




**KOÇ  
ÜNİVERSİTESİ**

A light blue world map with white outlines of continents and countries, serving as a background for the title text.

# **Global Contentious Politics Database Annotation Manuals**

Version 1.0

2021-04-09

**EMERGING  
WELFARE** 

The logo for Emerging Welfare, featuring the text "EMERGING WELFARE" in bold black uppercase letters, followed by a stylized logo composed of green and black geometric shapes.

**GLOBAL  
CONTENTIOUS  
POLITICS** 

The logo for Global Contentious Politics, featuring the text "GLOBAL CONTENTIOUS POLITICS" in bold black uppercase letters, followed by a stylized logo composed of purple and black geometric shapes.

## Table of Contents

1	Introduction.....	1
2	Document Level Protest Annotation Manual.....	4
2.1	Identifying Protest Events.....	4
2.1.1	Types of collective action.....	4
2.1.1.1	Political mobilizations.....	4
2.1.1.2	Social Protests.....	5
2.1.1.3	Group Confrontations.....	5
2.1.2	Types of Events.....	5
2.1.2.1	Demonstrations.....	5
2.1.2.2	Industrial Actions.....	5
2.1.2.3	Group clashes.....	6
2.1.2.4	Political violence and militancy.....	6
2.1.3	Necessary characteristics of events to be annotated.....	6
2.1.3.1	Necessity of civilian actors.....	6
2.1.3.2	Time and location necessities.....	6
2.1.4	Examples of cases that should be considered protest events.....	7
2.1.5	Examples of cases that should not be considered protest events.....	9
2.1.5.1	Events that fall outside our protest event definition.....	10
2.1.5.2	Events that are included in the event definition but lack necessary characteristics.....	12
3	Sentence Level Protest Annotation Manual.....	14
3.1	Classifying Sentences.....	14
3.1.1	Event sentences.....	14
3.1.2	Non-event sentences.....	14
3.1.3	Planned event sentences.....	14
3.2	Notes and examples.....	15
4	Token Level Protest Annotation Manual.....	18
4.1	Introduction.....	18
4.2	Terminology.....	19
4.3	General Rules of Token Level Annotation.....	19
4.3.1	How annotation proceeds.....	19
4.3.2	When to tag event arguments?.....	20
4.3.3	Overlapping Annotation.....	21
4.3.4	Tag span rules:.....	22
4.3.4.1	Tag span continuity.....	22
4.3.4.2	Tag spans in complex expressions:.....	23
4.3.4.3	Tagging punctuation and articles.....	23
4.4	Document Information Tags.....	24
4.4.1	Title of the document.....	24
4.4.2	Document publication time.....	24
4.4.3	Document publication place.....	25
4.5	Event Reference Tags.....	25
4.5.1	Type of the Event.....	25
4.5.2	Event Mention.....	28
4.5.3	Semantic category of events.....	29
4.5.3.1	Demonstrations.....	29

4.5.3.2 Industrial actions.....	29
4.5.3.3 Group clashes.....	30
4.5.3.4 Armed militancy.....	30
4.5.3.5 Electoral politics.....	30
4.5.3.6 Other.....	31
4.6 Event's Time and Location Arguments.....	31
4.6.1 Time of the event.....	31
4.6.2 Place of the event.....	32
4.6.3 The facility in which the event takes place.....	33
4.6.4 Centrality of the event location (Urban or rural).....	34
4.7 Event Actor Arguments.....	35
4.7.1 Participant Arguments.....	35
4.7.1.1 Types of participants.....	35
4.7.1.2 Names of participants.....	36
4.7.1.3 Participant Count.....	36
4.7.1.4 Ideology of the Participants.....	37
4.7.1.5 Ethnic identity of participants.....	37
4.7.1.6 Religious Identity of Participants.....	37
4.7.1.7 Caste of the Participants.....	38
4.7.1.8 Socioeconomic Status of the Participants.....	38
4.7.2 Semantic Categories of Participants.....	38
4.7.2.1 Peasant.....	39
4.7.2.2 Worker.....	39
4.7.2.3 Small producer.....	39
4.7.2.4 Employer/executive.....	39
4.7.2.5 Professional.....	39
4.7.2.6 Student.....	39
4.7.2.7 Politician.....	39
4.7.2.8 Activist.....	39
4.7.2.9 Militant.....	40
4.7.2.10 People.....	40
4.7.2.11 Other.....	40
4.7.3 Organizer Arguments.....	40
4.7.3.1 Distinguishing between organizers and participants.....	40
4.7.3.2 The event organizer type.....	41
4.7.3.3 The Name of the Event Organizer.....	42
4.7.3.4 The Ideology of the Event Organizer.....	43
4.7.3.5 The Ethnicity of the Event Organizer.....	43
4.7.3.6 The Religion of the Event Organizer.....	43
4.7.3.7 Caste of the Event Organizer.....	43
4.7.3.8 Socioeconomic Status of the Organizer.....	43
4.7.4 Semantic Categories of Organizers.....	44
4.7.4.1 Political party.....	44
4.7.4.2 NGO.....	44
4.7.4.3 Union.....	44
4.7.4.4 Militant/armed organization.....	45
4.7.4.5 Chamber of professional.....	45
4.7.4.6 Person.....	45
4.7.4.7 Other.....	45

4.7.5 Target Arguments.....	45
4.7.5.1 Target Type.....	46
4.7.5.2 Target Name.....	46
4.8 Event Separation.....	47
4.8.1 Introduction.....	47
4.8.2 Situations in which events will be separated.....	47
4.8.3 Rules of event separation.....	48
5 Violent Protest Events Annotation Manual.....	51
5.1 Identifying violent protest events.....	51
5.1.1 Some protest events that are violent.....	51
5.1.2 Some protest events that are not violent.....	52
6 Protest Event Demands Annotation Manual.....	53
6.1 Non-Economic Demands.....	53
6.2 Economic Non-Welfare Demands.....	53
6.3 Economic Welfare Demands.....	53
6.4 Rules of Demand Annotation.....	54
7 Publications Based on the Annotation Manual.....	55

# 1 Introduction

Emerging Markets Welfare project investigates the effects of contentious politics on welfare state programmes in countries of the Global South. It hypothesizes that government responses to social contention is a significant factor that shapes welfare policies. It is in this respect that mapping the dynamics of social contention in a given country becomes crucial, and duly constitutes a fundamental component of the entire project. Investigating the causal relationship between social contention and government policy involves more than simple correlation, particularly if the focus is on a specific government action, namely welfare policies. The map of social contention adequate for such an understanding should thus go beyond laying out basic trends of ebbing and flowing of social contention over space and time, and provide insight into particularities such as the types of action repertoires, levels of violence, characteristics of actors or social groups that engage in contentious politics, the characteristics of the demands that they raise.

