# History Repeats: Historical Phase Recognition from Short Texts

Fabio Celli[1,*], Valerio Basile[2]

[1]Gruppo Maggioli, Via Bornaccino 101, Santarcangelo di Romangna, 47822, Italy
[2]Università di Torino, Via Pessinetto 12, 10149, Torino, Italy

**Abstract**

This paper introduces a new multi-class classification task: the prediction of the Structural-Demographic phase of historical cycles - such as growth, impoverishment and crisis - from text describing historical events. To achieve this, we leveraged data from the Seshat project, annotated it following specific guidelines and then evaluated the consistency between three annotators. The classification experiments, with transformers and Large Language Models, show that 2 of 5 phases can be detected with good accuracy. We believe that this task could have a great impact on comparative history and can be helped by event extraction in NLP.

**Keywords**

Cultural Analytics, Structural Demographic Theory, LLMs, NLP for the Humanities,

## 1. Introduction And Background

In the last decade, at least since Brexit [1], many countries in the world experienced a generalized polarization and phenomena of toxic language online have grown [2]. Hate speech [3], misogyny [4], conspiracy theories [5] and related phenomena are just visible manifestations of deep structural social crises, ushering in periods of shifting world order [6]. While crises may appear sudden, they are often rooted in underlying factors like demographics, geopolitics, technological advancements, and historical-economic cycles. Using scientific method, mathematical modelling and the Structural Demographic Theory (SDT) [7] it was possible to formalise secular cycles [8], that typically last between 75 to 100 years [9], and predict outbreaks of political instability in complex societies based on the rate of past crises [10]. The SDT defines three actors and five phases of the secular cycle. The three key actors are:

- The population, which is the source of the society's resources and manpower, represents approximately 90% of the entire society and is the part that follows instructions to produce goods and wealth, consuming only a small part of it.
- The elites, who typically cover around 8% of the society, are the groups of people in charge of finding potential solutions to the problems of the society and are eligible to become part of the state. Who is considered part of the elite and how someone gains or loses elite status depends on the type of government and the power dynamics within a society.
- The state, formed by roughly 2% of the society, is the government that enforces its will and manages resources from the population. It is composed by one or more elite groups, depending on the social structure, and it crystallizes the culture to keep the society alive.

The actors interact in five phases during the secular cycle, progressively increasing social and political instability:

1. The growth phase. During this phase a fresh and effective culture creates social cohesion, the economy is growing rapidly and the state is expanding its control over the population. This leads to increased economic prosperity and stability but raises the problem of sustainability. Periods of reconstruction immediately following wars, like post-war Italy in the 1950s, are examples of this phase.
2. The population immiseration phase. The population continues to grow in number while the economy slows down. This happens because over the long term the rate of return on capital is typically greater than the growth rate of population salaries [11], as result the elites gets richer and the population gets poorer. Moreover, demography has a strong impact on the wealth of the population: the more workers of the same type are available, the less likely their wages are to grow. The state's ability to extract resources from the population reaches its limits in this phase. This

*CLiC-it 2024: Tenth Italian Conference on Computational Linguistics, Dec 04 — 06, 2024, Pisa, Italy*
*Corresponding author.
fabio.celli@maggioli.it (F. Celli); valerio.basile@unito.it (V. Basile)
https://github.com/facells/fabio-celli-publications (F. Celli); https://www.unito.it/persone/vabasile (V. Basile)
0000-0002-7309-5886 (F. Celli); 0000-0001-8110-6832 (V. Basile)

