particularly suitable for analysis via the methods developed herein.

2 Prior Work

2.1 Qualitative Studies of the *Cahiers*

Almost from their advent, the *cahiers* were subjected to a radical variety of interpretations. Perhaps the most well-known of these is the interpretation given by Alexis de Tocqueville in his 1856 *The Old Regime and the Revolution*. Tucked away in a footnote to Book II entitled “Analysis of the Grievance Books of the Nobility in 1789”, Tocqueville paints a picture of the *cahiers* of the Nobility as illustrating a radical spirit desirous of sweeping changes across all aspects of social, political, and economic life. “In short,” he intones, “reading the grievance books of the nobles leads to the conclusion that they might have made the Revolution themselves had they only been commoners, for that was the only ingredient they lacked” (de Tocqueville 1856, 244). Recognizing that the validity of these conclusions rest

upon the validity of the *cahiers* themselves as representative of the true feelings of their drafters, Tocqueville insists that they were “drafted in complete freedom, in very public circumstances, by each of the orders. They were discussed at length among the interested parties and earnestly reflected upon by their drafters, for when government appealed to the nation in those days, it did not take it upon itself to supply both questions and answers” (de Tocqueville 1856, 234).

Francois Furet, however, expresses a common skepticism of the *cahiers* when he characterizes them as “difficult to interpret and probably misleading: under the umbrella of the people as a whole, it was really the lawyers who were expressing themselves; they had most often presided over the assemblies and drawn up the grievances” (Furet 1988, 58). And indeed, most scholars agree that “model” *cahiers* (containing pre-selected grievances) were often adopted without modification by local assemblies. Georges Lefebvre, however, points out that a not-insignificant number of the *cahiers* (at both the general and the parish level) eschew the models altogether, or modify and combine them to suit their purposes: “In the bailiwick of Nancy, for example, the use of eleven different models has been proved: 33 parishes adopted one model or another, 12 combined two or more, 32 either saw no models or declined to use them. In the bailiwick of Arques only a quarter of the lists were influenced by models” (Lefebvre 1947, 69). Thus, for the sake of this work, I share Tocqueville’s fascination with the *cahiers* as a unique documentary record from “the only [revolution] at the beginning of which the different classes of society were enabled to present an authentic account of the ideas they had conceived, and express the feelings which animated them, before the revolution had distorted or modified those ideas and feelings” (Tocqueville 1856, 262).

2.2 Quantitative Studies of the *Cahiers*

In the 20th century, two landmark studies of the *cahiers* shifted the focus of this scholarship away from interpretive studies of haphazardly-selected subsets of the *cahiers* and towards statistically-principled large- N analysis. With the archival and methodological developments made by these studies, historians of the French Revolution could for the first time test hypotheses and make well-founded inferences about the contents of the *cahiers*, in Popperian fashion.

Beatrice Hyslop, in a series of three works between 1933 and 1936 (Hyslop 1933, Hyslop 1934, Hyslop 1936), set the standard upon which all subsequent quantitative studies of the *cahiers* have been built. As part of her doctoral work at Columbia University, she was commissioned by the French government’s Commission d’Histoire Économique et Sociale de la Révolution Française to collect and catalogue the contents and details (publication dates, aberrations in the drafting, mistakes and omissions in the existing volumes) of the *cahiers* in one place, a task which led to her 1933 PhD dissertation, titled “Répertoire Critique des Cahiers de Doléances pour les Etats-Généreaux de 1789”.

With this rich set of cahiers and associated information, Hyslop performed a content analysis of nationalistic sentiment in the *cahiers* which for the first time was able to draw conclusions regarding the *entirety* of French society – all three estates and (nearly) every bailliage in every region of the country – utilizing both the general *cahiers* and a wide swath of the parish-level *cahiers* which reflected more directly the opinions of the rural peasants. This work was published in 1934 as *French Nationalism in 1789 According to the General Cahiers*, and remains a cornerstone of French Revolution scholarship.

Most important to the present study, however, is Hyslop’s 1936 *Guide to the General Cahiers of 1789*. This work, essentially synthesizing the lessons learned from her doctoral work and from the issues she encountered while performing the content analysis of French nationalism, acknowledged the fact that the scope of subsequent work would be limited if

researchers could only study the *cahiers* via her particular coding scheme (a fairly coarse set of codes indicating whether a particular grievance voiced nationalistic sentiment). Hence, in this work she translated (and updated) the findings from her 1933 cataloguing of the *cahiers* into English, provided a set of data tables listing sources, dates, corrections, and cross-references for each bailliage³, and included an appendix with hundreds of pages of *cahiers* which had previously been spread across dozens of out-of-print or hard-to-find sources. With this volume, researchers could develop their own coding schemes relative to a shared “base” dataset of *cahiers*, adapted to whatever hypotheses and aspects of 1789 France they were interested in studying.

Gilbert Shapiro and John Markoff’s 1998 *Revolutionary Demands*, the culmination of decades of work manually tagging and analyzing the *cahiers*, represents the state-of-the-art in quantitative (though non-automated) content analysis. Taking seriously Hyslop’s point that a coding scheme developed with a particular hypothesis in mind precludes other researchers from studying different hypotheses without re-doing the entire tagging process, Shapiro and Markoff explicitly aimed to make their coding scheme as general as possible. Central to this aim was their insistence on separating the researchers from the coders as much as possible. The coders in their study were purposefully kept entirely in the dark as to the hypotheses of the study, thus ensuring that the particular hypotheses and analytical categories of interest to the researchers would *not* impact the coding, allowing the codes to be used as-is by other researchers studying vastly different hypotheses.

One way Shapiro and Markoff ensured maximum reusability was by utilizing a *hierarchical* coding scheme, with very general characteristics of the grievances being recorded at the top of the hierarchy (See Columns 1 and 2 of Table 5 for a list of these top-level codes), and specific characteristics (*e.g.*, the mention of particular taxes or judicial institutions) at the lower levels. In addition, their scheme separates the *subject* of the grievances from the

³I have digitized these tables and included them in Appendix E.

particular *actions* demanded regarding these subjects, by having two separate sets of “Subject Codes” and “Action Codes”, with the latter further subdivided into “General Action Codes” applicable across all subjects (*e.g.*, abolition, reform, more/less bureaucratic oversight) and “Special Action Codes” which apply only to particular subjects (*e.g.*, “That the Third Estate be permitted access to posts and careers in the church”, which applies only to subjects within the “Religion” category. For a list of the top 10 Special Action Codes, see Column 5 of Table 4).

As yet another step in the direction of reproducible and transparent scientific study of the French Revolution, Shapiro and Markoff created a computer program, the French Revolution Analysis System, allowing researchers to explore and perform tests on the *cahiers* via boolean queries (*e.g.*, “(Grievance concerns the *corvées*) OR (Grievance concerns the *taille*)”). Although the program is no longer compatible with modern operating systems (tested on Windows 10, Mac OSX, and Ubuntu Linux 16.02), I was able to run it on a Virtual Machine running Windows 95 and export the list of tags *en masse* for each category, data which I have made available on the project’s GitHub repository. It is with the data made available via Shapiro and Markoff’s tables in combination with the tags exported from this program that I was able to utilize Natural Language Processing techniques, discussed in the next section.