The purpose of the second work package of the EMW Project is drawing that map of social contention. For achieving this purpose, we are creating a database of contentious politics events through extraction of information from news reports that are featured in the most prominent online sources each focus country has to offer. The Global Contentious Politics Database (GLOCON) will record contentious politics events (referred to as protest events for the sake of brevity) that take place within the borders of our focus countries with all the information available in the source about the events' time and place, actor, type, demands raised, violence level. As of the moment, the GLOCON database will contain protest event data from India, China, South Africa, Argentina and Brazil. It will feature data in three languages: English for India, China and South Africa data; Spanish for Argentina data, and Portuguese for Brazil data. The database will be created in a way to accommodate additions of other focus countries and/or news sources in the future.

The database creation will depend on automated text processing tools that will detect if a news article contains a protest event, locate protest information within the article, and extract pieces of information regarding the detected protest events. The basis of training and testing the automated tools is the GLOCON Gold Standard Corpus (GSC), which contains news articles from multiple sources from each focus country. The articles in the GSC will be manually coded by skilled annotators in both classification and extraction tasks with utmost accuracy and consistency that automated tool development demands. In order to assure these, the annotation manuals<sup>1</sup> in this document will lay out the rules according to which annotators will code the news articles. Annotators will refer to the manuals at all times for all annotation tasks and will conform to the rules that they contain. Despite the EMW Projects focus on the countries of the Global South, and the initial choice of a limited number of countries to be featured in the GSC, none of the rules or

---

1 This manual will be published along with the forthcoming publication Hürriyetoglu, A., Yörük, E., Mutlu, O., Duruşan, F., Yoltar, Ç., Yüret, D. and Gürel, B.: Cross-context news corpus for protest event-related knowledge base construction. *Data Intelligence* 3(2), 2021. doi: 10.1162/dint\_a\_00092. See the final section for the list of publications from the EMW Project that refer to it. Regarding the contents of the manual, please contact Fırat Duruşan (email: [fdurusan@ku.edu.tr](mailto:fdurusan@ku.edu.tr)).

principles contained in this manual is more or less applicable to certain countries, sources or periods than others. The aim of the GLOCON database is to be inclusive and capable of expanding. Securing consistency, reliability and validity of data in the face of temporal and spatial expansion requires that annotation principles are generally applicable, and that they are applied consistently.

The annotation process is composed of three main levels for each news report document. The document level annotation determines the news articles that contains information on actual (past or ongoing) protest events. The sentence level annotation aims to locate sentences that contain protest event related information. In the final phase, words or phrases that give concrete information about protest events are detected.

Corresponding to the document and sentence classification, and information extraction tasks, there are three main and two supplemental manuals which together cover the entire annotation process from document, through sentence, to the token level. The first manual is the Document Level Protest Annotation Manual (DOLPAM) which establishes the rules of determining news articles that contain protest events; in other words classifying news articles into those which contain protest events and those which do not. It lays out the protest event ontology, that is, the protest event definition which specifies the range of contentious politics events that are included in the scope of the project. It introduces and exemplifies different types of protest events, and defines the criteria to which a news report must conform in order to be labeled as a protest event article. The following Sentence Level Protest Annotation Manual (SELPAM) carries on with classifying the sentences in the documents that have already been classified as protest event articles. Similarly, it defines and exemplifies event sentences and enumerates the rules by which sentences are labeled as event sentences and non-event sentences. The third and final main manual is the Token Level Protest Annotation Manual (TOLPAM) which is the longest and most detailed of the three main level manuals. It defines the types of event related information that the project aims at collecting from news articles and explains how expressions within the event sentences which contain these pieces of information are tagged. The remaining two manuals are supplemental manuals that label further information about the events that are already extracted in the three main levels of annotation. Both define annotation tasks that are performed on the document level. The first is the Violent Protest Events Annotation Manual which lays out the rules of classifying news reports that contain protest events into categories of violent and non-violent. The following, Protest Event Demands Annotation Manual aims at setting the rules for labeling the demands and/or grievances associated with the protest events that are extracted in the news articles. More detailed information about each manual can be found under their respective headings.

Even though every particular level of annotation has its own annotation manual, the whole process must be thought of as an integrated whole as each level of annotation is premised on the results of the previous level. Hence, familiarizing oneself with all the manuals before starting annotation on any single level is recommended. Knowing in advance what the sentence and token level annotation tasks entail would help an annotator working on the document level considerably. That

said, it is neither practical nor advisable to try to learn all annotation procedures by heart. Memory is prone to mislead, and recurrent reference to the manuals is the preferred way of utilizing them. Thus, annotators must read the entire manual before starting annotation, and remember to refer to it when there is the slightest doubt about a rule or a difficult case.

The content of the annotation manual is built on the general principles and standards of linguistic annotation laid out in other prominent annotation manuals such as ACE, CAMEO, and TimeML. These principles, however, have been adapted, or rather modified heavily so as to accommodate the social scientific concepts and variables employed in the EMW project. The manual has been molded throughout a long trial and error process that accompanied the annotation of the GSC. It owes much of its current shape to the meticulous work and invaluable feedback provided by highly specialized teams of annotators, whose diligence and expertise greatly increased the quality of the corpus.

