Event Extraction from Historical Texts: A New Dataset for Black Rebellions

Viet Dac Lai[1], **Minh Van Nguyen**[1], **Heidi Kaufman**[2], **Thien Huu Nguyen**[1]
[1]Department of Computer and Information Science
[2]Department of English
University of Oregon, Eugene, OR 97403, USA
{vietl,minhnv,thien}@cs.uoregon.edu
hkaufman@uoregon.edu

Abstract

Understanding historical events is necessary for the study of contemporary society, culture, and politics. In this work, we focus on the event extraction task (EE) to detect event trigger words and their arguments in a novel domain of historical texts. In particular, we introduce a new EE dataset for a corpus of nineteenth-century African American newspapers. Our goal is to study the discourse of slave and non-slave African diaspora rebellions published in the periodical press in this period. Our dataset features 5 entity types, 12 event types, and 6 argument roles that concern slavery and black movements between the eighteenth and nineteenth centuries. Historical newspapers present many challenges for existing EE systems, including the evolution of meanings of words and the extensive use of religious discourse in newspapers from this era. Our experiments with current state-of-the-art EE systems and BERT models demonstrate their poor performance over historical texts and call for more robust research efforts in this area.

1 Introduction

In the last two decades, the emergence of digital humanities has transformed scholarship in the humanities. Historical documents are now massively digitized into photos and texts that allow researchers to query across collections and languages (Piotrowski, 2012). Despite the convenience of these applications (Yang and Eisenstein, 2016), a gap still exists between datasets and research methods. As such, humanities scholars do not solely interpret historical facts from statistical figures derived from massive data. Rather, they prefer reading texts and interpreting words in historical and cultural context, or by associating texts with the circumstances surrounding their publication. This working methodology requires an emphasis on the quality of the data over the quantity of the data. Recent advances of natural language processing (NLP) aim to bridge the gap between qualitative and quantitative analyses by identifying, extracting, and counting contextual data (Won et al., 2018; Wadden et al., 2019; Lin et al., 2020). This new approach provides contextual information about real-life entities (e.g., individuals, locations, times, documents) which can be later integrated into knowledge bases (Won et al., 2018) to aid historical research and discourse analysis.

In this work, we explore Information Extraction (IE) in NLP for humanities research in support of the important and complicated process of **knowledge extraction** from historical texts. Particularly, we investigate the Event Extraction (EE) task which identifies event trigger words of pre-determined event types (the most important words/phrases to evoke events) (Li et al., 2013), together with its arguments (e.g., participants, locations). For example, in the following sentence an EE system should be able to detect the word "*proclaimed*" as a trigger word of the event type "*Law_Approve*" and associate it with the arguments, i.e., agent (*Capitol*), beneficiary (*the slave*), and datetime (*now*).

Freedom to the slave should now be ***proclaimed*** *from the Capitol, and should be seen above the smoke and fire of every battle field.*

To enable the development and evaluation of EE models for historical text, benchmark datasets play an important role. However, most of the current datasets in EE (i.e., ACE-2005 (Walker et al., 2005) and TAC KBP (Mitamura et al., 2015)) are not suitable for this domain for several reasons. First, these datasets are collected from various sources without a target topic (Walker et al., 2005; Mitamura et al., 2015). Therefore, tracking the evolution of some specific movements or progress, which is of great interest to literary scholars and historians, is not a feasible goal. Second, documents in these datasets

Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021, pages 2390–2400
August 1–6, 2021.

are derived from recent articles and documents in which the use of words in the text differ from their uses in the past. For example, some words obtain new semantics over time, and the dominance of religion in the past led to extensive use of religion-related words and figurative language in historical publications. Last but not least, existing EE datasets mostly concern events in common human life, such as giving birth, transportation, and crimes. These events might not relate to the subjects literary scholars and historians want to study.

To redress this problem, we introduce a novel EE dataset for historical texts, called BRAD, focusing on Black Rebellions in African Diaspora (i.e., African American population). BRAD's documents are selected by a humanities expert and are annotated by EE experts for 5 entity types, 12 event types, and 6 argument roles. Finally, we evaluate the state-of-the-art EE models on BRAD. Our experiments show that the performance for historical texts of current EE models is significantly poorer than those for modern texts, necessitating further research into this area. We will also release our dataset and code to facilitate future research.