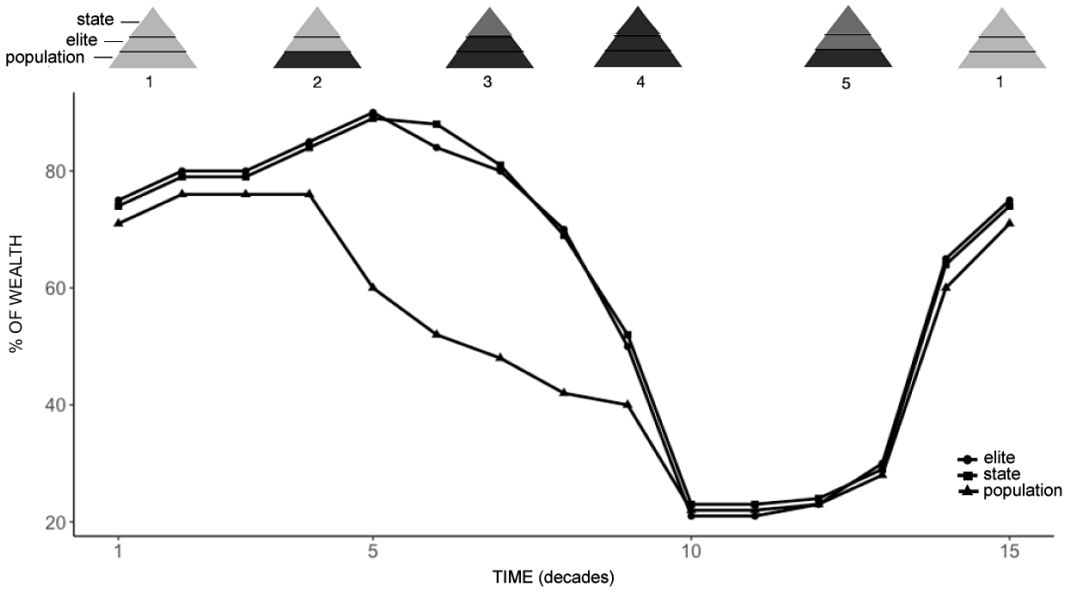


**Figure 1:** Time chart depicting the dynamics and phases described by the Structural-Demographic Theory.

can lead to increasing inequality, and social unrest begins. United States in the 1890s and 1970s are an example of this phase.

3. The elite overproduction phase. The population tries to access the elite ranks but overloads the social lift mechanisms and yields a reduced capability of the elite to solve problems in the society, which raise the probability to have societal instability. USSR in the 1950s and US in the 1990s are examples of this phase.
4. The state stress phase. The state's ability to govern the population and foster cooperation between population and elites begins to decline, and the elites become increasingly fragmented. This can lead to widespread violence and civil war. Moreover, the state tends to be in financial distress as a consequence of slowed economy and internal fragmentation, thus any triggering event that the state cannot manage can break into a crisis. Germany in the 1920s is an example.
5. The crisis, collapse or recovery phase. The state is either reformed by the elites who find an agreement or overthrown by internal or external forces. At the end of this phase a new social equilibrium is found and a new period of stability begins, restarting the cycle. Examples are France in the 1790s, UK in the 1940s, US in the 1860s under civil war and also in 1930s under New Deal reforms.

The dynamics described by the SDT are represented in figure 1 [12]. SDT has been used to explain a wide range of historical events, including the French Revolution, the American Civil War [13], the fall of the Qing Dynasty [14], the Russian Revolution and the instability in the US in recent years.

In this paper we propose a novel multi-class classification task: given a text describing the historical events of a decade, find the appropriate SDT phase label. To do so we exploited historical data from the Seshat project, produced textual descriptions for decades in the history of human societies and annotated each decade with SDT phases following specific annotation guidelines. We computed inter-annotator agreement between 3 annotators and experimented with LLMs in classification. The paper is structured as follows: in Section 2 we will describe the data, the guidelines for the annotation (Section 3), the classification experiments in Section 4, the conclusion and direction for future work in Section 5.