## 2 Document Level Protest Annotation Manual

The document level protest event classification task aims at determining which news articles contain contentious politics events. This manual will define the event ontology, that is the meaning and range of protest events, and how it should be decided that a news article contains such an event. All the news articles that are processed in this task will be classified into “protest” and “no protest” categories in the annotation interface we utilize for this task. Please remember that annotation decisions must be based exclusively on the text that is being processed and the rules laid out in annotation manuals. Although this point might sound obvious, there will be times when the annotator will have knowledge about the events mentioned in news articles prior to reading them, and it is only natural that such knowledge has a tendency to play a role on annotator’s decision. This will all the more be the case as annotation progresses and the annotator keeps gaining knowledge about the case countries. The focus must always remain on what the specific news article that is being annotated contains as information rather than prior or background knowledge of the annotator. Please read carefully and comprehend the event ontology below, and remember to come back to this manual whenever you encounter news articles which prove hard to decide.

### 2.1 Identifying Protest Events

The event we simply refer to as protest events are those events that are comprised within the scope of contentious politics. Contentious politics events refers to politically motivated collective action events which lay outside the official mechanisms of political participation associated with formal government institutions of the country in which the said action takes place. The term contentious politics is rather broad as to cover collective action types so diverse that it might seem counter-intuitive to group them under the same heading. In order to point out to this diversity, which the simple definition of contentious politics given above hardly reveals, different collective action, and event types will be identified below. Contentious politics as such, can be said to comprise three broad categories of collective actions and four categories of events. Remember that it is not necessary on this level to distinguish the categories from each other and pigeonhole the events encountered in news articles into one of them. This enumeration is done simply to establish the limits of the protest event ontology we operationalize.

#### 2.1.1 Types of collective action

##### 2.1.1.1 *Political mobilizations*

Political mobilizations refer to instances of collective action where political parties, organizations or individual representatives of these entities mobilize any group of participants for political goals, demands or grievances. Such political organizers might be in government or in opposition in their respective political contexts.



### **2.1.1.2 Social Protests**

The second class of collective action that contentious politics encompasses social protests in the narrow sense. A protest action is one through which individuals, groups, or organizations voice their objections, oppositions, demands or grievances to a person or institution of authority.

### **2.1.1.3 Group Confrontations**

The final broad category of collective action we consider within the scope of contentious politics is incidents of social confrontation where two or more social groups engage in political action against each other. Various sources of disagreement might exist between social groups, such as conflicting economic interests, ethnic, religious or gender-based identity differences, or ideological disagreements. Any instance of actual confrontation between groups based on such grounds is in scope of our understanding of contentious politics. Such confrontations might or might not involve state and/or political institutions.

## **2.1.2 Types of Events**

These broad categories of contentious politics manifest themselves in different types of events. Again, with the aim of simplification, four broad event type categories can be identified: demonstrations, industrial actions, group clashes and politically motivated violent actions targeting officials (including security personnel) or civilians. Note that there might be certain types of protest that are unique to a particular country, or a protest might have a unique name in a particular country. Please refer to the supervisors or country experts when you encounter a type of protest which is not covered in this manual.

### **2.1.2.1 Demonstrations**

A demonstration is defined as a form of political action in which a demand and/or grievance is raised outside the given institutionalised forms of political participation in a country. The aim of the event is to draw attention of politicians and/or general public to said goal by making it visible in public sphere which implies any space which is open to members of the public. Violent or peaceful forms of social protest that take place in open or closed public spaces such as streets, plazas, vicinity of prominent buildings fall in the category of demonstrations. Notable examples of demonstrations are as follows: Protest marches, silent marches, rallies, demonstrations, outdoor press declarations, gatherings, sit-ins, acts of civil disobedience, dharnas, bandhs, demonstrators clashing with security forces, commemorations and religious rituals (when they become means of protest, e.g. akhand paths), collective petitions (collecting signatures), shouting slogans, carrying banners, self-immolation (and threats of self-immolation), hunger strikes, fasts, barricading, picketing (road blocks), burning of vehicles, occupations of public buildings that are not included in the industrial action category and the like.

### **2.1.2.2 Industrial Actions**

Industrial actions are types of protest events that take place within workplaces and/or involve the production process in the protest. Notable examples include any kind of strike (comprising

slowdown strikes, wildcat strikes, sympathy strikes and green bans), workplace occupations, boycotts, picket lines and gheraos.

### **2.1.2.3 Group clashes**

Group clash events are instances of confrontation that stem from politicized conflicts (e.g. identity or economic interest based or ideological conflicts) between social groups such as fighting, lynching, ransacking, arson, any armed or unarmed clash between civilians of a political nature. Actions that target religious, ethnic or similar minorities fall within this category. Said actions can be unidirectional, that is, the target of the event might or might not retaliate during the event.

### **2.1.2.4 Political violence and militancy**

Finally, politically motivated violent events fall within our event definition. These actions are usually carried out by political and/or militant organizations which resort to violence, and target officials or civilians. Notable examples of this type of events are kidnapping, assassination, bombing, suicide bombing, hijacking and the like. Even though such actions are generally carried by organizations, there might be cases in which such acts are carried by individuals not necessarily affiliated to organizations or groups. Hence, if the act which intentionally resorts to violence carrying political goals, they are included in our definition even when they are carried out by a proverbial “lone wolf”.

## **2.1.3 Necessary characteristics of events to be annotated**

Having identified the types of contentious politics and associated event types, laying out necessary criteria in terms of their reporting in the news articles is in order. The protest event mentions that are encountered in the articles must have the characteristics enumerated below so that the article can be classified as a protest article.

### **2.1.3.1 Necessity of civilian actors**

Firstly, the participants or organizers of protest events must include at least one non-state actor. Non-state actors might be political or non-political organizations such as parties, associations or trade unions, or they might simply be individuals or a group of citizens such as residents of a locality or employees of a certain workplace.

### **2.1.3.2 Time and location necessities**

The most important characteristics of an event are its time and place. We expect events to have concrete times and places as these are among the most important pieces of information we aim to collect when we analyze the selected event related articles in the information extraction phase. Below you will find an abstract definition of these necessities. In order to better come to grips with these principles, you can refer to section 1.5.2. below, where they are elaborated on more concrete terms. Time and location are pieces of information which makes it certain that the event mention in the news article is that of a specific event which actually took place rather than a conceptual or general reference.