2 Data Collection and Annotation

In this project we use documents from the African American newspaper corpus. These documents involve news articles derived from nineteenth-century African American periodicals[1] published from 1827 to 1909.

To create an EE dataset, we first designed a set of event types and annotation guidelines, consulting our humanities expert who specializes in nineteenth-century literature. In particular, we focus on the four most important events for Black rebellions presented in our corpus, including **Humanity**: a humanity event concerns a violation or facilitation of basic human rights (e.g. living, freedom, property); **Law**: a law event characterizes an introduction, approval or appeal of a law; **Conflict**: a conflict event represents an act of violence; it includes the initialization, development, and consequences of a violent act; and **Justice**: a justice event captures an act of punishment of the government to the people who violate a law. These four events are further expanded into 12 event sub-types. Tables 7 and 8 present event types along with their descriptions and examples in BRAD. To capture arguments for such events, we introduce five entity types (i.e., Person, Organization, Geographical-Political Entities, Time, and Document). The first four entity types follow the definition in the ACE 2005 guideline (Walker et al., 2005) while the Document type represents government documents (e.g., Slavery Act) used in events. Finally, we define six argument roles that such entity types can play in our events, including Time, Location, Agent, Patient, Object, and Beneficiary. Tables 9 and 10 provide more descriptions and examples of these argument roles for each event type.

The African American corpus is a large corpus of 177,582 articles. We thus select documents that are relevant to our focused topic of Black diaspora rebellions. First, automatic selection is done by keyword matching to identify documents related to slavery and insurrection. As such, our humanities expert defined a set of keywords for the topic of rebellion. In the nineteenth century this cluster of words were used interchangeably to describe African diaspora rebellion events (e.g., "*rebel*", "*revolt*", "*strike*", "*insurrection*"). We used the Stanford CoreNLP toolkit to split and tokenize documents into sentences and words. Next, for each document in the corpus, we counted the number of words in the document that appears in the designated keyword set (called matching rate). The top 1000 documents with the highest matching rates are selected for further consideration. In the second step, the humanities expert examined the 1000 documents to identify relevant documents for Black rebellions, leading to the selection of 151 documents used for the EE annotation.

In the next step we recruited two graduate students to annotate the selected documents for EE. Each student was independently trained on the annotation guideline and performed a group of exercises to better recognize events and entities. The students annotated the 151 documents for entity mentions and event triggers, achieving Cohen's Kappa scores of 0.81 and 0.82 respectively. Note that these scores are very close to the near-perfect agreement range of $[0.81, 0.99]$. To further improve the quality of the dataset, our humanities expert will resolve the annotation conflicts that arise between the two students, leading to the final annotation version of entity mentions and event triggers in the 151 documents. In the next step, given the reconciled entity mention and event trigger annotation, the two students continue to annotate event

[1] Douglass Monthly, The Frederick Douglass Paper, Freedom Journals, The Christian Recorder, The Colored American.

arguments for the event triggers. Our evaluation shows a Cohen's Kappa score of 0.75 that indicates a strong agreement between the two annotators. Also, the lower agreement score for event arguments suggests that event argument annotation is more ambiguous than those for entity mentions and event triggers. Finally, our domain expert was consulted to resolve any conflicts in event argument annotation, producing the final version of our BRAD dataset with the 151 documents. To facilitate the development of EE models, we then split BRAD into three portions for training, development, and test data with 101, 25, and 25 documents, respectively. Table 1 presents the statistics while Table 2 and 3 presents the frequencies of event and entity types in our BRAD dataset.

	Train	Dev	Test	Total
#document	101	25	25	151
#sentence	3,847	925	866	5638
#token	117,278	27,860	26,920	172,058
#event trigger	2,720	606	933	4,259
#entity mention	14,389	3,287	3,749	21,425
#event argument	6,057	1,219	2,570	9,846

Table 1: Statistics of the BRAD dataset.

Event Type	#Event
Conflict_Attack	1,628
Conflict_Other	971
Humanity_Deprive	577
Humanity_Endow	376
Conflict_Injure	153
Law_Approve	145
Law_Repeal	141
Law_Propose	108
Justice_Arrest-Jail	78
Conflict_Protest	42
Justice_Execute	26
Justice_Sentence	14

Table 2: Distribution of event types in BRAD.

Annotation Challenges: During the EE annotation process of historical texts, we found several noteworthy challenges regarding the ability to achieve interpretive consensus of the texts.