## 2. Data

It is not easy to design a dataset for historical data. There are specific datasets for event detection from text [15], for paleoclimatology [16], for census analysis through time [17] and for information extraction from historical documents [18], but there are few long-term historical datasets for Structural-Demographic analysis. Crucially the Seshat project [19] produced a dataset that contains machine-readable historical information about global history. The basic concept of Seshat is to provide quanti-


**Figure 2:** Distribution of the sampling zones. There are two sampling zone per World region: North America (US, Mexico), Oceania (Hawaii, Madang - Papua New Guinea), South America (Ecuador, Peru), Europe (France, Italy), Africa (Egypt, Ghana), Middle East (Levant, Iraq), Eurasia (Turkey, Siberia), South Asia (Uttar Pradesh - India, Java - Indonesia), East Asia (Henan - China, Japan)

tative and structured or semi-structured data about the evolution of societies, defined as political units (polities) from 35 sampling points across the globe in a time window from roughly 10000 BC to 1900 CE, sampled with a time-step of 100 years. A sampling frequency of 100 years is too much coarse-grained, not suitable to track the internal phases of the secular cycle, thus we resampled the data with a sampling frequency to 10 years, manually integrating data and descriptions from Seshat and from Wikipedia. To do so, we followed these general guidelines:

- For each polity in Sesaht create a number of rows to represent each decade. There must be no gaps between decades. If needed, add polities to fill the gaps searching in Wikipedia.
- Read the description of the polity provided in Seshat, identify dates and map the content to the corresponding decade.
- Search Wikipedia to find more information about the polity that can be mapped into decades. Fill in as much decades as possible. When dates are uncertain within a specific time period, use the median decade of that period.
- Summarize the content to fit about 400 characters. Focus on the following types of events: wars or battles; reforms; rulers; population; elites; disasters or epidemics; alliances or treaties; socioconomic context; famines or financial stress; protests or movements; changes of elite; religions and philosophies When possible, report the references about the information found.

We also extended the data to include the polities until the 2010s CE. In order to limit the long and time-consuming manual data wrangling, we reduced the number of sampling zones from 35 to 18 but at the same time we kept the original variety of world regions [20]. This, combined with the extension of the time window, allowed us to obtain 366 polities (roughly the same number of polities as Seshat) and 3540 rows with a textual description. We will call "Chronos" the dataset we produced. It contains the following features:

- *timestamp* of each decade,
- the *Age* indicating the periods of history (prehistoric, ancient, medieval, early-modern, modern, post-modern),
- the *sampling zone* as reported in Figure 2,
- the *world regions* related to the sampling zones,
- a *Polity ID* formatted with a standard method: 2 letters to indicate the area of origin of the culture, 3 letters to indicate the name of the polity, 1 letter to indicate the type of society (c=culture/community; n=nomads; e=empire; k=kingdom; r=republic) and 1 letter to indicate the periodization (t=terminal; l=late; m=middle; e=early; f=formative; i=initial; *=any). For example "EsSpael" is the late Spanish Empire, "ItRomre" is the early Roman Republic and "CnWwsk*" is

the period of the Warring States under the Wei Chinese dynasty,
- a *short textual description* of the decade in Italian and English.

Short texts can contain one or more events and references. Consider the following examples extracted from the Chronos dataset:

1. *introduction of iron from Vietnam by 300 BC [Bellwood P. 1997. Prehistory of the Indo-Malaysian Archipelago: Revised Edition pp. 268-307]. Old Malay as lingua franca.*
2. *Siege of Constantinople in 626. The Byzantines won. Problems in the succession to the throne: Kavadh II is killed in 628. Years of war with Bizantines had exhausted the Sasanids who were further weakened by economic decline; religious unrest and increasing power of the provincial landholders. King Yazdegerd III (r. 632-651) could not stand against the Islamic conquest of Persia.*

Example 1 contains a socio-economic context about the Buni culture of Indonesia and example 2 contains events about war, rulers, socio-economic context, religion and elite change about the late Sasanian Empire. The events in the short textual description are specific to the SDT and help annotators in their decisions about the historical phase labels. For example a good socio-economic context may be a clue of a growth phase and a disaster may trigger a crisis phase. For this reason we did not exploit the labels proposed in literature, such as second-level HTOED categories or the HISTO classes [21]. However, we acknowledge that this is an aspect that requires further research. All events included in the texts were manually detected, and the data collectors were trained to recognize key events from the examples provided in the literature about SDT [12].