Most essentially, the event to be labeled must certainly have taken place. This means that only the past or ongoing events are to be included. The most certain indicator of this is the tense the article uses in narrating the story. The events that are in scope are most commonly narrated in past tense. For ongoing events the narrator might prefer to use simple present or continuous tenses which also indicates certainty of the events occurrence.

Secondly, we expect events to be current affairs and event articles to provide a relatively specific time information which leaves no doubt as to event's occurrence. Definite time indicators such as complete dates, days, hours, or time expressions such as yesterday, last week etc. would certainly qualify as time indicators. We also allow vague time indicators which inform us that the event is a current affair as in "Last month's Naxal violence in the region" or "recent riots in Gujarat".

We expect concrete information about the location of the event to be present in the article. The location information might be the place of an event, that is, a geographical location more specific than the country where the event took place such as state, city, town, district, village etc. If such a place is not explicitly mentioned, the facility information, that is, information regarding the type of space where the event takes place (such as a building, street, plaza, stadium etc.) will also be sufficient. However, some protest actions such as those that take place in cyberspace do not have physical locations. Likewise, certain other events do not take place in specific locations, such as product boycotts and signature campaigns. These and similar events which, by their nature, do not take place in specific physical locations must be annotated as protest events despite lacking location information.

#### **2.1.4 Examples of cases that should be considered protest events**

In this section, a number of specific protest event cases that are in our scope will be enumerated and examples will be provided. Note that the cases below are far from exhaustive in terms of our event definition. They will merely provide a number of instances where the inclusion of the event in the article is not immediately obvious. You might think of them as corner cases of sorts.

- i. Sometimes protest events exist in the news articles even though they are not mentioned in the title or are not the main topic of the article. Put differently, we are not primarily concerned with the title or main topic of the articles that we analyze, but rather whether they contain a protest event or not.
- ii. Political parties or their leaders mobilizing groups of people are manifestly included in our scope. This might look less obvious when said politicians or parties are in the government and therefore the whole affair might look official or less of a civilian character. We include these and all cases where the event is actively participated by members of the public. An example would be the political party mobilizing its supporters for an election rally or during an official ceremony celebrating a national day. Similarly, political party activities which aim at mass participation such as mass feedings or youth festivals are venues for mass mobilizations and thus are included in our event definition.

- iii. Protest actions by members of political parties, including members of parliament are included if these actions take place publicly, and outside the state institutions and normal procedures of government. The actions that take place without public participation within the confines of state institutions, such as legislative bodies, that are unusual and/or not associated with the normal operation of government are included as protest events. Events like hunger strikes or armed violence that are manifestly unusual in terms of the normal functioning of state institutions are accepted as protest events even when they physically take place within state institutions.
- iv. Events that are part of armed conflicts between states and (non-state) organizations are included regardless of the scale of the conflict, provided that they are initiated by non-state actors. Clashes between state and non-state actors which are initiated by state actors are included if it is actively responded to by the non-state actor. Military coups are exceptions to this rule (see the following section).
- v. Though rare, there are instances of state officials acting outside their official capacity and protesting. Public employees striking or engaging in work related protests are usual and easy to recognize but more subtle cases might involve them appear as carrying out their duties. Be on the lookout for acts of insubordination or misconduct that is carried out as advancing an explicitly stated political agenda that is specific to those who carry out the action. Cases in which security personnel attack civilians or refuse to prevent acts of violence among civilians might be difficult to identify as protest events as they might be associated with official policies of state repression.
- vi. Sometimes it might not be immediately obvious that an act of violence, such as a case of murder or clash between groups have a political character which would make the case a protest event in our understanding. Assassinations of or attacks against prominent public figures be it political figures or professionals of public stature (high level bureaucrats, lawyers, writers etc.) are very likely to be protest events. They should be treated as protest events unless it is certain that they happen to be caused by personal or non-political reasons.
- vii. Group clashes based on economic interests might not have an obvious political angle and look like instances of petty turf war. Exemplary cases include fights between Uber and taxi drivers or clashes between different fishermen communities. These and similar clashes are very much included in our scope.
- viii. We expect the included events to have definitely taken place or be ongoing events as mentioned above. But in certain instances, the threat or attempt of certain actions have protest effect in and of themselves despite not being carried out for one reason or another. Threats of violent actions such as assassination, bombing, attack and self-immolation are such cases which have significant effects that make them eligible for inclusion in our event definition, even if they are not carried out or somehow obstructed from being fulfilled.

- ix. Non-political contexts can well become scenes of political protests. Sports events, concerts, even religious rituals might become scenes or media of protest actions. Keep in mind that the protest events that take place in such non-political settings are also in our scope, provided that they have political motivations or target authorities.
- x. The events that are alleged by civilian or official actors to have taken place in the news reports are included even when the said claims are not verified. E.g. “The police have reportedly rounded up 11 members of the BKU for allegedly attacking one Kishan Singh”
- xi. There are certain protest events that are not space-bound, that is to say they do not take place on a specific physical space by definition. These are regarded as an exception to the location necessity rule. Protest actions that take place on world wide web that are sometimes referred to under the name of “hacktivism” such as instances of cyber-attacks of a political nature are among these and will be labeled as events. Also, events like product boycotts, sometimes referred to as consumer activism, do not take place on specific locations and must be labeled as protest events despite not meeting the location requirement.
- xii. Collective petitions and declarations that are open to members of the public for participation as signatories are included as events.
- xiii. Sometimes, riot events might be reported without explicit place or location mentions, except in the names of the participants or the residents of the area, e.g. “Tibetan protesters involved in rioting last year”. Since riot is associated by localities by definition, the participant mentions that contain place names will be regarded as location information in and of themselves. That is, these events will be labeled as protest events despite the location of the event is not explicitly mentioned.
- xiv. News articles compiled from any newspaper will contain reports about countries other than EMW project’s case countries. In this level of annotation, we do not exclude news reports about countries other than our cases. As long as the protest event mentioned in the article conforms with the definition and criteria laid out in this manual, it will be labeled as a protest event.

### **2.1.5 Examples of cases that should not be considered protest events**

In order to draw a more complete picture, this section will provide a number of illustrative cases which are not included in our event definition. Note, again, that this is not an exhaustive list but a collection of cases where it is apt to mention specifically that the events we encounter in the news articles are outside our scope.