2.3 Machine Learning and Natural Language Processing

With the advent of computers and the ability to store and process massive corpora automatically, the computer science subfield of “Natural Language Processing” began to develop algorithms for efficiently parsing, clustering, and performing syntactic and semantic analyses of natural text. Topic Modeling (Blei et al. 2003) is an unsupervised Natural Language Processing technique which aims to discover distinct “topics” within the corpus and classify each portion of the text (in the case of the present work, individual grievances) by how well



Figure 1: Results of a Latent Dirichlet Allocation (LDA) topic model run on the New York Times corpus with the number-of-topics parameter $K = 10$. Figure from Blei (2012).

it seems to correspond to these topics. For example, a topic model run on the New York Times corpus (Figure 1) effectively “recovers” the sections of the newspaper (*e.g.*, Finance, Sports, Food) based solely on the textual contents of individual articles, without any prior knowledge of these sections or “steering” by the researcher.

Though the constituent technologies (topic models, linguistic parsing algorithms, supervised machine learning algorithms) have been available since the late 1980s, the utilization of natural language processing methods for social-scientific content analyses did not begin until 2009, with Stewart and Zhukov’s automated content analysis of speeches by Russian political elites between 1998 and 2008 (Stewart and Zhukov 2009) and Gentzkow and Shapiro’s analysis of news articles published in 2005 across 433 newspapers (Gentzkow and Shapiro 2010). Jensen et al. (2012) incorporate a time-series element, analyzing changes in linguistic polarization in the Congressional Record between 1873 and 2009.

Perhaps most relevant for our present study is a forthcoming study (Blaydes et al. 2018) which uses Topic Modeling to analyze the different focuses that Christian and Muslim writers

gave to different topics when writing “mirrors for princes”, essentially guidebooks for rulers with advice on how to rule effectively. In this study, 46 texts (21 from the “Islamic tradition” and 25 from Christian Europe) in English translation are analyzed using a Bayesian Hierarchical Topic Modeling algorithm (Grimmer 2010), and results are compared between the Islamic and Christian texts.

After examining the most-indicative words within each topic (the hierarchical structure allows them to estimate 4 broad topics, each of which contains 15 more specific sub-topics), they label the topics as “The Art of Rulership” (Topic 1), “Personal Virtues, Habits, and Relationships” (Topic 2), “Religion and Faith” (Topic 3), and “Political Geography and the Natural World” (Topic 4). With these labels, they are able to analyze the relative proportions of the texts which focus on these different areas, and compare these proportions between the Christian and Islamic texts. Using this approach, they find for example that while the two traditions place a similar amount of emphasis on Topics 2 and 3 overall, the Christian tradition places significantly more emphasis on “Political Geography and the Natural World”, while the Islamic tradition places more emphasis on “Political Geography and the Natural World”.

In addition to this “cross-sectional” analysis, they perform a time-series analysis, analyzing changes in these topic-emphasis proportions over time and relating them to historical events. For example, they find that the Medieval Islamic canon exhibits a rapid increase in focus on the “Religion and Faith” topic, occurring at the same time as a rapid increase in Turkic incursions into and conquests of Muslim cities and villages noted by historians. Delving deeper into this period, they find that the historical literature describes this era as one in which

Turkic rulers, like the Seljuks, sought to legitimize their political authority by presenting themselves as defenders of the faith, even waging “jihad against non-Muslim populations in Anatolia and Central Asia” (Blaydes et al. 2016, quoting

Lapidus 1996, 13).

Thus we see an example of the potentially fruitful role that automated content analysis can play in historiographic practice: the historical claim (“Turkic rulers . . . sought to legitimize”) can be substantiated or belied to some degree by analyzing machine-learned patterns within the text of a corpus. Though my present study concerns a cross-sectional corpus (*i.e.*, I don’t have series of “developing” *cahiers* over time), one could easily imagine time-varying extensions to the project (perhaps utilizing the entirety of the *Archives Parlementaires* with the associated dates for each text), some of which I describe in Section 5.

3 Exploratory (Unsupervised) Analysis

3.1 Obtaining and Cleaning the Data

English translations of the *cahiers* are few and far between⁴. A rich corpus of the full French text of nearly all *cahiers* known to exist, however, has been meticulously pieced together via the combined efforts of hundreds of researchers. For the purposes of this work I utilize the most complete and up-to-date list of *cahiers* and their sources available at the time of writing, that given in Appendix III of Shapiro and Markoff (1998)⁵

As a first step, I translate between the entries in the *cahier* list and the actual locations in the sources. For the *Archives Parlementaires*, Shapiro and Markoff employ a standardized index providing the start and end locations plus an indicator character ‘A’ or ‘B’ denoting whether the *cahier* begins on the left (‘A’) or right (‘B’) side of the page. To operationalize this index, I (a) obtained the Text Encoding Initiative (TEI) XML-formatted versions of the *Archives’* first six volumes, (b) employed Regular Expressions⁶ in Python to obtain a volume

⁴In my research we were able to locate 26 English translations scattered across 5 different works, a miniscule proportion of the approximately 60,000 total *cahiers*.

⁵I have digitized and reproduced this table in Appendix C herein.

⁶https://en.wikipedia.org/wiki/Regular_expression

number and page range for each *cahier*, and then (c) extracted each individual *cahier* via Regular Expressions run on the XML from (a)⁷. To ensure that no portions of the *cahiers* were lost, the indicator characters were ignored and any text appearing before or after the *cahier* of interest was filtered out later on in the pipeline.

Of the 357 general *cahiers* from the *Archives Parlementaires* listed in the table, I exclude 9 joint *cahiers* of the three estates and 5 *cahiers* of cities (Lyon, Metz, Rouen, Strasbourg, and Valenciennes), leaving me with 158 Noble *cahiers* and 185 Third Estate *cahiers*. In the next stage, the *cahiers* are further divided into individual grievances. I again employed Regular Expressions to detect the beginning and end of each grievance, checking for example if and where the string "Art. " followed by a numeral appears in the document⁸.

3.2 Model and Estimation

For my first analysis I minimize the imposition of particular hypotheses on the corpus by utilizing an unsupervised topic modeling approach. Following a similar procedure to that employed in Blaydes et al. (2018), I fit a Bayesian hierarchical topic model (Grimmer 2010) with the aim of comparing the top words in each topic across estates (Nobility vs. Third Estate), to gain insight on what aspects of each topic differ between the three estates (*e.g.*, whether an estate focuses more on bureaucratically- or constitutionally-oriented grievances regarding the legal system). Even a “null” result such as there being no significant differences between the Nobility and Third Estate *cahiers* would be interesting here, in that it would provide additional support for Tocqueville’s interpretation of the nobles as being essentially equal in radicalism to the Third Estate.

⁷Using Regular Expressions for XML parsing is typically frowned upon, as XML is a Chomsky Type 2 (Context-Free) language while Regular Expressions are Chomsky Type 3 (Regular). Unfortunately, however, the TEI XML files provided by the French Revolution Digital Archive are not valid XML, so Regular Expressions are required in this case, an issue researchers working with the files should take note of.

⁸This subtask could be automated (with improved performance) via a simple supervised machine learning approach, which I leave for future exploration.

The model is based on the following data-generating process (Grimmer 2010): Consider a “nationwide” Dirichlet distribution α , a prior distribution *over* distributions of topic importances which puts more weight on more likely distributions and, importantly, allows information sharing between the otherwise-independent bailliage \times estate pairs. Then each bailliage-estate b draws a particular distribution of topic importances $\pi_b = (\pi_{b1}, \pi_{b2}, \dots, \pi_{bK})$ over K topics from this α^9 and produces a set of N_b grievance topics $T_b = \{t_{b1}, t_{b2}, \dots, t_{bN_b}\}$, where each t_{bi} is independently drawn from this topic importance distribution π_b . Finally, the actual observed text of each grievance $g_{b1}, g_{b2}, \dots, g_{bN_b}$ is drawn from the set of topics T_b via a von Mises-Fisher distribution, chosen due to its agnosticism to the length of the observed documents, a property lacking in the case of the Normal distribution (Banerjee et al. 2005).