First, for the domain expertise, we find that the use and meaning of words evolves over time and across geographical regions, potentially introducing new meanings or making one meaning more popular than the others. Language is always in perpetual flux. As such, understanding texts from the past requires analysis of the context in which texts were written. In order to be effective, the annotations must be attentive to these contexts. For example, in the following sentence, "*Congress*" and "*her*" are two mentions of the *USS Congress* battleship launched by the United State Navy in 1841. Without historical knowledge, our current perception might interpret "*Congress*" as the legislative branch of the United States. In fact, the second clause mentions the wooden hull that helps to clarify it as the battleship that sunk in 1862 during the US Civil War. Such misinterpretation might lead to incorrect annotations and analyses.

Entity Type	#Entity
PERSON (PER)	12,599
ORGANIZATION (ORG)	3,836
GEOPOLITICAL (GPE)	2,873
DOCUMENT (DOC)	1,121
DATETIME (TIME)	996

Table 3: Distribution of entity types in BRAD.

"*The **Congress** was visited and received the shots and shells in all part of **her** wooden hull*".

Second, we find that annotation disagreements are more likely to occur in the interpretation of event triggers. In BRAD, we allow event triggers to involve multiple words that cause span mismatches between annotations for some confusing cases (e.g., annotating the whole phrase "*make the black man equal*" as an event trigger or annotating "*make*" and "*equal*" as two separate triggers). Another form of popular disagreement involves mismatches on event types. Consider the following sentence as an example:

"*Believing his life to be in danger, Patmon stepped back, drew his revolver, and told the fellow to surrender, or he would **shoot** him* ."

Two annotators agree that the word "*shoot*" is an event trigger. However, one annotator considers this as an event of type *Conflict_Attack* as it is a part of the conflict between the overseer ("*Patmon*") and the slave ("*fellow*", "*him*"); the other annotator, on the other hand, treats "*shoot*" as a *Humanity_Deprive* event as the overseer is threatening to kill the slave (i.e., taking the right to life).

Data Analysis: To illustrate the ambiguity in BRAD, Table 4 shows five words with the highest frequency as event triggers (i.e., Event Count), along with the percentage of times these words are labeled as event triggers in the dataset (i.e., Event Rate) (Sims et al., 2019). This table demonstrates the likelihood that words with the highest event counts might not be annotated as event triggers in BRAD, thereby necessitating EE models to find a

Word	#Event	Event Rate
war	183	23.5%
rebellion	112	47.3%
insurrection	66	78.6%
revolt	77	60.2%
emancipation	58	82.9%
take	64	32.3%
put	35	47.3%
send	25	45.5%

Table 4: Event rates of the words with the highest event counts in BRAD.

	ACE 2005	BRAD
god	4.5%	17.5%
lord	0.3%	8.7%
heaven	0.7%	8.7%
mighty	0.2%	8.7%
sacred	0.2%	9.5%
curse	0.5%	5.6%
christian	1.7%	18.3%

Table 5: Percentages of documents containing religion-related words in BRAD and ACE 2005.

method of effectively capturing context in order to perform correct predictions.

Moreover, we find extensive use of religion-related words in BRAD compared to existing EE datasets. For example, considering the words "*lord*", "*heaven*", and "*christian*", the percentages of documents in ACE 2005 containing these words are only 0.3%, 0.7%, and 1.7% while those percentages for BRAD are 8.7%, 8.7%, and 18.3% respectively. Such language difference suggests the potential need to adapt existing language models to better capture the nature of historical texts which, in turn, will facilitate a more accurate performance of EE.

3 Experiment

There are three major EE tasks that BRAD supports for historical texts, including entity mention detection (EMD), event trigger detection (ED), and event argument extraction (EAE). This section aims to reveal the complexity of the EE tasks in BRAD by evaluating the performance of existing state-of-the-art models for EE on this dataset. In particular, we focus on the following state-of-the-art models for EE that leverage the pre-trained language model BERT (Devlin et al., 2019) for the text encoding and jointly perform predictions for all EE tasks in an end-to-end fashion (i.e., joint inference):

DyGIE++ (Wadden et al., 2019): This model utilizes dynamic span graphs to exploit long-range cross-sentence relationships for span representation propagation for joint IE.