An event is excluded from annotation either if it is outside the scope of our event ontology of contentious politics or its reporting in the news article lacks the necessary event characteristics that are defined above.

### **2.1.5.1 Events that fall outside our protest event definition**

- i. Events that cannot be considered as political protest or mobilization are obviously not included. Petty crimes, fights or clashes between civilians that lack political, economic or identity-based motivations, or actions targeting officials or political figures simply for personal reasons are thus outside our scope. At times certain cases of violent crimes might involve actions which can readily be associated with political violence, such as using bombs to rob banks or pirate attacks. Care must be taken to make sure such events have political motivations.
- ii. We do not consider individual public actions such as petitions or litigations of public character that lack collective support and remain on the individual level as protest events.
- iii. Events which are initiated by state or government actors that are not actively participated by civilians are outside our scope. These might be military operations on non-state organizations or other states, acts of repression on members of the public (note that the latter's active resistance would count as an event), official ceremonies, celebrations, commemorations organized by state/government authorities, events that take place within the auspices of the parliament and would be part of common government procedures (e.g. politicians leaving the parliament, having heated arguments, obstructing decision making etc.).
- iv. Protest actions by members of legislatures during meeting sessions are mostly habitual events and are associated with the routine functioning of these bodies. Acts such as disrupting speeches or groups of members disrupting sessions, verbal protests, banging on desks, rushing towards speakers, walking out of meetings, even fistfights etc. are all usual events in parliaments which have little to no consequence outside the legislatures. Such events are not included in our protest event definition.
- v. Events that are parts of election campaigns and, unlike election rallies, do not mobilize civilians are not considered protest events. Examples would be canvassing, distribution of election brochures, leaflets etc. which are not actively participated by members of the public.
- vi. Simply declaring certain demands and grievances, as in submitting memorandums, giving declarations to press, without the said act of declaration itself becoming a public spectacle is not considered an event. However, a press declaration outdoors is a public spectacle and a prominent way to protest in certain countries and as such is included in our event definition.
- vii. Sometimes the news stories are about general contexts or situations of conflict between a set of actors that are very prone to cause or likely to contain protest events but are not events themselves. Discerning events from situations might be trickier than it seems at first sight as the latter might be so concrete and specific as to be given concrete times and places. Annotators must be careful with expressions such as tension, conflict,

disagreement, enmity etc. as these do not denote actions (events) but rather situations. If the article does not include a concrete confrontation event (a clash for instance) which such conflicts give rise to, it should not be annotated. Examples of conflictual situations which do not count as events in and of themselves include: the tense atmosphere of prolonged conflict between a state and a militant organization which is strained in summer months when the state engages in military operations, the conflict between the managers and workers of a subcontracting firm over unpaid wages, the distress created among the workers of a shipyard by recurrent job related accidents, the tension between two religious communities living in the same town before a prominent religious holiday etc. The protest events we would observe in such situations, and thus label as such, would rather look like: militants' retaliation attack on military outposts, workers beating up their bosses, the shipyard workers occupying and stopping the work in the workplace, the mobs belonging to each community ransacking each other's small businesses etc.

- viii. Similar to the previous case, some news stories contain statements which give information about the background or the context of one or more of the story elements. Such statements might contain general allusions to, or summaries of multiple events. Such events might be very likely to be contained in our event definition. In such cases generally speaking the annotator must be careful to mark events that might elude a superficial reading. However, we are after concrete events that we can identify rather than summaries of event rich contexts. Extra care should be taken to mark only the events which are specific, concrete and can be singled out. Distinguishing concrete events from eventful contexts can be tricky as the difference is hard to define comprehensively and the clues in the text can be subtle. Consider the difference between the sentences below:

- The city is a far cry from the hotbed of riots that it was until the beginning of the decade.
- The city is a far cry from the hotbed of riots that it was during the last months of 2009.

The second sentence is very likely to be referring to a specific series of riots while, the first sentence implies that the city used to have a certain characteristic which made it more prone to be the scene of riots. The second sentence would make the article containing it a protest event article.

- ix. Events that are initiated by foreign governments, armies or any institution of a foreign state are outside our scope.
- x. Cases of military coups are not included in our protest definition. Cases of rebellion, mutiny, and civil war that are instigated by state officials (e.g. part or whole of military) are non-civilian actions and thus goes beyond contentious politics. In certain cases of civil wars, where state sovereignty is contested, initially non-state armed organizations can become state-like during the course of the conflict (e.g. ISIS). Such cases are also to be considered

non-protest events. The annotators should consult the domain experts about cases of civil wars where they find it difficult to decide.

- xi. Events of mass migration have been salient issues all over the globe, particularly in Middle East and Europe and Americas, in the last decades. Passage of masses of migration caravans through cities on their trajectory, camping at and attempting to cross international borders, and residing collectively in refugee camps might resemble at times contentious political events. Even though migration due to reasons of poverty, human rights abuses or war has aspects that challenge state authorities in designation and passage countries, we do not treat it as a contentious political event on its own. However, migrating masses have a tendency to engage in contentious political actions that we do include in our event ontology. We include actions such as marches, slogan shouting, carrying banners/posters, clashing with security forces, and group clashes that involve migrants in host countries as contentious politics events, but try to exclude migration itself while annotating.

#### ***2.1.5.2 Events that are included in the event definition but lack necessary characteristics***

- i. As stated above, the event described in the article must have taken place or be an ongoing event to be annotated. There are a number of cases in which the event in question is/was expected to take place but did not. These cases are outside our scope and will not be annotated.
  - a) The announcement or mention of events that are to take place in the future and hence reported in the future or conditional tenses are not to be included.
  - b) Sometimes the articles may mention events that are supposed to take place the same day the article is published, and hence report them with a different, stronger sense of certainty compared to what would characterize the news about future events. In these cases also, the norm in the news reports is to use future or conditional tenses, i.e. whether and how the events “will” or “would” take place. That is to say, the clue would again be in the tense used in the narrative.
  - c) Instances of vowing (an earnest promise to perform a specified act or behave in a certain manner) or threatening to carry out certain actions that are not carried out. The exception here is the situation mentioned in the 8. case in “Some event cases that are included” section whereby the threat of the action itself has the considerable effect of a protest, i.e. bomb threats, death threats and self-immolation or other suicide threats which also frequently become spectacles themselves.
  - d) Planned, threatened or intended actions that are thwarted or disrupted before happening such as a strike or demonstration that is disallowed by authorities, or apprehension of militants that are planning to carry out an action are not considered as events. The exception to this are cases of threats and attempts of violence, as it was mentioned in the previous section.