The advantages of this model over a non-hierarchical topic model are numerous, but in particular the explicit modeling of bailliage-estate ideology via the π_b distributions allows me to get an indication of the idiosyncrasies of each bailliage-estate. If one of the topics i corresponds to “radicalism” in some sense, for example, then a bailliage-estate (say, the Third Estate of Grenoble) with a high density on i can be interpreted as one with a high degree of “radicalism”. In an extension to the model available via the `expAgenda` package in R, such a relationship can be further explored by including covariates like the frequency of peasant revolts (Markoff 1990 – see analysis in Section 3.4 below), degree of urbanization (Shapiro et al. 1973), or rate of emigration (Greer 1951) within each bailliage.

3.3 Results

First, I present the 10 most-indicative words for each topic in Table 1. The topic labels I inferred from these terms are as follows:

⁹Importantly, the Dirichlet prior ensures that π_b lives in the K -simplex, or in other words that $\sum_{i=1}^K \pi_{bi} = 1$ and thus π_b is a valid probability distribution.

- (1) *Assembly Procedures*: The structure and decision-making processes that will be utilized in the upcoming Estates-General. Most likely concerns the issue of voting by order or by head, and how to represent the Third Estate.
- (2) *Judicial Reform*: Grievances/recommendations regarding criminal and civil law. Often mentioning specific judges or abuses observed in the particular bailliage.
- (3) *Rights: Seignorial, Commercial, and Civil*: Grievances regarding feudal structure, freedom of commerce, and (whether implicitly or explicitly) the Rights of Man and the Citizen.
- (4) *Ecclesiastical Reform*: Grievances concerning the structure and revenue-collection of the Church. Sometimes specifically corrupt dioceses or parish priests are mentioned.

Topic 1 (Assembly Procedures)	Topic 2 (Judicial Reform)	Topic 3 (Rights)	Topic 4 (Church Reform)
etat	vill	droit	ecclesiast
general	criminel	seigneur	vicair
provinc	civil	commerc	dioces
nation	juge	libert	paroiss
assembl	ministr	privileg	conseil
provincial	libert	office	revenus
consent	municipal	citoyen	abbay
bailliag	proc	religion	communaut
ordre	justic	lettre	avril
mesur	abus	tarif	ordon

Table 1: 10 Most Indicative Words for Each Topic

Next, I plot the relative frequencies of each topic for both the Nobility and the Third Estate in Figure 2. From this plot we immediately see two patterns. First, the small differences between the two estates (relative to the large differences between topics) lend credence to Tocqueville’s interpretation of the Noble *cahiers* as being quite similar both linguistically and in terms of radicalism of demands to the Third Estate.

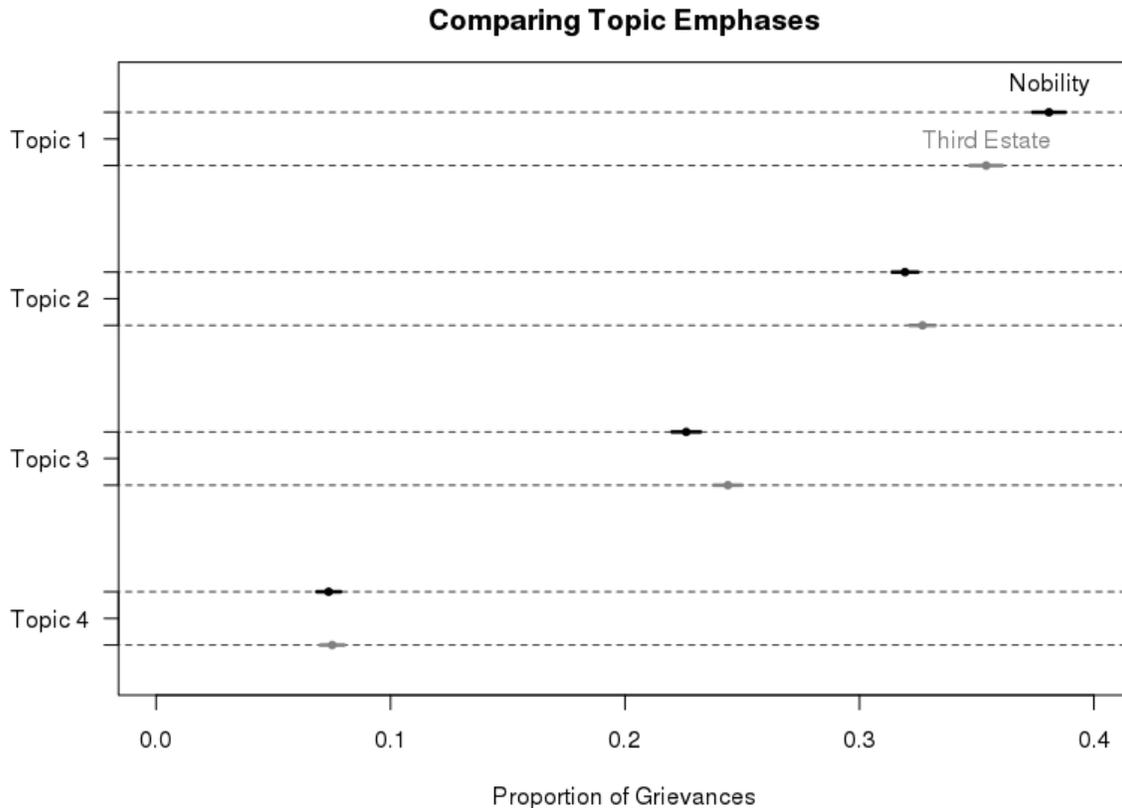


Figure 2: Relative Topic Frequencies by Estate

However, although differences between the estates with respect to Judicial and Ecclesiastical Reforms are not statistically significant, we *do* observe statistically significant differences in emphasis between the estates with respect to Assembly Procedures and Rights. While the Nobility has a more “narrow” focus on assembly procedures in the immediate future (an assembly which many Nobles will have the chance to attend while very few members of the Third Estate will have the same opportunity), the Third Estate is significantly more vocal about the seigneurial, commercial, and civil freedoms which affect their day-to-day lives at home. Thus, through an unsupervised “exploratory” analysis of the *cahiers*, I am able to verify Tocqueville’s assertions in broad strokes while also “zooming in” and gaining a more nuanced understanding of the particular institutions of the *ancien régime* for which

his analysis was most and least accurate.

Strikingly, the two topics on which the estates differ comport with the comparative analysis in Chapter 15 of Shapiro and Markoff (1998), titled “Consensus and Conflict at the Onset of Revolution”. Of the 41 subjects they analyze for differences between the estates in Table 15-4 (reproduced here in Table 2), only four exhibit statistically significant differences: “Estates-General”, “Economy (miscellaneous)”, “Criteria of mobility”, and “Seigneurial regime”. Given the strong association of the words in Topic 1 with the procedures of the Estates-General, and of the words in Topic 3 with Seigneurial rights (with “seigneur” being the most indicative word for this topic besides “droit”), we therefore have a quite remarkable result: that the two topics which significantly differed in my analysis between the Nobility and Third Estate precisely align with two of the four manually-tagged topics that Shapiro and Markoff’s analysis finds to have significant differences between the estates.