OneIE (Lin et al., 2020): This model first identifies spans of entity mentions and event triggers. The detected spans are then paired to jointly predict entity types, event types, relations, and argument roles for IE. Global features are used to capture cross-task and cross-instance dependencies and are employed in the decoding phase with beam searches to improve extraction performance.

As such, we adapt the official implementations of such models from their original papers for our EE task in BRAD by ignoring the relation extraction task and re-tuning them on the BRAD development set. For both models, we employ the pre-trained BERT model (i.e., the bert-base-cased version) to encode input texts. Besides, motivated by the language difference between historical and modern texts, we further explore a variant of the BERT model by fine-tuning it on the African American corpus via the masked language modeling task (Devlin et al., 2019). Note that we exclude the 151 documents of BRAD in this fine-tuning process. This fine-tuned BERT model will also be fed into DyGIE++ and OneIE to perform EE in BRAD.

Result: Table 6 reports the performance of the models on the test set of BRAD over five subtasks: Entity Mention Detection (Entity), Event Trigger Identification, i.e., not concerning event types (Trig-I), Event Trigger Classification (Trig-C), Event Argument Identification, i.e., not concerning argument roles (Arg-I), and Event Argument Classification (Arg-C). For comparison, we also include the original performance of the models on the popular EE dataset ACE 2005. There are three major observations from the table. First, the performance of current EE models on BRAD is significantly and substantially worse than those on ACE across different tasks. It thus suggests that EE for historical texts in BRAD is a challenging task and more research effort is necessary to boost the EE performance for this domain. Second, comparing the performance of the models with different versions of BERT (i.e., original vs fine-tuned), it is clear that fine-tuning BERT on historical texts is beneficial for improving the performance of EE models on BRAD (especially for OneIE where the improvement is consistent across different EE subtasks with large margins). This observation suggests that pre-training BERT on modern texts is unable to capture the nuance of language use in

Task	Model	ACE	BRAD			BRAD$^{\#}$		
		F1	P	R	F1	P	R	F1
Entity	DyGIE++	90.7	85.6	75.9	80.5	84.3	78.6	81.4
	OneIE	90.3	85.4	77.0	81.0	85.0	79.4	82.1
Trig-I	DyGIE++	76.5	81.5	50.1	62.0	77.4	56.9	65.6
	OneIE	78.6	80.9	47.0	59.4	80.8	52.9	63.9
Trig-C	DyGIE++	73.6	62.7	38.5	47.7	61.6	40.6	49.0
	OneIE	75.2	63.5	36.9	46.7	64.9	42.2	51.2
Arg-I	DyGIE++	55.4	55.5	28.0	37.2	47.8	28.6	35.8
	OneIE	60.7	57.7	33.9	42.7	58.9	40.4	47.9
Arg-C	DyGIE++	52.5	48.8	24.6	32.7	42.5	25.4	31.8
	OneIE	58.6	49.6	29.1	36.7	52.1	31.6	39.4

Table 6: The performance of models on the test sets of BRAD and ACE 2005. The BRAD$^{\#}$ columns report the performance with BERT fine-tuned on the African American corpus.

history, thus impairing the models and requiring appropriate adaptation to boost the EE performance. Finally, we note that the human performance (F1 scores) for Entity, Trig-C, and Arg-C on BRAD are 95.43, 88.3, and 79.8 respectively. The large performance gaps between human and current EE systems thus presents many research opportunities for future work on BRAD.

4 Related work

Prior work in NLP for historical texts has mainly focused on spelling and text normalization (Pettersson et al., 2014; Bollmann et al., 2017; Flachs et al., 2019). Recently, some studies have undertaken research on historical texts with NLP tasks such as POS tagging (Yang and Eisenstein, 2016) and information extraction (Pettersson et al., 2016). However, none of this work has explored EE.

EE is an active research area due to the availability of EE datasets e.g., for general (Walker et al., 2005; Mitamura et al., 2015) and biomedical (Kim et al., 2011) domains. Most of prior studies focus on in-domain EE (Ahn, 2006; Li et al., 2013; Nguyen and Grishman, 2015; Chen et al., 2015; Nguyen et al., 2016; Yang et al., 2019; Wadden et al., 2019; Lai et al., 2020c; Nguyen et al., 2021). Some recent studies in EE have also addressed extensible learning settings for EE to new event types, e.g. zero-shot learning (Huang et al., 2018), few-shot learning (Lai et al., 2020a,b), or new domains (Naik and Rosé, 2020). The closet works to ours involve recent efforts to create new datasets for EE (Satyapanich et al., 2020; Ebner et al., 2020; Wang et al., 2020; Trong et al., 2020; Le and Nguyen, 2021). However, these works do not consider historical texts as we do.

