

Constructing an Annotated Corpus for Protest Event Mining

Peter Makarov[†] Jasmine Lorenzini* Hanspeter Kriesi*

[†]Institute of Computational Linguistics, University of Zurich, Switzerland

*Department of Political and Social Sciences, European University Institute, Italy

makarov@cl.uzh.ch {jasmine.lorenzini,hanspeter.kriesi}@eui.eu

Abstract

We present a corpus for protest event mining that combines token-level annotation with the event schema and ontology of entities and events from protest research in the social sciences. The dataset uses newswire reports from the English Gigaword corpus. The token-level annotation is inspired by annotation standards for event extraction, in particular that of the Automated Content Extraction 2005 corpus (Walker et al., 2006). Domain experts perform the entire annotation task. We report competitive intercoder agreement results.

1 Introduction

Social scientists rely on event data to quantitatively study the behavior of political actors. Public protest (demonstrations, industrial strikes, petition campaigns, political and symbolic violence) accounts for a large part of events involving sub-state actors. Protest event data are central to the study of protest mobilization, political instability, and social movements (Hutter, 2014; Koopmans and Rucht, 2002).

To advance the machine coding¹ of protest data, we have been building a manually annotated corpus of protest events. Our protest event coding follows guidelines adapted from successful manual coding projects. All coding decisions are supported by careful token-level annotation inspired by annotation standards for event extraction. Both event cod-

¹We use the social science term *coding* to refer to information extraction. Automated extraction is *machine coding*, manual extraction is *human coding*. Event coding happens at the document level. We use the linguistic term *token-level annotation* to refer to attaching metadata to tokens in the document.

ing and token-level annotation are performed by domain experts. We find that domain experts without specialist linguistic knowledge can be trained well to follow token-level annotation rules and deliver sufficient annotation quality.

Contentious politics scholars often need more fine-grained information on protest events than can be delivered by available event coding software. Our event schema includes issues—the claims and grievances of protest actors—and the number of protesters. We also code protest events that are not the main topic of the report. This is often desirable (Kriesi et al., 1995), although event coding systems would not always code them by design.

We code newswire reports from the widely used English Gigaword corpus and will release all annotations.²

2 Related Work

2.1 Machine coding of events

The machine coding of political event data from newswire text goes back to early 1990s and has been first applied to the study of international relations and conflicts (Gerner et al., 1994; Schrodt and Hall, 2006). Many widely used systems—e.g. TABARI (O’Brien, 2010) / PETRARCH³, VRA-Reader⁴ (King and Lowe, 2003)—have relied on pattern matching with large dictionaries of hand-crafted patterns. A system scans a news lead attempting to match an event, source and target actors—thereby extracting *who did what to whom*; the date of the event is taken

²https://github.com/peter-makarov/apea_corpus

³<https://github.com/openeventdata/petrarch>

⁴<http://vranet.com>

to be the date of publication. Common ontologies CAMEO (Gerner et al., 2002) and IDEA (Bond et al., 2003) define dozens of event types and hundreds of actors. Proprietary event coder BBN ACCENT, which uses statistical entity and relation extraction and co-reference resolution, considerably outperforms a pattern matching-based coder (Boschee et al., 2013; Boschee et al., 2015). O’Connor et al. (2013) present an unsupervised Bayesian coder, which models the gradual change in the types of events between actors.

Pattern-matching coders have been found to predict event types on a par with trained human coders (King and Lowe, 2003) and sufficiently accurate for near real-time event monitoring (O’Brien, 2010). That event coding is hard and coding instructions are often not rigorous enough manifests itself in low intercoder reliability (Schrodt, 2012). Boschee et al. (2015) report an intercoder agreement of F1 45% for two human coders coding 1,000 news reports using only the top event types of the CAMEO ontology.

2.2 Machine coding of protest events

Pattern matching-based systems have been employed to assist humans in coding protest events (Imig and Tarrow, 2001; Francisco, 1996). Some (Maher and Peterson, 2008) use only machine-coded protest events. More recently, statistical learning has been applied to the coding of protest events. Hanna (2014) trains a supervised learning system leveraging the events hand-coded by the Dynamics of Collective Action⁵ project (Earl et al., 2004). Nardulli et al. (2015) employ a human-in-the-loop coding system that learns from human supervision.