It is also noteworthy that the two topics found to have *non*-significant in my unsupervised analysis align with non-significant topics in Shapiro and Markoff’s analysis. First, my non-significant “Judicial Reform” topic corresponds to Shapiro and Markoff’s “Justice” top-level code, within which none of the sub-topics are found to be significant: “Justice (general)”, “Criminal prosecution”, “Courts”, “Civil law”, “Due process”, “Enforcement agents of courts”, and “Legal professions”. Similarly, my non-significant “Church Reform” topic corresponds to Shapiro and Markoff’s top-level “Religion” code, within which again they find no significant differences in any of the sub-topics: “Religion (miscellaneous)”, “Religion (general)”, “Church finances”, “Church-state relations”, “Tithe”, and “Morality”.

3.4 Adding Covariates

While this “baseline” comparison of the two estates via automated text analysis already reveals a great number of interesting aspects of the *cahiers*, as evidenced by its accordance

Code0	Code1	Topic	Agreement	Predicted	Residual
0	CO	Colonies	0.025	0.031	-0.006
0	NC	Non-Catholics	0.101	0.117	-0.016
1	X	General	0.154	0.152	0.002
C	0	Constitution (misc)	0.279	0.274	0.005
C	1	Constitution (general)	0.190	0.207	-0.017
C	EG	* Estates-General	0.343	0.412	-0.070
C	NA	Nation	0.460	0.491	-0.030
C	PL	Political liberties	0.284	0.312	-0.028
C	PM	Powers of monarch	0.560	0.546	0.014
E	0	* Economy (misc)	0.086	0.172	-0.086
E	1	Economy (general)	0.354	0.338	0.016
E	AG	Agriculture	0.126	0.143	-0.016
E	CO	Commerce	0.247	0.252	-0.005
E	FI	Finance	0.180	0.194	-0.014
E	IN	Industry	0.149	0.155	-0.005
E	TR	Transportation	0.092	0.101	-0.009
G	1	Government (general)	0.045	0.039	0.006
G	AA	Administrative agencies	0.178	0.173	0.005
G	FI	Finances	0.166	0.178	-0.012
G	KI	King	0.275	0.278	-0.003
G	MI	Military	0.120	0.159	-0.039
G	RL	Regional and local gov't	0.242	0.253	-0.011
G	TA	Taxation	0.190	0.219	-0.029
J	1	Justice (general)	0.198	0.193	0.005
J	CM	Criminal prosecution	0.247	0.252	-0.006
J	CO	Courts	0.184	0.207	-0.023
J	CV	Civil law	0.173	0.171	0.002
J	DP	Due process	0.359	0.378	-0.019
J	EA	Court enforcement agents	0.274	0.263	0.011
I	LP	Legal professions	0.150	0.143	0.007
R	0	Religion (misc)	0.053	0.040	0.013
R	1	Religion (general)	0.111	0.011	0.100
R	CF	Church finances	0.240	0.289	-0.049
R	CL	Clergy	0.224	0.230	-0.006
R	CO	Church organization	0.141	0.123	0.018
R	CS	Church-state relations	0.188	0.198	-0.010
R	DI	Tithe	0.170	0.186	-0.016
R	MO	Morality	0.255	0.267	-0.012
S	I	Stratification (general)	0.050	0.096	-0.047
S	CM	* Criteria of mobility	0.212	0.300	-0.087
S	SE	* Seigneurial regime	0.148	0.203	-0.055

Table 2: Table 15-4 from Shapiro and Markoff (1998), illustrating significant and non-significant topics with respect to disagreement between Third Estate and Nobility *cahiers*. * indicates statistical significance.

with the theories of Tocqueville and Shapiro and Markoff described above, the next stage of a substantive content analysis entails explaining *why* these variations occur – what socio-demographic variables are able to explain the observed differences between the Nobility and the Third Estate *cahiers*?

To investigate this question, I first decided to explore the role that radicalism of the peasantry – measured via the number of peasant revolts in a bailliage – had on the contents of the *cahiers* of the Nobility and the Third Estate. Did the Nobility or Third Estate in bailliages with more peasant revolts pay more attention to (for example) reforms, or to the Rights of Man and the Citizen, as measured through emphasis on Topics 2 and 4 and Topic 3 in our results, respectively, out of fear that without reforms their countrysides would explode? Or did the drafters in these more unstable bailliages *under*-emphasize these topics, perhaps hoping to push instead for more “law and order” at the Estates-General without making concessions to the peasantry?

To answer this question, I merged my topic-emphasis data with data on the frequency of peasant revolts between 1788 and 1793 from Markoff (1996)¹⁰. Aggregate monthly counts of these revolt events are given in Figure 3.

The results of my covariate analysis are presented in Figure 4. As can be seen from the plots, there is essentially no relationship between the number of peasant revolts in a bailliage and the distribution of emphasis that the bailliage places on our four topics. Indeed, even when the data is split into pre- and post-Cahier-drafting sets (thus mirroring Shapiro and Markoff’s methodology more directly and allowing us to move a step closer towards studying causality) I still observe no relationship, as illustrated in Figures 5 and 6.

As a final test that most closely matches the method of Chapter 21 in Shapiro and

¹⁰Two of the bailliages, Vitry-Le-Francois and Villers-La-Montagne, had no revolt data in the dataset I was able to obtain. This is most likely due to the fact that this dataset was itself based on Optical Character Recognition of an on-paper data table, which probably got cut off before the end, as these are the alphabetically-last bailliages in the data.

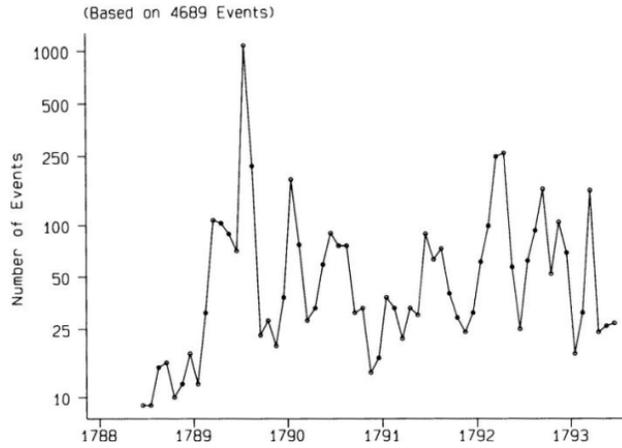


Figure 3: Monthly revolt counts from Markoff (1996).