5 Acknowledgement

This work is supported by an I3 grant from the University of Oregon Provost's Office (Incubating Interdisciplinary Initiatives) and the Army Research Office (ARO) grant W911NF-21-1-0112. This research is also based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No. 2019-19051600006 under the Better Extraction from Text Towards Enhanced Retrieval (BETTER) Program. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of ARO, ODNI, IARPA, the Department of Defense, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation therein. This document does not contain technology or technical data controlled under either the U.S. International Traffic in Arms Regulations or the U.S. Export Administration Regulations.

6 Conclusion

We present BRAD, a new dataset for EE on historical texts that focuses on Black rebellions in the American Africa corpus. Our experiments demonstrate the poor performance of current models for EE on BRAD compared to those on modern texts, thus creating room for future research on EE for historical texts. We also illustrate one approach to improve current EE systems for historical texts via fine-tuning existing pre-trained language models. In the future, we plan to enlarge our datasets with more annotated documents and event types.

References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*.

Marcel Bollmann, Joachim Bingel, and Anders Søgaard. 2017. Learning attention for historical text normalization by learning to pronounce. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Yubo Chen, Liheng Xu, Kang Liu, Daojian Zeng, and Jun Zhao. 2015. Event extraction via dynamic multi-pooling convolutional neural networks. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Seth Ebner, Patrick Xia, Ryan Culkin, Kyle Rawlins, and Benjamin Van Durme. 2020. Multi-sentence argument linking. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Simon Flachs, Marcel Bollmann, and Anders Søgaard. 2019. Historical text normalization with delayed rewards. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Lifu Huang, Heng Ji, Kyunghyun Cho, and Clare R Voss. 2018. Zero-shot transfer learning for event extraction. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Jin-Dong Kim, Tomoko Ohta, Sampo Pyysalo, Yoshinobu Kano, and Jun'ichi Tsujii. 2011. Extracting bio-molecular events from literature—the bionlp'09 shared task. *Computational Intelligence*, 27(4):513–540.

Viet Dac Lai, Franck Dernoncourt, and Thien Huu Nguyen. 2020a. Exploiting the matching information in the support set for few shot event classification. In *Proceedings of the 24th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*.

Viet Dac Lai, Thien Huu Nguyen, and Frank Dernoncourt. 2020b. Extensively matching for few-shot learning event detection. In *Proceedings of the First Joint Workshop on Narrative Understanding, Storylines, and Events*.

Viet Dac Lai, Tuan Ngo Nguyen, and Thien Huu Nguyen. 2020c. Event detection: Gate diversity and syntactic importance scores for graph convolution neural networks. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Duong Le and Thien Huu Nguyen. 2021. Fine-grained event trigger detection. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*.

Qi Li, Heng Ji, and Liang Huang. 2013. Joint event extraction via structured prediction with global features. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (ACL)*.

Ying Lin, Heng Ji, Fei Huang, and Lingfei Wu. 2020. A joint neural model for information extraction with global features. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Teruko Mitamura, Zhengzhong Liu, and Eduard H Hovy. 2015. Overview of tac kbp 2015 event nugget track. In *Proceedings of the Text Analysis Conference (TAC)*.

Aakanksha Naik and Carolyn Rosé. 2020. Towards open domain event trigger identification using adversarial domain adaptation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Minh Van Nguyen, Viet Lai, and Thien Huu Nguyen. 2021. Cross-task instance representation interactions and label dependencies for joint information extraction with graph convolutional networks. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Thien Huu Nguyen, Kyunghyun Cho, and Ralph Grishman. 2016. Joint event extraction via recurrent neural networks. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Thien Huu Nguyen and Ralph Grishman. 2015. Event detection and domain adaptation with convolutional neural networks. In *The 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*.

Eva Pettersson, Jonas Lindström, Benny Jacobsson, and Rosemarie Fiebranz. 2016. Histsearch-implementation and evaluation of a web-based tool for automatic information extraction from historical text. In *HistoInformatics@ DH*.