2.3 Corpora annotated with protest events

A major benchmark for event mining is the Automated Content Extraction (ACE) 2005 Multilingual Training Corpus (Walker et al., 2006). The corpus, distributed by the Linguistic Data Consortium, comes with token-level annotations of entities, relations, and events. Its event ontology includes the CONFLICT event type. Its sub-type ATTACK overlaps with violent protest; the other sub-type, DEMONSTRATE, is close to our understanding of demonstrative protest. Some important protest event types

⁵<http://web.stanford.edu/group/collectiveaction>

e.g. petition campaign, industrial strike, symbolic protest, are not included. Unlike our project, the targets of ATTACK events are annotated (but not the targets of DEMONSTRATE events). Issues and the number of participants are not annotated.

Of some interest is the corpus of Latin American terrorism⁶ used in the Message Understanding Conference evaluations 3 and 4 (Chinchor et al., 1993). It comes with a highly complex event schema that includes detailed information on the actor, human and physical targets, and distinguishes several types of terrorist acts. The corpus predates information extraction by statistical learning from annotated text and thus does not contain token-level annotation.⁷

3 Annotated Corpus of Protest Events

The main motivation for this work has been the connection of event coding, which is performed at the level of the document, to token-level annotation. In that respect, we follow the trend towards annotating for social science tasks at below the document level (Card et al., 2015; Žukov-Gregorič et al., 2016). Unlike these projects, we have chosen to train domain experts to perform careful token-level annotation. The downside of having coders annotate in a linguistically unconstrained manner—an approach sometimes advocated for annotation tasks performed by domain experts (Stubbs, 2013)—is that the resulting annotation requires extensive standardization. This is challenging in the case of a complex task like ours.

The overall coding procedure is thus twofold. The coders perform traditional event coding, which involves the identification of protest events and classification of their attributes (type, actors, etc.). In parallel, the coders carry out token-level annotation, which we motivate as a means of supporting coding decisions with the help of text. The coder connects the two by linking coded events to their mentions in the text. Figure 1 shows sample annotation.

All our coders are doctoral students in political science. All are non-native English speakers with a high command of English. One project leader is a trained linguist.

⁶http://www-nlpir.nist.gov/related_projects/muc/muc_data/muc34.tar.gz

⁷DAPRA’s Deep Exploration and Filtering of Text (DEFT) program has in recent years produced event extraction resources (Aguilar et al., 2014), which currently are only available for closed system evaluation. It uses the ACE event ontology.

event no	event type	location	date	actor	issue	size
1, 2, 3, 4	Occupation/Blockade	France	29.09.2003	Occupational group	For labor rights	100-999 people

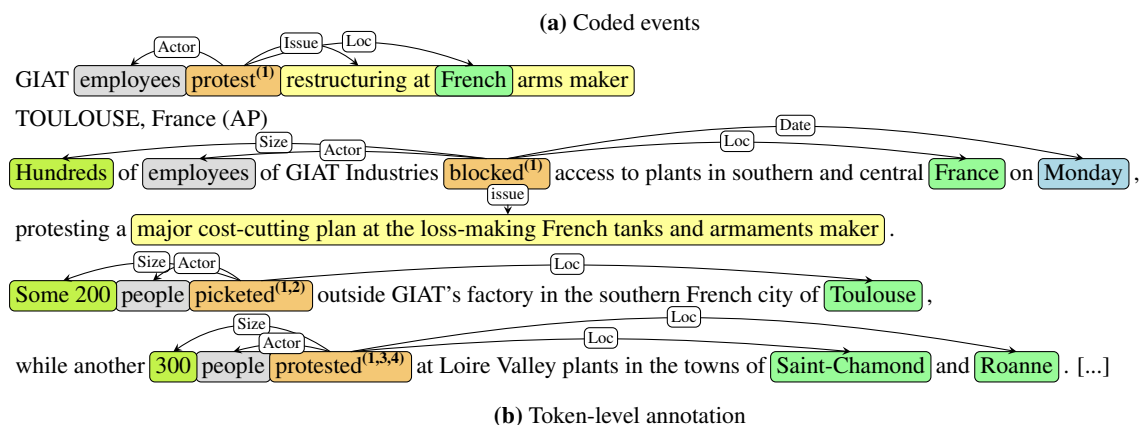


Figure 1: An annotation example. (1a) coded events. The coder has identified four protest events—all of the same structure. Event one is a campaign event, the other events are the episode events of event one (e.g. “protested” refers to two episode events, which are differentiated based on the two distinct city-level locations). (1b) in-text annotations. Event mentions are in orange. In the superscript are the indices of the coded events that an event mention refers to. In the annotation interface, the coder draws links from event mentions to the mentions of event attributes (actor, location, date, etc.).