## 3. Annotation and Evaluation

The main problem with the annotation of phases of historical cycles is its interpretability. While everyone agrees the 1789-1799 period in France was a time of crisis, reaching a consensus on the impact of the 1860s French intervention in Mexico proves more difficult. Did it trigger a phase of impoverishment or of elite overproduction? Moreover, did the rise of Mao Zedong as leader of China in the 1950s began a phase of growth or continued the previous crisis?

We defined the following guidelines for the annotation:

1. Read the textual description to identify key events: wars, reforms, rulers, population, elites, disasters, epidemics, alliances or treaties, socio-economic context, famines or financial stress, protests or movements, religions.
2. Use polity identifiers to find the start and end points of cultures. The end of a culture represents a crisis period.
3. Starting from the beginning of a culture, initially assign the sequence of labels of a standard secular cycle model: 1,1,2,2,3,3,4,4,4,5 and then evaluate whether to keep or change the labels in each decade. It is possible to have longer or shorter cycles. There can be only one label 5 (crisis) per cycle. A polity can have one or more cycles.
4. Having in mind the key events in the textual description, select one of the following labels to describe the decade: 1=growth. A society is generally poor when it experiences renewal or change followed by demographic (but not always territorial or economic) growth. Reforms, alliances, wars won or similar events are potential indicators of this phase. 2=impoverishment of the population. Potential economic and/or territorial expansion slows while demography continues to expand. The elite takes much of the wealth and defines the status symbols. Stability and external attacks are potential indicators of this phase. 3=Overproduction of the elites. The wealthy seek to translate their wealth into positions of authority and prestige. The population becomes poor. Movements, protests, and wars are potential indicators of this phase. 4=State stress. The elites want to institutionalize their advantages in the form of low taxes and privileges that lead the state into fiscal difficulties. Wars, protests and changes in the elite are potential indicators of this phase. 5=Crisis. a triggering event such as a war, revolt, famine or disaster that the state is unable to manage leads to a new configuration of society. Emigration of elites, subjugation to other societies, civil wars or profound reforms are potential indicators of this phase.
5. Use the progressive order of the phases if no textual description is available for the decade.
6. Make sure there is a progressive order of the labels (e.g. phase 3 must follow phase 2). All labels can be repeated in the following decade except the crisis phase, which conventionally lasts one decade.

| Trial | Examples | Raters | Labels | K |
|---|---|---|---|---|
| base | 93 | 3 | 5 | 0.206 |
| trained | 93 | 3 | 5 | 0.455 |

**Table 1**
Inter-Annotator Agreement (Fleiss' Kappa) on the annotation of secular cycle phases.

A single annotator annotated the entire corpus, then

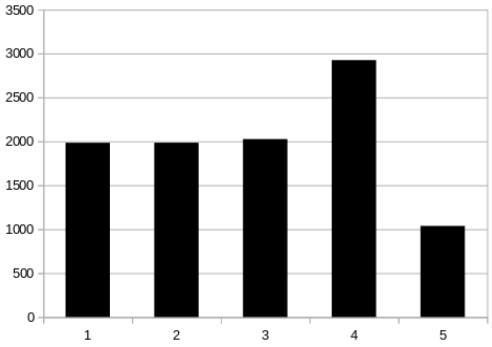


**Figure 3:** Distribution of the labels in the Chronos dataset.

we evaluated the annotation with two different trials involving students, not expert in history. We compared a subset of data annotated by two students to the same subset annotated by the principal annotator. The first trial was done just following the guidelines after a general explanation of the SDT. The second trial was done, with different students, following the guidelines after a training session, where the annotation was discussed and agreed upon. Results, reported in Table 1, show that with a training session the agreement rises considerably (from slight to moderate). The base agreement level is comparable to the one observed in the annotation of hate speech among 5 trained judges on a non-binary scheme, which obtained a Fleiss K=0.19 [22] [23]. The distribution of the labels in the Chronos dataset is depicted in Figure 3. In the standard secular cycle model, the stress phase (label 4) is the most common, followed by the crisis phase (label 5), which is the least common. The other three phases (labels 1, 2, and 3) occur with roughly equal frequency in the data.