Eva Pettersson, Beáta Megyesi, and Joakim Nivre. 2014. A multilingual evaluation of three spelling normalisation methods for historical text. In *Proceedings of the 8th workshop on language technology for cultural heritage, social sciences, and humanities (LaTeCH)*.

Michael Piotrowski. 2012. Natural language processing for historical texts. *Synthesis lectures on human language technologies*, 5(2):1–157.

Taneeya Satyapanich, Francis Ferraro, and Tim Finin. 2020. CASIE: Extracting cybersecurity event information from text. In *Proceedings of the Association for the Advancement of Artificial Intelligence (AAAI)*.

Matthew Sims, Jong Ho Park, and David Bamman. 2019. Literary event detection. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Hieu Man Duc Trong, Duc Trong Le, Amir Pouran Ben Veyseh, Thuat Nguyen, and Thien Huu Nguyen. 2020. Introducing a new dataset for event detection in cybersecurity texts. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

David Wadden, Ulme Wennberg, Yi Luan, and Hannaneh Hajishirzi. 2019. Entity, relation, and event extraction with contextualized span representations. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*.

Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2005. ACE 2005 multilingual training corpus. In *Technical report, Linguistic Data Consortium*.

Xiaozhi Wang, Ziqi Wang, Xu Han, Wangyi Jiang, Rong Han, Zhiyuan Liu, Juanzi Li, Peng Li, Yankai Lin, and Jie Zhou. 2020. MAVEN: A Massive General Domain Event Detection Dataset. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Miguel Won, Patricia Murrieta-Flores, and Bruno Martins. 2018. Ensemble named entity recognition (NER): evaluating NER tools in the identification of place names in historical corpora. *Frontiers in Digital Humanities*, 5:2.

Sen Yang, Dawei Feng, Linbo Qiao, Zhigang Kan, and Dongsheng Li. 2019. Exploring pre-trained language models for event extraction and generation. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Yi Yang and Jacob Eisenstein. 2016. Part-of-speech tagging for historical English. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT)*.

Type	Description	Examples
LAW Propose	A PROPOSE event occurs when an actor (Agent) introduces a bill, proposition, or treaty which benefits a group of people (Beneficial).	Below we give the salient points of the bill of an entertainment recently **given** in the interest of a certain church about to be organized in a certain town in New Jersey. The bill **introduced** in Congress last week by the congressman from North Carolina , to abolish the 15th amendment. It 's only effect will be to create support for the bill of congressman Crumpacker which **proposes** a reduction of representation in those States
LAW Approve	An APPROVE event occurs when a bill or order (Object) is passed by either the head of the government or a representative committee (Agent).	Be it **enacted** by the General Assembly of Maryland. Vermont has **passed** her Liberty Bill , New York has under discussion , and Massachusetts will soon report and **pass** her Act . But it is said that for the Government to **adopt** the abolition policy, would involve the loss of the support of the Union men of the Border Slave States.
LAW Repeal	A REPEAL event occurs when an active law (Object) is completely repealed by a state actor (Agent).	... that I determined to revoke the act of the Federal Constituent Assembly , whereby Slavery was **abolished**. Even the New York Tribune protests against making this war for the **destruction** of slavery, and insists that such a war would alienate a large body of the Northern people at present who adhere to the Government in the prosecution of the war. They want to se the Government march a powerful array into the traitorous States, proclaim liberty to every slave, and **wipe out** the last vestige of that barbarous system from the land ...
CONFLICT Protest	A PROTEST event occurs when people (Agent) come into a public area to demand some action. PROTEST events include, but are not limited to, protests, sit-ins, and riots as the result of a previous protest.	Almost simultaneously with the appearance of the minstrels there arose from every kennel in the neighborhood timely protest **barked** forth vigorously by a hundred curs, who, in common with their masters, cursed their common luck. It has attempted to supplant Government with anarchy, and the fury of a brutal **mob** for the beneficent operation of law, and the legally appointed law-makers. ...while the majority of the men were absent at a public **demonstration** at Myrtle-avenue Park, in another part of the city.
CONFLICT Attack	An ATTACK event occurs when a person or a organization (Agent) performs an violent act causing harm or damage to another person or organization (Patient).	Make the slave first, midst, and last Follow no longer the partial and side issues; **strike** for the abolition of slavery. . . . and hovering about Williamsport in an unaccountable manner - while the rebel troops are **burning , destroying** , pressing loyal men into service , or driving them from the houses they hoped to possess , and the wheat-fields they expected to reap , under the protecting folds of the Federal flag . The States which **rebelled** , after having been most thoroughly **whipped** in a great war , came back into the Union upon their promises to abide by the Constitution and Laws of the same .
CONFLICT Injure-Die	A Injure-Die event is defined as a death or wound of a person (Patient) which is the result of a violence event by another person (Agent).	The life of loyal men are being **sacrificed** by scores, and will, by and by, be **sacrificed** by thousands. Why should the nation **pour out its blood** and lavish its treasure by the million, consent to protect and preserve the guilty cause of all its troubles? Our loss is estimated at two hundred **killed** and **wounded**.