- e) Cases of bomb hoaxes and similar deceptions are not to be considered protest events.
  - f) Sometimes a news article will be about a judicial procedures such as prosecution or court cases about protest events. Care should be taken when deciding in these cases so that the articles which only contain protest mentions which do not refer to the actual events but to legal or procedural concepts are labeled as non-protest articles. An article in which the **only** protest mention is e.g. “cases of rioting and murder are registered against them”, is not a protest news article.
- ii. The second necessary aspect of the events to be annotated concerns their spatial characteristics.
- a) As it was mentioned in the necessary characteristics section, vague time expressions which nevertheless inform us that an event is a current affair are acceptable, but event location info is crucial to make the event mention sufficiently concrete. The events that do not get a specific location information attached are outside our scope. In cases where event place information (i.e. a specific geographical location more specific than the country) does not exist, a facility (building, campus, factory, etc.) provides the necessary concrete spatial dimension. Event mentions or summaries which do not come with any location information in terms either of geographical place or facility within the article are not to be annotated.
  - b) Events which gained historical significance and/or have become so iconic that they are mentioned with their specific name are a frequently encountered case of lack of location info. Gezi Events, 9/11 Attacks, Phoolan Devi Assassination are prominent examples. When these and similar events are not current and mentioned only by virtue of their historical significance, it is usually straightforward to exclude them. But when they are relatively recent, it would be harder to decide whether to include them or not. The location rule is a more certain guide in deciding in such cases.
  - c) Another case which might difficult to discern whether to include or not is when the news article contains a story about the survivors, victims or people affected by a certain event. In these cases, too, the articles are not annotated if the location information of the event whose victims are the focus of the story is not provided.

## 3 Sentence Level Protest Annotation Manual

This task aims at identifying and labeling sentence(s) that contain protest events in the news articles. It follows the document level protest labeling task which identifies news articles that contain protest events as defined in the DOLPAM. Once the news reports are classified as containing protest events, what remains is to identify where in the article the relevant event information is presented. In determining this location, the linguistic unit we evaluate is sentence. The sentences which contain event information will be labeled as event sentences. Those which do not contain event related information will be labeled as non-event sentences. Finally, those sentences which contain mentions of events which have not happened, i.e. are future events, will be labeled as planned event sentences. Read each sentence of the protest documents in the annotation tool and label accordingly.

### 3.1 Classifying Sentences

As an abstract procedural framework, the task can be divided into two phases. The first is to determine whether the sentence contain any event reference, thus making it eligible to be labeled as either 1 or 2. Once the non-event sentences are labeled as 0 and thus excluded, what remains is to decide whether the event took place in the past, is an ongoing event or has not taken place. The past and ongoing events are labeled as 1, the rest 2.

#### 3.1.1 Event sentences

Event sentences, must contain an expression which refers directly to any protest event that makes the document eligible for being classified as a protest article. Such reference can be any word or phrase which denotes the said event. They can be direct expressions of the event or the pronouns which stand for the event (e.g. protest, march, demonstration, strike, clash, incident, event etc.). These expressions are labeled as event triggers in the next annotation task. Thus, the detailed definition of event triggers is in the TOLPAM below so please refer to it for a better grasp of the concept. The event sentences must clearly indicate that the events in them have definitely happened in the past or are ongoing events. For additional information about this principle and our event definition, please refer to the DOLPAM.

#### 3.1.2 Non-event sentences

Non-event sentences, are the ones which does not contain any event reference past, present, and future. Note that they might contain event related information but this is not enough in and of itself to label them as event sentences insofar as they do not contain a direct reference to the event itself.

#### 3.1.3 Planned event sentences

Planned event sentences are those sentences which contain reference to events that have been planned, announced, threatened or attempted by a given actor but have not taken place for one

reason or another. They can be events that are expected to happen in the future, events that had been planned in the past but have not taken place, or events that any social agent threatens or declares to carry out. As it was mentioned in the previous manual, threats and attempts of violence are to be treated as protest events so they will be labeled as protest event sentences rather than planned.

## 3.2 Notes and examples

- i. The event sentences must contain at least one expression which denotes the event. These expressions can be:
  - a) Expressions which denote an action which is the whole or part of a protest event e.g. clash, strike, demonstrate, protest, attack, hit, blockade, shout slogans, sit-in, kidnap, assassinate etc.
  - b) Generic or dummy expressions which stand for the event e.g. incident, event etc.
  - c) Pronouns and demonstratives such as “it” and “this” when they stand for the events.
- ii. The expressions listed below are not to be considered event expressions that are defined above, that is, they do not make a sentence an event sentence by themselves:
  - a) Expressions which denote the participants or actors such as protesters, agitators, demonstrators, attackers etc.
  - b) Expressions which denote situations which cause or contain events, but neither are events themselves nor stand for events e.g. situation, tension, conflict etc. (Please refer to the DOLPAM for a detailed explanation of the difference between events and situations.)
  - c) Expressions which denote the consequences of events, e.g. death, injury, damage, captivity etc.
  - d) The word “protest” when it denotes the feeling or attitude of objection e.g. “There were murmurs of protest among the committee members after the meeting was postponed for the second time.”
- iii. The sentences which refer to the protest events in the abstract or refer to them in general are not event sentences. e.g. “Police forces were heavily deployed in the area to deter any such agitation” or “The inhabitants of the city are no strangers to violent clashes.”
- iv. The necessary characteristics of protest events such as presence of civilian actors, time and location necessities hold for event sentences as well, however, we do not expect the sentences themselves to satisfy them. If the event that is mentioned in the sentence satisfies the criteria in the overall document, the sentence is to be labeled 1. On the other hand, if the event mention in the sentence conforms to the criteria neither in the sentence not in the document, the sentence containing it will not be labeled 1.