## 4. Classification and Discussion

In order to test the robustness of the Chronos dataset, we performed cross-validation classification experiments. The setting is straightforward: each line of the dataset is considered independently from one another, and we apply a supervised classification model to predict the human-annotated label, i.e., the phase (from 1 to 5).

In this experiments, we ignored lines for which no textual description is available and we used the chance baseline of $F1 = 0.2$. As learning model, we fine-tuned RoBERTa large[1] [24] for the English textual descriptions and Italian BERT XXL[2] for the Italian texts. We used a learning rate of $10^-6$ and applied early stopping and model checkpointing, validating each fold on 10% of the training set.

We performed 5-fold cross validation and measured the precision, recall, and F1 score of the predicted labels compared against the gold standard. Table 2 shows the results of the experiments.

| English | | | |
|---|---|---|---|
| Phase | Precision | Recall | F1-score |
| 1 | 0.542 | 0.486 | **0.513** |
| 2 | 0.338 | 0.256 | **0.291** |
| 3 | 0.242 | 0.048 | 0.080 |
| 4 | 0.319 | 0.601 | **0.416** |
| 5 | 0.330 | 0.364 | **0.346** |
| Italian | | | |
| Phase | Precision | Recall | F1-score |
| 1 | 0.489 | 0.510 | **0.499** |
| 2 | 0.321 | 0.211 | **0.254** |
| 3 | 0.191 | 0.044 | 0.071 |
| 4 | 0.290 | 0.660 | **0.403** |
| 5 | 0.397 | 0.186 | **0.254** |

**Table 2**
Results of 5-fold multiclass classification experiments. Results above the baseline (0.2) are marked in bold.

The classification performance shows that the textual descriptions in our dataset are sufficient to predict the corresponding phase to a certain extent, however in quite an imbalanced way. In particular, the classification of phases 1 and 4 achieves moderately good results, while phase 3 in particular is almost never predicted, despite the rather balanced distribution of labels in the dataset.

**English**

| Gold labels \ Predicted labels | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0.486 | 0.106 | 0.021 | 0.228 | 0.159 |
| 2 | 0.219 | 0.256 | 0.054 | 0.346 | 0.124 |
| 3 | 0.164 | 0.170 | 0.048 | 0.478 | 0.141 |
| 4 | 0.082 | 0.100 | 0.041 | 0.601 | 0.175 |
| 5 | 0.141 | 0.075 | 0.006 | 0.414 | 0.364 |

**Italian**

| Gold labels \ Predicted labels | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 0.186 | 0.201 | 0.062 | 0.008 | 0.542 |
| 2 | 0.074 | 0.510 | 0.110 | 0.032 | 0.274 |
| 3 | 0.041 | 0.248 | 0.211 | 0.055 | 0.446 |
| 4 | 0.046 | 0.190 | 0.139 | 0.043 | 0.582 |
| 5 | 0.063 | 0.142 | 0.081 | 0.054 | 0.660 |

**Figure 4:** Confusion matrices of the classification of English (above) and Italian (below) decade descriptions.

The confusion matrices in Figure 4 further highlight

[1] https://huggingface.co/FacebookAI/roberta-large
[2] https://huggingface.co/dbmdz/bert-base-italian-xxl-cased

interesting trends. While the biases of the models in terms of phases are clear, it is worth noticing that misclassification happens often between contiguous phases.

```
Structural Demographic Theory predicts outbreaks of political instability in complex societies, based on three actors: the population, the elite, and the state. Each decade is associated with one of five phases:

1. The 'growth' phase, when a fresh and effective culture creates social cohesion, the economy is growing rapidly and the state is expanding its control over the population;

2. The 'population immiseration' phase, when the population continues to grow while the economy slows;

3. The 'elite overproduction' phase, when the population tries to access the elite ranks but overloads the social lift mechanisms and yields a reduced capability of the elite to solve problems in the society;

4. The 'state stress' phase, when the state's ability to govern the population and foster cooperation between population and elites begins to decline, and the elites become increasingly fragmented;

5. The 'crisis, collapse or recovery' phase, when the state is either reformed by the elites or overthrown by internal or external forces;

Act as a highly intelligent historian chatbot. You will be given the description of a decade and you are asked to predict the phase number. Please output only a number from 1 to 5.

Decade: textual description

Phase:
```