3.1 Event schema and ontology

A protest event has an event type, a date range, a country, a number of participants, a set of actor types, and a set of issue types. We distinguish ten protest event types, twenty-six issues, and twelve actor types. The types are organized hierarchically. We have not used a large ontology of entities and events that one typically finds in event coding. Our aim has been to ensure that each type occurs sufficiently often and the reliability of coding does not suffer due to codebook complexity.

The choice and definition of the types reflect our primary interest in European protest. Having examined a sample of recent European protest events coded using a larger codebook, we have selected some frequent actor types and issues and reworked the event ontology. For example, all specific issues now come with the stance on the issue fixed: *against cultural liberalism, for regionalism, etc.*

We code only asserted specific past and currently unfolding events—in contrast to the ACE 2005 corpus⁸ and despite the practice of coding planned future events in manual coding work.

3.2 Token-level annotation

In devising token-level annotation guidelines, we have relied on the ACE English Annotation Guidelines for Events and Entities⁹ (Doddington et al., 2004). We have borrowed many ideas, e.g. the annotation of event mentions largely as one-word triggers, which we have found to work well in practice. The ACE guidelines are written for annotators with a background in linguistics, not domain experts. We have found that it is often possible to convey more or less the same idea in less technical language, e.g. simplifying *present-participle in the nominal pre-modifier position* to *participle modifying a noun*, and by providing extensive examples.

Not all ACE rules could be adapted in this way. We do not distinguish between heads and multi-word extents, but rather annotate the one which appears easier for a given attribute. For example, we annotate collective actors (“postal workers”, “left-wing protesters”) as head nouns only and not full noun phrases, which would be more in line with the ACE guidelines but is challenging even for trained linguists. On the other hand, issue annotations are predominantly multi-word expressions.

The linking of coded events to token-level annotation is at the core of our approach. To consolidate

⁸Yet, much like the lighter-weight DEFT ERE (Entities, Relations, Events) annotation standard (Aguilar et al., 2014).

⁹<https://www ldc.upenn.edu/collaborations/past-projects/ace/annotation-tasks-and-specifications>

the information about an event scattered across multiple sentences, we would need an annotated event co-reference. Yet, annotating event co-reference (as well as non-strict identity relations) is hard (Hovy et al., 2013). In the annotation interface, the coders explicitly link coded events to event mentions that refer to them, and many events can be linked to the same event mention. Thus, unlike the ACE 2005 corpus, we do not explicitly define the co-reference relation between event mentions, but read it off the explicit references. We do not annotate entity co-reference.

3.3 Workflow

We code newswire texts by the Agence France Press (AFP) and the Associated Press World (APW) of the English Gigaword Third Edition Corpus (Graff et al., 2007). We sample from two-month periods in the 2000s. For a given period, we select all documents that mention a European location in the date-line and are found protest-relevant by a classifier that we have trained, for a related project, on newswire stories including AFP and APW. For each month and agency, we randomly sample forty documents of which a project leader picks ten to twenty documents for coding. In this way, we typically get groups of documents from various dates, each group covering the same story.

Each document gets coded by at least two coders. A project leader performs adjudication. We estimate that coding one document takes an average of fifteen minutes. Our budget allows for coding up to 300 documents. In the first rounds of coding, we have not used any pre-annotation.

3.4 Inter-coder reliability

We achieve competitive inter-coder agreement for the first batch of documents (Table 1). During the coding of this batch, the coders received general feedback on token-level annotation (Table 1b), which partly explains the high agreement. For reference, we show the agreement achieved by the ACE coders on newswire documents annotated with events of type CONFLICT. Crucially, the ACE documents are almost twice as long on average, which drags down agreement. While the agreement on coded events is expectedly low, our coders agree substantially on coding subsets of event attributes (Table 1d).

This work	ACE'05 nw Conflict
-----------	--------------------

a) **Dual-pass statistics:** number of documents coded independently by two coders, average number of tokens per document.