- v. Be mindful of the sentences which contain event references but mean or imply overall that the said events did not take place, or were not protest events. These sentences should be labeled [0]. e.g. "The governor criticized the press for covering a petty street fight as a political clash." "The aid convoys are advised to use those border crossings where there are no clashes and disruptions."
- vi. Planned event references must be made to concrete possibilities. In other words, there must be a threat, an announcement or an attempt of an event. Perceptions of risks or threats are not to be labeled as planned events. e.g. "With the union representatives voicing their dissatisfaction with the negotiations, the possibility of a transportation strike looms large." "The embassies have warned their personnel to avoid crowded places due to the risk of a terrorist attack." "'We believe the organization was planning a major strike during the upcoming celebrations', another police official said."
- vii. Similar to the above point, the threats voiced by a given actor must be specific enough to warrant being labeled a planned event. Some threats to take an unspecified action might not involve protest events. e.g. "The party spokesperson threatened dire consequences if the ruling coalition goes forward with the motion" "Their leader vowed that they would retaliate in kind to any such action."
- viii. As explained in more detail in DOLPAM, remember that the threats or attempts of violent actions such as assassination, bombing, self-immolation, are to be considered as events themselves and therefore the sentences that contain them will be labeled 1.
- ix. The speeches given during a rally or a protest gathering are not protest events themselves.
- x. As in the previous document level annotation task, reports of events occurring in countries different than the source country will be annotated on this level.
- xi. While annotating news reports about migration events, take care to remember the explanation on the topic in the DOLPAM section, and try to distinguish contentious politics events that fall in our scope which occur during mass migration from the act of migration itself. Sentences which contain expressions like "caravans marching" should be considered carefully as to determine whether the march in question is the movement of migrants, or an actual protest march. There will be cases in which the two are intermingled and therefore not readily distinguishable. In such ambiguous cases, we prefer to err on the side of exclusion and not consider such actions protest events when there is no explicit mention of a protest action.
- xii. As it was mentioned in the DOLPAM, event references which are legal concepts and mentioned in the context of judicial processes about actual events are not to be considered protest events themselves and will not be considered event sentences. e.g. "The activists were charged with rioting, obstructing the police and damage to public property." In this sentence, the protest event mentions are not to actual events but their legal concepts, hence this is not an event sentence.

xiii. As it was mentioned in section 2.1.5.1. of DOLPAM as well, the valid event references which we label as protest events must be to actual events rather than general, contextual qualities of other story elements. Consider the example below.

“The organization has been known to have involved in bombings in Mumbai.”

“The organization has been known to have involved in the bombings that took place in Mumbai.”

The second sentence contains event information that makes which makes it an event sentence as “the bombings that took place in Mumbai” refers to specific bombing events. On the contrary, in the phrase “bombings in Mumbai” of the second sentence, such a specific reference is not present.

## 4 Token Level Protest Annotation Manual

### 4.1 Introduction

Linguistic annotation of the protest event related information in the news reports involves the task of labeling linguistic units (words or phrases) that gives such information within the text of the report. The third main task of the GLOCON GSC annotation process is this task of labeling the tokens comprising sub-sentence linguistic units that appear in the text as protest event related information. The documents that will be annotated on this level are annotated on the document and sentence levels, and thus are known to contain protest events. The task should start with the detection and labeling of event references (see terminology), which act as anchors of the whole annotation process on this level. Alongside event references, other words or phrases that include values or entities that give event-related information such as its time, place, participants, organizers, targets and others that will be enumerated below. These entities or values will be referred to as event arguments. Each relevant information type and the tag(s) designated to them for labeling will be described briefly. General rules about linguistic information extraction and specific rules about each tag will be provided in detail with many examples in a variety of languages and from countries in the scope of GLOCON database.

This manual aims at assuring consistency and clarity in annotation. Since it contains a lot of detailed information and rules about a multitude of variables, please keep in mind to come back to it for reference throughout all your annotation effort. While it is advisable to read the manual to get a general sense of the specifics of the annotation process, it is intended for consistent and repeated reference so do not feel obliged to learn or memorize all the rules contained in it. It is built on the event definition (or ontology) developed in the Document Level Protest Annotation Manual (DOLPAM). You should be familiar with this ontology in order to be able recognize the event related expressions in the documents so please refer to it before carrying on with this manual.

The annotation process of the linguistic information is performed in the “Annotation editor” of the annotation environment (FLAT) and consists of five foci; event, participant, target, organizer and semantic characteristics. These are named as focus and found under “Annotation Focus” in FLAT. The document opens by default in the “Annotation editor” and “Event” focus. The tag names are presented in lower case and multi-word tags are separated with underscore. These tags are presented with title case and spaces instead of underscores in the annotation environment. Read the manual carefully and report any issue you may observe or any suggestion you may have to the advisors.

## 4.2 Terminology

- **Event:** Events that fall in the scope of contentious politics, these are mainly protests and conflicts among groups. Refer to DOLPAM for a detailed description of what we consider as event.
- **Event reference:** Words or phrases that express an in-scope event in the text. Two types of event references are defined in this manual. “Event types” are first occurring event references that express the specific nature of the events, whereas “event mentions” are any subsequent words or phrases which express and/or stand for the event, including pronouns and demonstratives. We have a particular event definition (See the DOLPAM for further details on this definition) and only the news articles that conform to this definition should be annotated. If you think the news article does not contain an event as defined by DOLPAM, do not do any annotation on that news article. Simply mark the “Event” field in the Metadata Editor of the annotation environment, FoLiA Linguistic Annotation tool (FLAT), as “No” and notify the advisors.
- **Event argument:** Entities or values that are connected to protest events references and give information about them in event sentences.
- **Event sentence:** In the news article, “event sentences” are all the sentences which contain the event\_type and event\_mention information (Please see “Type of Event” section for further information about event\_type and event mention). They are generally located in the first paragraph of the news article and usually contain vital information such as event time, place and main participants etc.
- **FLAT-sentence:** The annotation environment may represent text fragments as partial sentences, single sentences, or several sentences in the same line, which is given a number on the environment. These lines are referred as FLAT-sentence.