**Figure 5:** Prompt for zero-shot classification experiments with LlaMa70B.

This suggests that a more refined, regression-based learning setting could be more favorable to this kind of data. Finally, we performed a pilot experiment with a large language model, namely LlaMa 3 70B[3], prompting the model to elicit zero-shot classifications of the phases given the textual descriptions in English. The prompt we used for the model is shown in Figure 5. No particular decoding strategy was applied for this experiment.

Despite the dimension of this model, the classification performance was poor, 5–10 F1 points below the supervised classification results at the best try. Interestingly, the zero-shot classification exhibited a similar pattern in terms of individual labels, with the model strongly biased towards phase 1 and 4, and unable to properly predict phases 2 and 3.

We suggest that, while phases 1 and 4 have similar types of events in most societies (i.e. reforms or won wars in phase 1, famines or financial problems in phase 4) there is much more variability for phases 2, 3 and 5. It must be noted that these experiments only scratches the surface of the learning capabilities of the Chronos dataset. In particular, in this setting, the temporal interdependence of the decades is not considered, and specific algorithms should be applied in the future to capture this temporal structure.

## 5. Conclusion and Future

We introduced a new classification task named historical phase recognition. We believe that, once we improve their performance, classification algorithms trained for this task will allow us to automatically annotate many more polities with secular cycles with a potential disruptive improvement in the study of comparative history. We believe that inter-annotator agreement can be further improved by having domain experts annotate the data. Additionally, the automatic extraction of events from short historical texts, or the definition of guidelines for their annotation, can be a valuable tool both in the annotation and classification tasks. By combining these two approaches, we can improve the dataset and make it more reliable.

For the future we plan to improve the performance of classification by including the temporal interdependence factors, and to improve the inter annotator agreement, also calculating the agreement between labels generated by models and by humans. In the future it would be interesting to add event structure annotations such as TimeML in Chronos. The poor performance in zero-shot classification using an LLM is likely a function of the sophisticated reasoning and world knowledge required to perform the task. The LLM could benefit from more advanced prompting strategies (e.g. few-shot or chain-of-thoughts) or even supervision in the form of fine-tuning.

The Chronos dataset is accessible online in viewer/-commenter mode[4]. Edit and download access is available under request.

[3] https://huggingface.co/meta-llama/Meta-Llama-3-70B

[4] https://docs.google.com/spreadsheets/d/1OW6CtmUudN3WTJ1VvWRZYZdTWVEjDJGns6Q8_I6EBwk/edit?usp=sharing

## Acknowledgments

This work was supported by the European Commission grant 101120657: European Lighthouse to Manifest Trustworthy and Green AI - ENFIELD.

## References

[1] F. Celli, E. Stepanov, M. Poesio, G. Riccardi, Predicting brexit: Classifying agreement is better than sentiment and pollsters, in: Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES), 2016, pp. 110–118.

[2] M. Lai, F. Celli, A. Ramponi, S. Tonelli, C. Bosco, V. Patti, Haspeede3 at evalita 2023: Overview of the political and religious hate speech detection task, in: M. Lai, S. Menini, M. Polignano, V. Russo, R. Sprugnoli, G. Venturi (Eds.), Proceedings of the Eighth Evaluation Campaign of Natural Language Processing and Speech Tools for Italian. Final Workshop (EVALITA 2023), Parma, Italy, September 7th-8th, 2023, volume 3473 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023.

[3] D. Nozza, F. Bianchi, G. Attanasio, Hate-ita: Hate speech detection in italian social media text, in: Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH), 2022, pp. 252–260.

[4] E. W. Pamungkas, A. T. Cignarella, V. Basile, V. Patti, et al., Automatic identification of misogyny in english and italian tweets at evalita 2018 with a multilingual hate lexicon, in: CEUR Workshop Proceedings, 1, CEUR-WS, 2018, pp. 1–6.