# docs	avg # tokens per doc	# docs	avg # tokens per doc
30	363.5	53	705.9

b) **Token-level annotation:** F1-score agreement for exact match (F1_E) and match by overlap (F1_O). For ACE: Location=Place, Actor=Attacker or Entity, Time=any of the Time-* types.

	avg # per doc	F1 _E	F1 _O	avg # per doc	F1 _E	F1 _O
Trigger	4.3	75.6	76.4	5.1	48.9	62.1
Location	2.7	80.2	81.5	3.0	51.9	60.0
Time	1.7	87.4	87.4	2.0	54.3	66.7
Actor	2.5	79.7	82.7	3.0	52.0	60.2
Issue	1.2	54.0	76.2	-	-	-
Size	0.7	76.9	92.3	-	-	-

c) **Event co-reference:** average number of coded events per document, CoNLL F1-score agreement. For our corpus, we take two event mentions to co-refer if they are linked to the same events.

avg # per doc	CoNLL F1	avg # per doc	CoNLL F1
2.0	55.39	3.4	40.11

d) **Coding (this work):** F1-score agreement on (event type, location, date range). Coders agree if types and locations match, and date ranges overlap. F1-score agreement on event attributes given that two events agree on type, location, date range as described.

Event type / Loc / Date	F1 58.2	
	avg # unique labels per doc	F1 given type/loc/date match
Actors	0.82	78.1
Issues	0.83	78.1
Size	0.55	93.2

Table 1: Inter-coder agreement for two coders. For reference, we show agreement scores for the ACE coders on the ACE 2005 newswire reports annotated for events by both coders s.t. at least one report of each pair has CONFLICT events. We only consider specific asserted past/present events and their arguments.

4 Conclusion and Future Work

We have presented our work on a corpus of protest events, which combines event coding with careful token-level annotation. The corpus comes with coded issues and numbers of participants. Overall, we observe substantial inter-coder agreement.

Little work has been done on the evaluation of event coders (Boschee et al., 2013),¹⁰ and none on widely available data despite interest (Schrodt, 2016). We would encourage the use of our corpus as an evaluation benchmark. That would require mapping our ontology of events and entities to CAMEO categories.

As we often code groups of documents covering the same sets of events (Section 3.3), the corpus could be extended to include cross-document event co-reference annotations.

¹⁰There has been work comparing datasets of automatically coded events (Ward et al., 2013).

Acknowledgments

We would like to thank Swen Hutter, Elisa Volpi, Joldon Kutmanaliev, Chiara Milan, Daniel Schulz, and the two anonymous reviewers. This work has been supported by European Research Council Grant No. 338875.