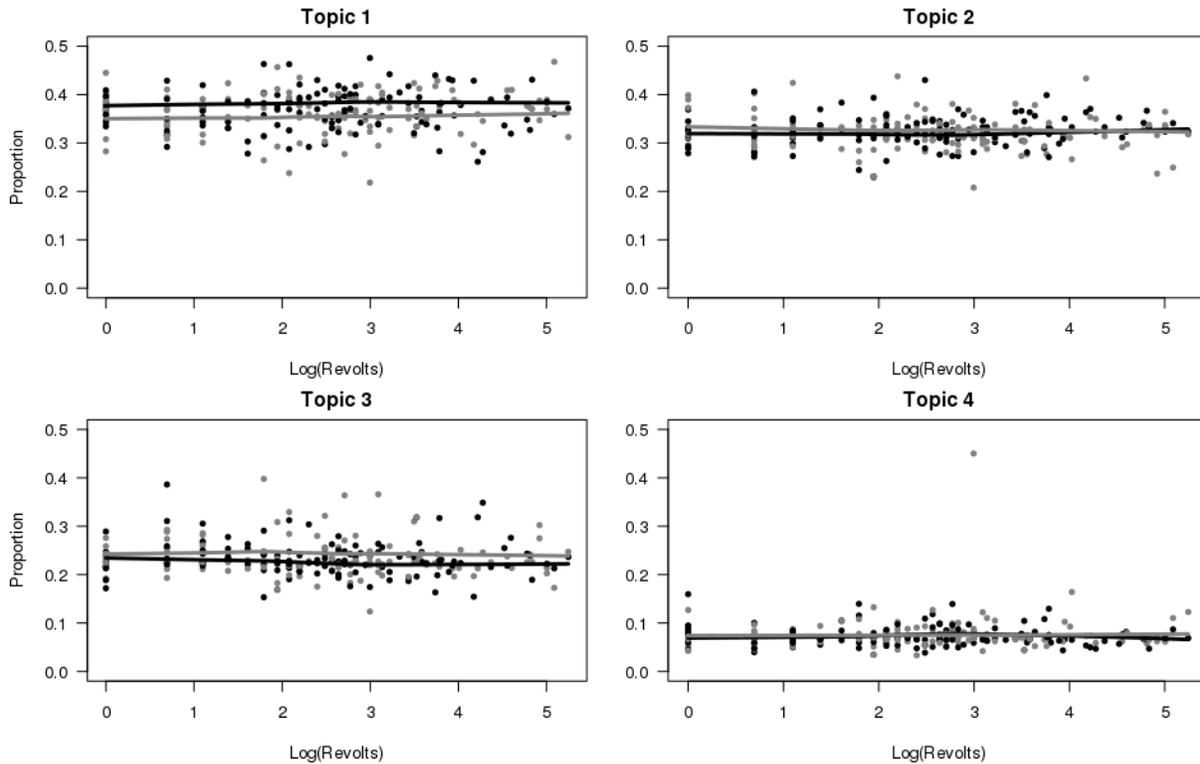


Figure 4: The relationship between peasant revolt frequency and distribution of emphasis among our four topics, for both Nobility and Third Estate *cahiers*. As in Figure 2, black indicates Nobility while grey indicates Third Estate data.

Markoff (“Peasant Grievances and Peasant Insurrection”)¹¹, I collapse the number-of-revolts

¹¹I emphasize here that this is an implementation of Shapiro and Markoff’s *methodology*, not a *replication* of their results, since the latter would require me to focus my analysis on only the Parish *cahiers*, which I

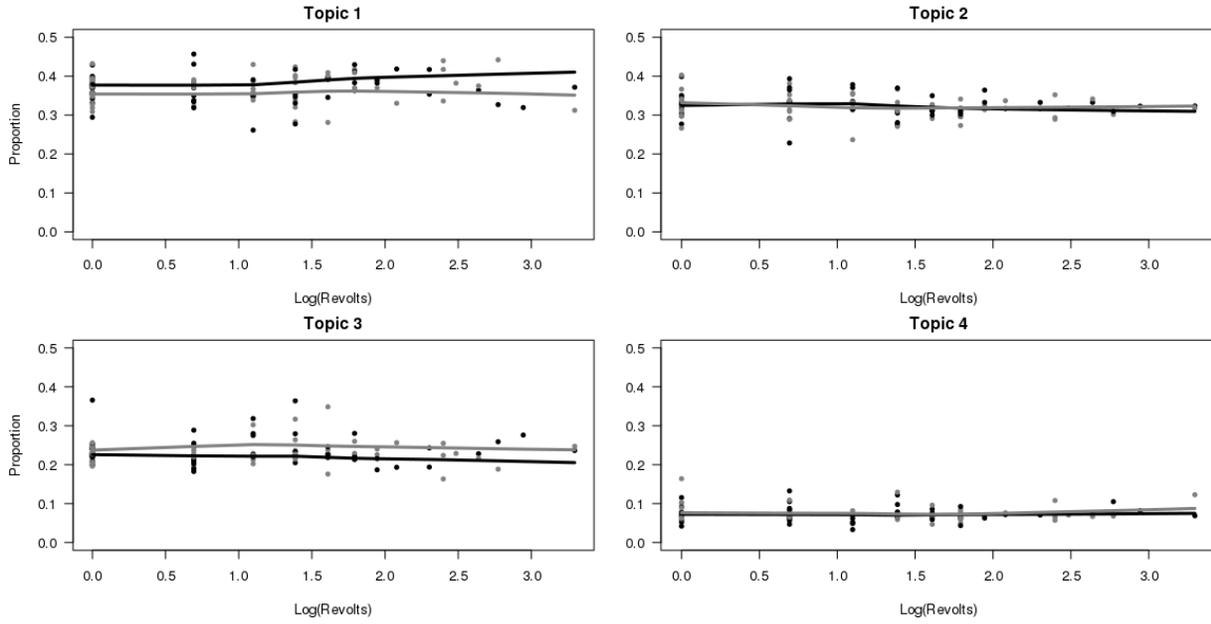


Figure 5: The relationship between **pre-cahier**-drafting peasant revolt frequency and distribution of emphasis among our four topics, for both Nobility and Third Estate *cahiers*. As in Figure 2, black indicates Nobility while grey indicated Third Estate data.

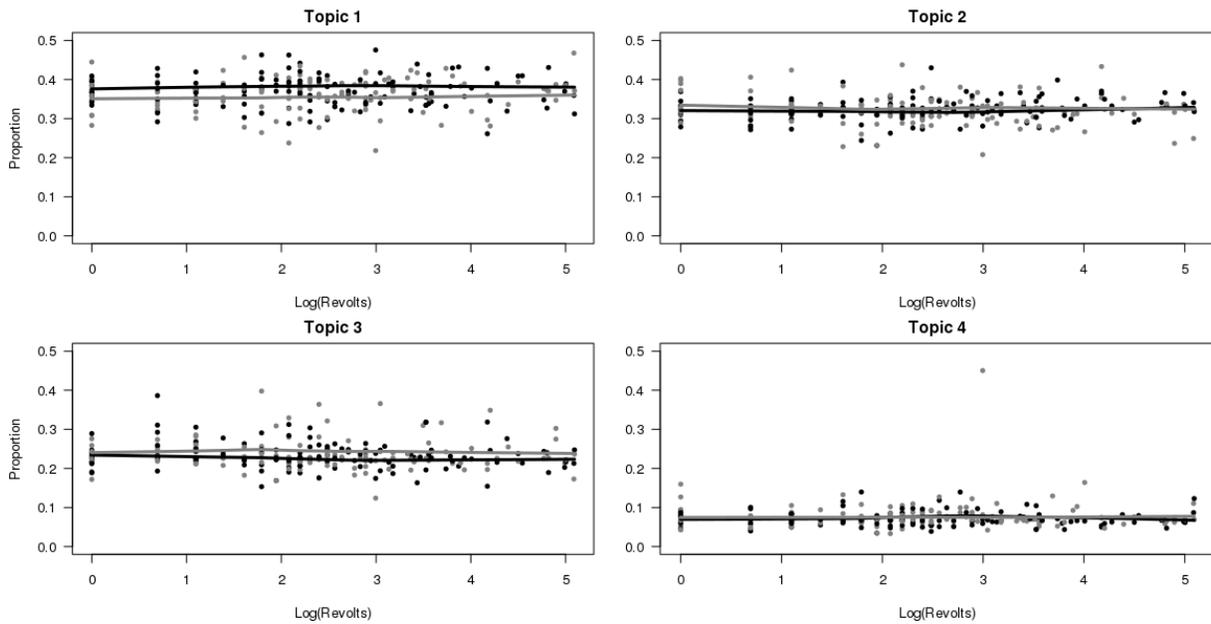


Figure 6: The relationship between **post-cahier**-drafting peasant revolt frequency and distribution of emphasis among our four topics, for both Nobility and Third Estate *cahiers*. As in Figure 2, black indicates Nobility while grey indicated Third Estate data.

plan to do in future work.

predictor variable into a binary any-revolts variable, regress the topic proportions on this binary variable separately for the Third Estate and Nobility *cahiers*, then compare the regression coefficients using the method of Clogg et al. (1995). Results from this analysis are presented in Table 3, and again we see a lack of any significant effect of revolts on topic emphasis, nor any significant difference between Nobility and Third Estate *cahiers* with respect to the effect of revolts on topic emphases. For the latter comparison, I used the following equation from Clogg et al. (1995):

$$z = \frac{\beta_{Nobility} - \beta_{TE}}{\sqrt{SE(\beta_{Nobility})^2 + SE(\beta_{TE})^2}}$$

then computed the two-tailed p -values (since I'm testing $H_0 : \beta_{Nobility} = \beta_{TE}$) in standard fashion as $2(1 - \Phi(|z|))$, where $\Phi(\cdot)$ is the Normal CDF. I repeated this analysis separately for pre-drafting and post-drafting revolts with substantively similar results (available upon request).