Table 7: Event types with their descriptions and examples in the BRAD dataset (to be continued in Table 8). Event trigger words are shown in bold.

Type	Description	Examples
CONFLICT Other	Conflict Other events are reserved for events that are related to conflicts, but not classified as one of the conflict event types above, including declaring war, threatening someone, forming an army, a movement, and a march.	Let the slaves and free colored people be **<u>called into service</u>**, and **<u>formed</u>** into a liberating army, to march into the South ... Their efforts in this direction have been crowned by entire **success**. He had **called** loud and earnestly upon the Government for reinforcements; but the Government was practically deaf to the **call**, and left him and his brave companions either to perform a miracle, or to be completely overwhelmed by superior numbers.
JUSTICE Arrest-Jail	An Arrest-Jail Event occurs when the movement of a person (Patient) is constrained by a state actor (Agent).	It appears that he obtained his information direct from German where a supposed agent of the company had been **<u>arrested</u>**, having in his possession incriminating documents. It is said to be possible to **imprison** a man for debt in Massachusetts. He **put her in jail** at Eastville and she stayed there for some time .
JUSTICE Sentence	A SENTENCE Event takes place when a punishment for a person or an organization (Patient) is issued by a state actor (Agent).	... and any person so offending shall be guilty of a felony, and shall, on conviction, be **<u>sentenced to confinement</u>** in the penitentiary of this State, for a period not less than ten nor more than twenty years from the time of **<u>sentence</u>** pronounced on such person. If any slave or servant be convicted or any crime the punishment whereof may be death or **confinement** in the penitentiary ..., but that it has been promptly put down and the guilty parties summarily **punished**.
JUSTICE Execute	An EXECUTE Event occurs when the life of a person (Patient) is taken by a state actor (Agent).	Hector Grant James Horney, and Esther Anderson, white servants, were **<u>executed</u>** at Chester, Kent county. All these , if the demand of the Administration and its friends is gratified , are to be **hanged** ; for the **punishment** of treason by our law is **death** , ... He made some confessions, and managed finally to escape, but was arrested, taken to El Dorado, tried, and **shot** - not, however, by regular process.
HUMANITY Deprive	An DEPRIVE Event occurs when someone's right (Patient) is taken away, disrespect, or discouraged in any form of expression including but not limited to law, action, and statement.	We thank Dr. CROFTS for the assurance of his sympathy, and hope often to receive his earnest words in behalf of our **<u>enslaved</u>** people. Before the **slaved** is freed, this and a hundred other plans will be critically canvassed , and the discussion of each will elicit some truth . ... shall the four millions slaves , now **robbed** of all their rights, and **degraded** to a level with brute beast...
HUMANITY Endow	An ENDOW Event occurs when someone's right is enriched or appreciated in any form of expression including but not limited to law, action, action and statement.	And as for lynching - let all the officers of the law, with all the powers of the law, **<u>defend</u>** the rights and life of every prisoner. Before the slaved is **freed** , this and a hundred other plans will be critically canvassed They are going into every community which **offers freedom** and **protection** to their citizens , where law is justly administered and where the rights of man are **respected** ; and there are many such sections in this country ; there will be the future homes of the Negroes .

Table 8: Event types with their descriptions and examples in the BRAD dataset. Event trigger words are shown in bold.