References

- [Aguilar et al.2014] Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. A comparison of the events and relations across ACE, ERE, TACKBP, and FrameNet annotation standards. In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53.
- [Bond et al.2003] Doug Bond, Joe Bond, Churl Oh, J Craig Jenkins, and Charles Lewis Taylor. 2003. Integrated data for events analysis (IDEA): An event typology for automated events data development. *Journal of Peace Research*, 40(6):733–745.
- [Boschee et al.2013] Elizabeth Boschee, Premkumar Natarajan, and Ralph Weischedel. 2013. Automatic extraction of events from open source text for predictive forecasting. In *Handbook of Computational Approaches to Counterterrorism*, pages 51–67. Springer.
- [Boschee et al.2015] Elizabeth Boschee, Jennifer Lautenschlager, Sean O’Brien, Steve Shellman, James Starz, and Michael Ward. 2015. BBN ACCENT Event Coding Evaluation.updated v01.pdf (ICEWS Coded Event Data). <http://dx.doi.org/10.7910/DVN/28075>.
- [Card et al.2015] Dallas Card, Amber E Boydston, Justin H Gross, Philip Resnik, and Noah A Smith. 2015. The media frames corpus: Annotations of frames across issues. In *ACL*, volume 2, pages 438–444.
- [Chinchor et al.1993] Nancy Chinchor, David D Lewis, and Lynette Hirschman. 1993. Evaluating message understanding systems: an analysis of the third message understanding conference (muc-3). *Computational linguistics*, 19(3):409–449.
- [Dodgington et al.2004] George R Dodgington, Alexis Mitchell, Mark A Przybocki, Lance A Ramshaw, Stephanie Strassel, and Ralph M Weischedel. 2004. The Automatic Content Extraction (ACE) Program-Tasks, Data, and Evaluation. In *LREC*, volume 2, page 1.
- [Earl et al.2004] Jennifer Earl, Andrew Martin, John D McCarthy, and Sarah A Soule. 2004. The use of newspaper data in the study of collective action. *Annual review of Sociology*, pages 65–80.
- [Francisco1996] Ronald A Francisco. 1996. Coercion and protest: An empirical test in two democratic states. *American Journal of Political Science*, pages 1179–1204.
- [Gerner et al.1994] Deborah J Gerner, Philip A Schrodt, Ronald A Francisco, and Judith L Weddle. 1994. Machine coding of event data using regional and international sources. *International Studies Quarterly*, pages 91–119.
- [Gerner et al.2002] Deborah J Gerner, Philip A Schrodt, Omür Yilmaz, and Rajaa Abu-Jabr. 2002. Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*.
- [Graff et al.2007] David Graff, Linguistic Data Consortium, et al. 2007. *English Gigaword Third Edition LDC2007T07*. Linguistic Data Consortium.
- [Hanna2014] Alex Hanna. 2014. Developing a system for the automated coding of protest event data. Available at SSRN: <http://ssrn.com/abstract=2425232> or <http://dx.doi.org/10.2139/ssrn.2425232>.
- [Hovy et al.2013] Eduard Hovy, Teruko Mitamura, Felisa Verdejo, Jun Araki, and Andrew Philpot. 2013. Events are not simple: Identity, non-identity, and quasi-identity. In *NAACL HLT*, volume 2013, page 21.
- [Hutter2014] Swen Hutter. 2014. Protest event analysis and its offspring. In Donatella Della Porta, editor, *Methodological practices in social movement research*. Oxford University Press.
- [Imig and Tarrow2001] Doug Imig and Sidney Tarrow. 2001. Mapping the Europeanization of contention: evidence from a quantitative data analysis. *Contentious Europeans: Protest and politics in an emerging polity*, pages 27–49.
- [King and Lowe2003] Gary King and Will Lowe. 2003. An automated information extraction tool for international conflict data with performance as good as human coders: A rare events evaluation design. *International Organization*, 57(03):617–642.
- [Koopmans and Rucht2002] Ruud Koopmans and Dieter Rucht. 2002. Protest event analysis. *Methods of social movement research*, 16:231–59.
- [Kriesi et al.1995] Hanspeter Kriesi, Ruud Koopmans, Jan Willem Duyvendak, and Marco G Giugni. 1995. *New social movements in Western Europe: A comparative analysis*, volume 5. U of Minnesota Press.
- [Maher and Peterson2008] Thomas V Maher and Lindsey Peterson. 2008. Time and country variation in contentious politics: Multilevel modeling of dissent

- and repression. *International Journal of Sociology*, 38(3):52–81.
- [Nardulli et al.2015] Peter F Nardulli, Scott L Althaus, and Matthew Hayes. 2015. A progressive supervised-learning approach to generating rich civil strife data. *Sociological Methodology*.
- [O’Brien2010] Sean P O’Brien. 2010. Crisis early warning and decision support: Contemporary approaches and thoughts on future research. *International Studies Review*, 12(1):87–104.
- [O’Connor et al.2013] Brendan O’Connor, Brandon M Stewart, and Noah A Smith. 2013. Learning to extract international relations from political context. In *ACL (1)*, pages 1094–1104.
- [Schrodt and Hall2006] Philip A Schrodt and Blake Hall. 2006. Twenty years of the Kansas event data system project. *The political methodologist*, 14(1):2–8.
- [Schrodt2012] Philip A Schrodt. 2012. Precedents, progress, and prospects in political event data. *International Interactions*, 38(4):546–569.
- [Schrodt2016] Philip A Schrodt. 2016. Comparison metrics for large scale political event data sets. <http://www.polmeth.wustl.edu/files/polmeth/schrodtepsa15eventdata.pdf>.
- [Stubbs2013] Amber C Stubbs. 2013. *A Methodology for Using Professional Knowledge in Corpus*. Ph.D. thesis, Brandeis University.
- [Žukov-Gregorič et al.2016] Andrej Žukov-Gregorič, Bartal Veyhe, and Zhiyuan Luo. 2016. IBC-C: A dataset for armed conflict event analysis. In *ACL*.
- [Walker et al.2006] Christopher Walker, Stephanie Strassel, Julie Medero, and Kazuaki Maeda. 2006. ACE 2005 multilingual training corpus. *Linguistic Data Consortium, Philadelphia*.
- [Ward et al.2013] Michael D Ward, Andreas Beger, Josh Cutler, Matt Dickenson, Cassy Dorff, and Ben Radford. 2013. Comparing GDELT and ICEWS event data. *Analysis*, 21:267–297.