With these results, we can tentatively conclude that unlike the Parish *cahiers*, the contents of which Shapiro and Markoff (1998) finds to be significantly associated with the number and type of revolts in the parishes, the Nobility and Third Estate *cahiers* show no significant association with the existence or frequency of these peasant insurrections. In general, because these results are conditional upon the salience of the automatically-discovered topics, significant results may indeed obtain for a different set of topics. However, based on my earlier argument for the salience of these topics and their strong correspondence with topics manually created in Shapiro and Markoff's analysis, I conclude that we have again uncovered an interesting result, this time regarding the (lack of) impact of a socio-political variable on the contents of the *cahiers*. I also emphasize that this analysis can be repeated for other covariates of interest to scholars of the French Revolution, such as those mentioned above (urbanization, population density, and rate of emigration) which are immediately obtainable

and which I am making available on the project's GitHub page.

	Topic 1		Topic 2	
	Nob Emphasis	TE Emphasis	Nob Emphasis	TE Emphasis
Any Revolts	0.009 (0.012)	0.008 (0.016)	-0.001 (0.008)	0.006 (0.014)
Constant	0.374*** (0.011)	0.347*** (0.016)	0.320*** (0.008)	0.321*** (0.013)
Observations	135	139	135	139
R^2	0.004	0.002	0.0001	0.001
Adjusted R^2	-0.003	-0.006	-0.007	-0.006
Coeff. Diff. z -score		0.048		-0.424
Coeff. Diff. p -value		0.962		0.672
<i>Note:</i>	* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$			

	Topic 3		Topic 4	
	Nob Emphasis	TE Emphasis	Nob Emphasis	TE Emphasis
Any Revolts	-0.012 (0.009)	-0.013 (0.016)	0.005 (0.011)	-0.0003 (0.008)
Constant	0.237*** (0.008)	0.257*** (0.015)	0.069*** (0.011)	0.075*** (0.008)
Observations	135	139	135	139
R^2	0.015	0.005	0.001	0.00001
Adjusted R^2	0.007	-0.002	-0.006	-0.007
Coeff. Diff. z -score		0.037		0.367
Coeff. Diff. p -value		0.970		0.714
<i>Note:</i>	* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$			

Table 3: Results from the regressions of topic emphasis on an any-revolts dummy for both Third Estate and Noble *cahiers*. Results from the coefficient-comparison test (Clogg et al. 1995) are presented on the final two lines of each regression summary, and indicate no significant differences between the relationship for Nobility and Third Estate *cahiers*.

4 Extension: Supervised Analysis and Calibration

Although the unsupervised model gives us insights into the relative importance of *ancien régime* reforms among the three estates, without any *a priori* hypotheses, this exploratory endeavor should be coupled with content-analytic statistical testing once hypotheses have been generated. Hence, in this section, I develop a *supervised* content analysis model which enables accurate statistical inference from a small number of training examples (document tags) regarding the characteristics of the entire corpus. With this model, one can quantify exactly how the quality of inference degrades as the number of training examples gets smaller and smaller, producing a “sample size determination curve” researchers can then use to derive the minimal number of manual tags necessary to achieve a desired level of accuracy¹².

4.1 Obtaining and Cleaning the Data

A supervised study of the *cahiers* in this mode requires both separate data (the individual tags from Shapiro and Markoff (1998)) and a separate data-cleaning process, as one needs to process the *cahier* texts in such a way that they can unambiguously pair each tag with a specific *cahier*. As presented in the FRAS program manual, the tag data (exported from the program as described in Section 2) contains the following information for each *cahier*:

- A unique numerical grievance ID
- The bailliage number (mapped to bailliage names via the table in Appendix B)
- The estate (N for Nobility, T for Third Estate, P for Parishes)
- The page number: ‘A’ or ‘B’ for the left and right sides of the page, in the case of the *Archives Parlementaires*

¹²The method developed in this section is being pursued for a separate forthcoming work.

- Location on the page of the beginning of the discussion, measured in tenths of an inch from the top
- Abbreviation of the source and volume number (mapped to the full source information via the list in Appendix D)
- Abbreviated version of the bailliage name
- (In the case of Parish files only) PAR XX, where XX is a serial index identifying the particular parish that expressed the grievance. The full names of the parishes and the document sources are listed in Appendix C.

Computationally, since the TEI XML format does not record spatial details such as inches from the top of the page, the start and end locations of each tagged grievance needed to be approximated based on some combination of (a) the locations in tenths-of-an-inch given in the tag export data, (b) the total number of grievances recorded as starting on each page, and (c) the automated grievance partitions computed via Regular Expressions during the unsupervised study.

Though a seemingly daunting task *prima facie*, recent advances in Neural Networks (a powerful but under-theorized machine learning framework which, in practice, radically outperforms “standard” machine learning algorithms in learning complex high-dimensional and non-linear functions¹³) allow one to model the task as an “end-to-end” learning problem, feeding in the quantities (a), (b), and (c) and providing negative feedback when a grievance produced by the algorithm does not match the intended grievance. In this manner, once the Neural Network has been sufficiently trained, retrieving the grievances for each tag becomes a trivial problem.

¹³For a clear and thorough introduction, see Nielsen (2015), or Goodfellow et al. (2016) for a more in-depth description of the topic. For a survey of Natural Language Processing-specific applications of Neural Networks, see Goldberg (2017).

The only remaining task for such an analysis would then be the following: starting with *all* of the (*cahier*, tag) pairs, check that the Neural Network indeed converged by verifying whether the resulting tag distributions match the distribution given in Table 4. The remainder of the procedure then entails simply removing some proportion p of the *cahiers* in a stepwise fashion, having the supervised machine learning algorithm infer the missing tags in each step. While Support Vector Machines are the most common algorithms for this type of classification problem, the Random Forest algorithm outperformed Support Vector Machines in Stewart and Zhukov (2009), thus I opt for Random Forests to perform this inference step. Finally, to evaluate the performance of the supervised algorithm as a function of the proportion of tags removed p , a simple metric such as Mean Squared Error can be used:

$$MSE(p) = \sum_{i \in S}^N (t_i(R_p) - a_i(R_p))^2,$$

where R_p is the set of pN grievances removed from the N total grievances, S is the set of possible tag values, $t_i(R_p)$ is the true proportion of tag i among the removed grievances R_p (*i.e.*, the values in Table 4 scaled to only incorporate the pN removed grievances), and $a_i(R_p)$ is the proportion of tags with value i in the algorithm’s set of guessed tags for the grievances in R_p . With this quantity $MSE(p)$ plotted for all values of $p \in [0, 1]$, the researcher thus obtains a curve indicating precisely the extent to which the quality of the automatic tagging degrades as fewer and fewer training examples are supplied. By choosing a maximum desired error threshold, then, they can “read off” how many manual tags are necessary directly from the plot.