Event	Argument Role	Entity Type	Description	Examples
LAW Propose	Agent	PER ORG	The person or organization who proposes the law	The resolutions were **proposed** by the gentleman from Ohio. Slavery has been **brought** into the House .
	Beneficiary	PER ORG	The person or organization who benefits from the proposal	
	Object	DOC	The proposed law	
	Time	TIME	When the proposal takes place	
	Location	GPE	Where the proposal takes place	
LAW Approve	Agent	PER ORG	The person or organization who approves the law	Freedom to the slave should now be **proclaimed** from the Capitol . The act was duly **approved** by the Executive , published, and announced to the civilized words.
	Beneficiary	PER ORG	The person or organization who benefit from this approval	
	Object	DOC	The approved law	
	Time	TIME	When the approval takes place	
	Location	GPE	Where the approval takes place	
LAW Repeal	Agent	PER ORG	The person or organization who repeals the law	They shall **strike down** Slavery . ... we can not see why the institution of private property was to be **abolished**.
	Beneficiary	PER ORG	The person or organization who benefits from this repeal	
	Object	DOC	The repealed law	
	Time	TIME	When the repeal takes place	
	Location	GPE	Where the repeal takes place	
CONFLICT Protest	Agent	PER ORG	The person or organization who protests	The red cap was **paraded** through Cape Haytien . The men were present at a public **demonstration** in Brooklyn .
	Patient	ORG	The organization that the agent protest against	
	Time	TIME	When the protest takes place	
	Location	GPE	Where the protest takes place	
CONFLICT Attack	Agent	PER ORG	The attacking person or organization	Fremont is **scouring** the rebels beyond the borders of Missouri . On Wednesday morning the rebels prepared to **storm** our works in Plymouth .
	Patient	PER ORG	The target of the attack	
	Time	TIME	When the attack takes place	
	Location	GPE	Where the attack takes place	
CONFLICT Injure-Die	Agent	PER ORG	The person or organization who attempts to attack or kill	They **cut** men **in half**, and pieces from exploded shells, **killed** and **wounded** several . Most of the negroes , we regret to hear, are said to have been **massacred**.
	Patient	PER ORG	The person or organization who is injured or dead	
	Time	TIME	When the injury/death takes place	
	Location	GPE	Where the injury/death takes place	

Table 9: Descriptions and examples of argument roles for each event type in the BRAD dataset (to be continued in Table 10). Event triggers are bold and underlined. Arguments are highlighted using colors that match with their roles.

Event	Argument Role	Entity Type	Description	Examples
CONFLICT Other	Agent	PER ORG	The acting person or organization	The poor men of the South have been **pressed into the army** to fight the battle of slavery.
	Patient	PER ORG	The person or organization who is the object of the act	
	Time	TIME	When the action takes place	
	Location	GPE	Where the action takes place	
JUSTICE Arrest-Jail	Agent	PER ORG	The arresting agent or jailer	A few weeks ago , a man named Hancock was **arrested** in Union county, Arkansas . Several free colored men were **captured** with the rebels in Fort Fisher .
	Patient	PER	The person who is arrested	
	Time	TIME	When the arrest/jail takes place	
	Location	GPE	Where the arrest/jail takes place	
JUSTICE Sentence	Agent	PER ORG	The judge or court	It has been represented there are **confined** in the Government jail forty-five prisoners , who are not charged with crime, but are represented as being slaves
	Patient	PER ORG	The person who is sentenced	
	Time	TIME	When the sentencing takes place	
	Location	GPE	Where the sentencing takes place	
JUSTICE Execute	Agent	PER ORG	The person/organization who orders or carry out the execution	A man by the name of Martin . was tried in El Dorado on a similar charge and **hanged**. They seized him, and being then convinced of his guilt, **shot** him in the woods
	Patient	PER	The person who is executed	
	Time	TIME	When the execution takes place	
	Location	GPE	Where the execution takes place	
HUMANITY Endow	Agent	PER ORG	The person/organization who endows the patient	It is our right to **liberate** the slaves of an enemy. ... and **give freedom** to persons held to labor in the slave states .
	Patient	PER	The person who is endowed	
	Time	TIME	When the endowment takes place	
	Location	GPE	Where the endowment takes place	
HUMANITY Deprive	Agent	PER ORG	The person/organization who deprives the patient	They have **outraged**, and **robbed**, and **murdered** our peaceful citizens . ... by the atrocities of the rebels **delivering back into bondage** thousands of slaves .
	Patient	PER	The person who is deprived	
	Time	TIME	When the deprivation takes place	
	Location	GPE	Where the deprivation takes place	

Table 10: Descriptions and examples of argument roles for each event type in the BRAD dataset. Event triggers are bold and underlined. Arguments are highlighted using colors that match with their roles.