[5] S. S. Tekiroglu, Y.-L. Chung, M. Guerini, Generating counter narratives against online hate speech: Data and strategies, arXiv preprint arXiv:2004.04216 (2020).

[6] R. Dalio, Principles for dealing with the changing world order: Why nations succeed or fail, Simon and Schuster, 2021.

[7] J. A. Goldstone, Demographic structural theory: 25 years on, Cliodynamics 8 (2017).

[8] A. V. Korotaev, Introduction to social macrodynamics: Secular cycles and millennial trends in Africa, Editorial URSS, 2006.

[9] P. Turchin, S. A. Nefedov, Secular cycles, in: Secular Cycles, Princeton University Press, 2009.

[10] P. Turchin, A. Korotayev, The 2010 structural-demographic forecast for the 2010–2020 decade: A retrospective assessment, PloS one 15 (2020).

[11] T. Piketty, Capital in the twenty-first century, Harvard University Press, 2014.

[12] D. Hoyer, J. S. Bennett, H. Whitehouse, P. François, K. Feeney, J. Levine, J. Reddish, D. Davis, P. Turchin, Flattening the curve: Learning the lessons of world history to mitigate societal crises, osf.io (2022).

[13] P. Turchin, A Structural-Demographic Analysis of American History, Beresta Books Chaplin, 2016.

[14] G. Orlandi, D. Hoyer, H. Zhao, J. S. Bennett, M. Benam, K. Kohn, P. Turchin, Structural-demographic analysis of the qing dynasty (1644–1912) collapse in china, Plos one 18 (2023) e0289748.

[15] R. Sprugnoli, S. Tonelli, One, no one and one hundred thousand events: Defining and processing events in an inter-disciplinary perspective, Natural language engineering 23 (2017) 485–506.

[16] B. J. Van Bavel, D. R. Curtis, M. J. Hannaford, M. Moatsos, J. Roosen, T. Soens, Climate and society in long-term perspective: Opportunities and pitfalls in the use of historical datasets, Wiley Interdisciplinary Reviews: Climate Change 10 (2019) e611.

[17] R. Abramitzky, L. Boustan, K. Eriksson, J. Feigenbaum, S. Pérez, Automated linking of historical data, Journal of Economic Literature 59 (2021) 865–918.

[18] F. Boschetti, C. Andrea, D. Felice, G. Lebani, P. Lucia, P. Paolo, V. Giulia, M. Simonetta, et al., Computational analysis of historical documents: An application to italian war bulletins in world war i and ii, in: Proceedings of the LREC 2014 Workshop on Language resources and technologies for processing and linking historical documents and archives (LRT4HDA 2014), ELRA, 2014.

[19] P. Turchin, H. Whitehouse, P. François, D. Hoyer, A. Alves, J. Baines, D. Baker, M. Bartokiak, J. Bates, J. Bennet, et al., An introduction to seshat: Global history databank, Journal of Cognitive Historiography 5 (2020) 115–123.

[20] F. Celli, Feature Engineering for Quantitative Analysis of Cultural Evolution, Technical Report, Center for Open Science, 2022.

[21] R. Sprugnoli, S. Tonelli, Novel event detection and classification for historical texts, Computational Linguistics 45 (2019) 229–265.

[22] F. Del Vigna, A. Cimino, F. Dell'Orletta, M. Petrocchi, M. Tesconi, Hate me, hate me not: Hate speech detection on facebook, in: Proceedings of the first Italian conference on cybersecurity (ITASEC17), 2017, pp. 86–95.

[23] F. Poletto, V. Basile, M. Sanguinetti, C. Bosco, V. Patti, Resources and benchmark corpora for hate speech detection: a systematic review, Language Resources and Evaluation 55 (2021) 477–523.

[24] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized BERT pretraining approach, CoRR abs/1907.11692 (2019). URL: http://arxiv.org/abs/1907.11692. arXiv:1907.11692.