Rank	Third Estate	Nobility	Parishes	Subject
1	68.182	49.398	14.666	Relegate the customs to the frontiers of the kingdom.
2	55.556	6.627	4.941	That Third Estate be permitted access to posts and careers in military.
3	54.545	56.627	5.340	Grant the government the power to tax for a limited period only.
4	37.374	22.289	12.582	Bring courts closer to the people concerned.
5	36.364	40.964	0.995	Abolish censorship provided that editors and writers be held liable.
6	28.788	32.530	4.006	Connaissance de la dette (publication of the debt).
7	27.778	30.120	0.836	That beneficiaries be required to reside where they hold a benefice.
8	27.778	4.217	3.317	Let there be a single tax administration roll.
9	26.263	2.410	1.564	That Third Estate be permitted access to posts and careers in church.
10	24.747	16.265	1.684	That Sveques (bishops) be required to reside in the district of their post.

Table 4: The ten most popular demands in the Third Estate’s cahiers de doléances, from Shapiro and Markoff (1998)

Code	Category	Third Estate		Nobility		Parishes		Total
		Count	%	Count	%	Count	%	
0	Miscellaneous	364	0.80	154	0.60	348	1.28	866
1	General	2109	4.65	1281	5.00	966	3.54	4356
C	Constitution	3862	8.52	3803	14.84	1156	4.24	8821
E	Economy	7048	15.54	2421	9.44	5369	19.69	14838
G	Government	14856	32.76	9662	37.69	11271	41.34	35789
J	Judiciary	9245	20.38	5033	19.63	3583	13.14	17861
R	Religion	4197	9.25	1803	7.03	2280	8.36	8280
S	Stratification System	3672	8.10	1476	5.76	2294	8.41	7442
Total		45353		25633		27267		98253

Table 5: Summary statistics for the grievances cataloged in Shapiro and Markoff (1998).

5 Conclusion

I have developed unsupervised and supervised models for automated content analysis, enabling both exploratory analyses to be conducted with no manual tagging and hypothesis-testing-based analyses to be conducted with a provably minimal number of manual tags.

The methods immediately suggest that fruitful insights can be derived from corpora too large for manual content analyses or for which a dearth of resources has precluded manual content analyses. Bearing a somewhat striking resemblance to the *cahiers*, the letters, speeches, and pamphlets of the “100 Flowers Period” of Maoist China (Nieh 1981) remain understudied. To this end, I have digitized and performed preliminary analyses of these documents, with results and data available upon request.

Slightly further afield, lyrics from the African-American musical tradition – from 17th-century spirituals to present-day hip hop – constitute an immeasurably rich source of sentiments regarding socio-political norms and institutions from a traditionally marginalized and excluded population. Thus I have also scraped a large corpus of more than 2 million song lyrics from both the web and digitized songbooks (available upon request), and I plan to conduct an automated content analysis on this data in the near future. The structure of this data presents a new set of challenges but an important new dimension of analysis: change over time. As each song is tagged with a date, I can not only analyze relative sentiment between groups (say, regions of the U.S.), but also changes in sentiment across time (say, before and after the abolition of slavery, the Civil Rights movement, or the election of Barack Obama)¹⁴.

References

- [1] Andrews, George. 1935. Review of *French Nationalism in 1789 According to the General Cahiers*, by Beatrice Hyslop. *The Journal of Modern History*, 7(2): 212–213. [http:](http://)

¹⁴For more on music lyrics as grievance documents, see <https://parkjaehyuk.github.io/curr-projects.html>, Project 2

[//www.jstor.org/stable/1872881](http://www.jstor.org/stable/1872881)

- [2] Banerjee, Arindam, Inderjit S. Dhillon, Joydeep Ghosh, and Suvrit Sra. 2005. “Clustering on the Unit Hypersphere using von Mises-Fisher Distributions”. *Journal of Machine Learning Research*, 6: 1345–82. <http://www.jmlr.org/papers/volume6/banerjee05a/banerjee05a.pdf>
- [3] Blaydes, Lisa, Justin Grimmer, and Alison McQueen. 2018. “Mirrors for Princes and Sultans: Advice on the Art of Governance in the Medieval Christian and Islamic Worlds”. *Journal of Politics*, forthcoming. http://stanford.edu/~jgrimmer/BGM_final.pdf
- [4] Blei, David. 2012. “Topic Modeling and Digital Humanities”. *Journal of Digital Humanities*, 2(1): 8–11. <http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/>
- [5] Blei, David, Andrew Ng, and Michael I. Jordan. 2003 “Latent Dirichlet Allocation”. *Journal of Machine Learning Research*, 3: 993–1022. <http://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf>
- [6] Clogg, Clifford, Eva Petkova, and Adamantios Haritou. 1995. “Statistical Methods for Comparing Regression Coefficients Between Models”. *American Journal of Sociology*, 100(5): 1261–1293. <http://www.journals.uchicago.edu/doi/10.1086/230638>
- [7] Doyle, William. 1980. *Origins of the French Revolution*. New York: Oxford University Press.
- [8] Elster, Jon. 2006. “Tocqueville on 1789: Preconditions, Precipitants, and Triggers”. In Welch, Cheryl (Ed.), *The Cambridge Companion to Tocqueville*, 49–80. New York: Cambridge University Press.
- [9] Furet, François. 1988. *The French Revolution 1770–1814*. Oxford: Blackwell.
- [10] Gentzkow, Matthew and Jesse Shapiro. 2010. “What Drives Media Slant?: Evidence from US Daily Newspapers”. *Econometrica*: 78(1): 35–71. <http://onlinelibrary.wiley.com/doi/10.3982/ECTA7195/abstract>
- [11] Gille, Bertrand. 1980. *Les Sources Statistiques de L’Histoire de France, des Enquêtes du XVII^e Siècle a 1870*. Geneva: Librairie Droz.
- [12] Goldberg, Yoav. 2017. *Neural Network Methods for Natural Language Processing*. London: Morgan and Claypool. <http://www.morganclaypool.com/doi/abs/10.2200/S00762ED1V01Y201703HLT037>
- [13] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Cambridge: MIT Press. <http://www.deeplearningbook.org/>

- [14] Greer, Donald. 1951. *The Incidence of the Emigration During the French Revolution*. Cambridge: Harvard University Press.
- [15] Grimmer, Justin. 2010. “A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases”. *Political Analysis*, 18: 1–35. <https://web.stanford.edu/~jgrimmer/ExpAgendaFinal.pdf>
- [16] Grimmer, Justin and Brandon Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts”. *Political Analysis*, 21(3): 267–297. <https://web.stanford.edu/~jgrimmer/tad2.pdf>
- [17] Hunt, Lynn. 1980. *Politics, Culture, and Class in the French Revolution*. Berkeley: University of California Press.
- [18] Hyslop, Beatrice. 1933. “Répertoire Critique des Cahiers de Doléances pour les États Généraux de 1789”. PhD Dissertation, Columbia University.
- [19] Hyslop, Beatrice. 1934. *French Nationalism in 1789 According to the General Cahiers*. New York: Columbia University Press.
- [20] Hyslop, Beatrice. 1936. *A Guide to the General Cahiers of 1789*. New York: Columbia University Press.
- [21] Jensen, Jacob, Ethan Kaplan, Suresh Naidu, and Laurence Wilse-Samson. 2012. “Political Polarization and the Dynamics of Political Language: Evidence from 130 Years of Partisan Speech”. *Brookings Papers on Economic Activity*, 1–81. <http://www.jstor.org/stable/41825364>
- [22] Lapidus, Ira. 1984. *Muslim Cities in the Later Middle Ages*. Cambridge: Cambridge University Press.
- [23] Lefebvre, Georges. 1932. *La Grande Peur de 1789*. Paris: Librairie Armand Colin.
- [24] Lefebvre, Georges. 1947. *The Coming of the French Revolution*. Princeton: Princeton University Press.
- [25] London, Jennifer. 2016. “Re-Imagining the Cambridge School in the Age of Digital Humanities”. *Annual Review of Political Science*, 19: 351–373. <http://www.annualreviews.org/doi/abs/10.1146/annurev-polisci-061513-115924>
- [26] Markoff, John. 1990. “Peasant Grievances and Peasant Insurrection: France in 1789”. *The Journal of Modern History*, 62(3): 445–476. <https://www.jstor.org/stable/1881173>
- [27] Markoff, John. 1996. *The Abolition of Feudalism: Peasants, Lords, and Legislators in the French Revolution*. University Park: Penn State University Press.

- [28] Mavidal, Jérôme and Émile Laurent. 1879. *Archives Parlementaires de 1787 à 1860, Recueil Complet des Débats Législatifs et Politiques des Chambres Françaises*. (Volumes 1–6). Paris: Bibliothèque National de France. <https://frda.stanford.edu/en/ap>
- [29] Nieh, Hualing. 1981. *Literature of the Hundred Flowers* (2 Volumes). New York: Columbia University Press.
- [30] Nielsen, Michael A. 2015. *Neural Networks and Deep Learning*. San Francisco: Determination Press. <http://neuralnetworksanddeeplearning.com/>
- [31] Popper, Karl. 1959. *The Logic of Scientific Discovery*. Abingdon: Routledge.
- [32] Řehůřek, Radim and Petr Sojka. 2010. “Software Framework for Topic Modelling with Large Corpora”. *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. <https://radimrehurek.com/gensim/about.html>
- [33] Shapiro, Gilbert, John Markoff, and Sasha Weitman. 1973. “Quantitative Studies of the French Revolution”. *History and Theory*, 12(2): 163–191.
- [34] Shapiro, Gilbert and John Markoff. 1998. *Revolutionary Demands: A Content Analysis of the Cahiers de Doléances of 1789*. Palo Alto: Stanford University Press.
- [35] Stewart, Brandon and Yuri Zhukov. 2009. “Use of Force and Civil-Military Relations in Russia: An Automated Content Analysis”. *Small Wars and Insurgencies*, 20(2): 319–343. https://scholar.harvard.edu/zhukov/files/2009_StewartZhukov_SWI.pdf
- [36] Tackett, Timothy. 1986. *Religion, Revolution, and Regional Culture in Eighteenth-Century France: The Ecclesiastical Oath of 1791*. Princeton: Princeton University Press.
- [37] Tackett, Timothy. 1996. *Becoming a Revolutionary: The Deputies of the French National Assembly and the Emergence of a Revolutionary Culture (1789-1790)*. Princeton: Princeton University Press.
- [38] Tackett, Timothy. 2004. “Introduction”. In Lefebvre, Georges, *The Coming of the French Revolution*, vii–xxx. Princeton: Princeton University Press.
- [39] Taylor, George V. 1972. “Revolutionary and Nonrevolutionary Content in the Cahiers of 1789: An Interim Report”. *French Historical Studies*, 7(4): 479–502. <https://www.jstor.org/stable/286194>
- [40] Tilly, Charles. 2008. *Contentious Performances*. New York: Cambridge University Press.
- [41] de Tocqueville, Alexis. 1856. *The Ancien Régime and the French Revolution*. New York: Cambridge University Press.

A Additional Data Tables from Shapiro and Markoff (1998)

B Bailliage List from Shapiro and Markoff (1998)

C List of *Cahiers* from Shapiro and Markoff (1998)

D *Cahier* Source Index for Shapiro and Markoff (1998)

[AP] Mavidal, J., and E. Laurent, eds. Archives Parlementaires de 1787 a 1860, Recueil c
(XML on FRDS)

[BSN1] Boissonnade' p- &d- Cahiers de doliances de la senechaussee d'Angouleme et du sie
(PDF downloaded from Google books. BSN1.pdf)

[HG] Hyslop, Beatrice Fry, ed. A Guide to the General Cahiers of 1789. New York: Columbi
(Have physical book)

[CHAR] Charmasse, A. de, ed. Cahiers des paroisses et es du bailliage d'Autun pour les E
(PDF+TXT downloaded from BnF Gallica. CHAR.pdf, CHAR.html)

[RF32] Revolution frangaise, revue d'histoire moderne et contemporaine, 32 (1897).

[VERN] Vernier,, J., ed. Cahiers de doleances du bailliage de Troyes (principal et secon
(PDF+TXT downloaded from HathiTrust. <https://babel.hathitrust.org/cgi/pt?id=osu.32435023>
VERN.pdf, VERN.html)

[RECA] Recamier, E., ed. Les Deputes des communes du Bugey en 1789 et en 1876. Paris: A.
(Downloaded from HathiTrust. [https://babel.hathitrust.org/cgi/pt?id=wu.89046624557;view=](https://babel.hathitrust.org/cgi/pt?id=wu.89046624557;view=RECA.pdf)
RECA.pdf, RECA.html)

[LESU] * Lesueur, F., and A. Cauchie, eds. Cahiers. de doleances des bailliages de Blois

[AL16] Lesprand, L'Abbe P., "Cahiers lorrains de 1789," Annuaire de la societe d'histoir

[BV44] Bulletin de la societe d'tudes scientifiques et archeologiques du Vat, 44 (1942-

[ROUY] Rouy, H.; ed. Souvenirs sedanais, Series 4, Vol. n. Sedan: Imprimerie de J. Laroc

[LAUR] * Laurent, Gustave, ed. Cahiers de doleances pour les Etats generaux de 1789.1: B

[BRID] * Bridrey, Emile, ed. Cahiers de doleances du bailliage de Qotentin (Coutances et

[DHML] Duhamel, L., ed. Documents rates ou inedits de l'histoire des Vosges. Vol. II. Pa

[ETN2] * Etienne, Charles, ed. Cahiers du bailliage de Dieuze. ("Cahiers des bailliages

[DSH] Thomas-Lacroix, P., ed. Les Cahiers de doleances de la senechaussee d'Hennebont. M

[FOUR] * Fournier, J., ed. Cahiers de doleances de la s6nechaussee de Marseille pom le&E

- [AL15] Lesprand, L'Abbe P. "L'Election du depute et cahier du ti etat de la ville de Met
[MART] 'Martin, E., ed. Cahiers de doleances du bailliage deMirecourt. Epinal and Paris:
[GODF] 'Godfrin, Jean, ed. Cahiers du bailliage de Nancy. ("Cahiers des bailliages des g
[BLIG] * Bligney-Bondurand, E., ed. Cahiers de doleances de la senechaussee de Nimes pom
[BLCH] * Bloch, Camille, ed. Cahiers de doleances du bailliage d'Orleans pom les Etats g
[CHAS] Chassin, Charles L., ed. Les Elections et les de Paris en 1789. Vol. III. Paris:

E List of *Cahiers* from Hyslop (1936)

F Source Index for Hyslop (1936)