## 4.3 General Rules of Token Level Annotation

### 4.3.1 How annotation proceeds

First read the whole of the news article you are going to annotate before starting the annotation. If the tagged expression occurs multiple times in the article, keep in mind that our priority is the information directly related with the event(s) in the article. We aim at annotating the most relevant entities to the event. This information is usually provided in the document title and/or in the event sentence(s) (see the terminology section for the definition of event sentence). Therefore, please follow the steps below to avoid unnecessary tagging or missing any relevant information:

- a) Identify the title of the news article and annotate all event-related information in the title.

- b) Identify the event sentence(s) and annotate all event-related information in those sentences. In cases where the tagged word is used more than once in the event sentence(s) then all those words should be tagged. E.g. “the problem rose in the village after a TDP supporter won the third ward of the village in the panchayat election held on august 14.” Both occurrences of “village” in the sentence should be tagged as `rural_location_identifier` (See the relevant section below for further details on how to use `rural_location_identifier` tags).
- c) There will be no annotation outside the event sentences, that is sentences which do not contain an `event_type` or `event_mention` tag will not be annotated.

Observe the following sentences as an example of this principle:

- “Samajwadi Party workers also held a demonstration at Jantar Mantar on Thursday morning demanding resignation of up chief minister Rajnath Singh and union home minister L. K. Advani for their negligence which allowed the assassination.
- “SP general secretary Usha Yadav said the party leaders also visited Phoolan Devi's residence at Ashoka road to express condolences to the bereaved family.”

In this example, only the first sentence is event related. The entities which are related to the protest (in the first sentence) such as SP, are not tagged in the second sentence because they occur in an irrelevant context.

- d) In case a tag is not used for an article, we will assume that this information was not provided in the article. Therefore, we rely on you to not miss any relevant information. Guessing may cause issues that will cost time and quality. In case you are not sure, you may prefer to give a lower confidence score to your annotation and add a comment which explains your doubts rather than leaving the information untagged. You will find the confidence level bar once you check the “confidence” box in the annotation window which pops up when you select a phrase to annotate it. Please do not skip explaining the reasons of the low confidence by adding a comment in order to avoid future confusion.

### 4.3.2 When to tag event arguments?

All event arguments listed in this manual will be labeled when they are syntactically and/or semantically connected with the event in a non-negative way. This principle can be elaborated in four ways:

- i. The immediate and most significant implication of this principle is that we will not code sentences which do not contain event references. Event information will only be tagged when they appear in event sentences, that is, when they appear as event arguments.



- ii. An entity or value must be connected to the event in the event sentence to be tagged as an argument. For instance a time value or place name that an event sentence contains must be given as the time or place of the event. Direct syntactic connection of this sort is usually straightforward. E.g. “The protesters gathered at the plaza in the early afternoon”. An indirect and semantic connection can also make certain entities and values event arguments. E.g. “Tuesday started as an ordinary weekday for thousands of residents, who were unaware of the chaos the afternoon’s attack would spark”. Here, even though there is no direct syntactic connection between the time value “Tuesday” and the event reference “attack”, the sentence establishes the semantic connection, making the former the time argument of the latter. That said, the connection between the event and argument must be unambiguous. Care should be taken not to infer relationships between events and entities that complex sentences might include but not connect. E.g. “A Delhi court found him guilty of perpetrating the attack that cost the lives of innocent civilians”. In this sentence there is no indication that the attack took place in “Delhi” and therefore it should not be tagged as event place.
- iii. The modality of certainty will not play a role on showing the relationship between events and arguments. This means that we annotate arguments even when the documents use qualifiers such as supposed(ly) or alleged(ly) before them. E.g. “the police have reportedly rounded up 11 members of the BKU for allegedly attacking one Kishan Singh” In this case even though the actor status of “11 members of the BKU” for the attack is “alleged”, we treat them as true event arguments and label accordingly.
- iv. Unlike the case of uncertainty modalities mentioned above, the even-argument connection will not be assumed in negative sentences. In sentences which establish a negative connection between events and arguments, we will not label negated entities as arguments e.g. “BKU leadership denied the organization’s involvement in the attack”. This sentence establishes a negative connection between BKU and the “attack” event. Thus, BKU is not taggable as the organizing actor of this event.

### 4.3.3 Overlapping Annotation

Sometimes certain expressions or parts of them give pieces of information which correspond to more than one tag. Overlapping annotation refers to tagging the same expression (or a part of the expression) with more than one tag. In the rules enumerated below, you will find the cases where overlapping annotation is required as well as cases where it is not allowed. Note that the rules are exhaustive, that is, overlapping annotation will be allowed for only the cases mentioned below.

- i. Tags on all kinds of event information in the document will overlap with the document\_title tag.
- ii. Participant type expressions which also relay information about other participant characteristics such as ideology, religion, ethnicity, caste and SES will be tagged with all applicable tags. e.g. “Teachers are determined to continue their strike unless their

demands are met.” [Teachers]=participant\_type, participant\_SES e.g. “The city witnessed a massive rally of thousands of angry Maoists today” [Maoists]=participant\_type, participant\_ideology

- iii. Organizer type or name expressions which contain information which are annotated with other tags in the organizer focus, i.e. organizer, ideology, religion, SES, caste and ethnicity. e.g. “Over 1.000 people were present in the Communist Party of India (Marxist) rally today” [Communist Party of India (Marxist)]=organizer\_name, [Communist]=organizer\_ideology, [Marxist]=Organizer\_ideology
- iv. Named entity expressions will only be annotated with “name” tags. There will be no “type” tags overlapping with facility, organizer and participant name expressions. e.g. [Safdarjung Hospital]=facility\_name (i.e. “Hospital” will not be tagged with facility\_type) e.g. [Karnataka State Government Employees Association]=organizer\_name (i.e. no other tag will be used in this expression.)
- v. Expressions which give different information which belong to more than one separate event, which is mentioned in the same sentence, will have all tags necessary. For instance, if the participants of event 1 are targets of event 2, overlapping annotation will be allowed for participant and target tags with their respective event number comments.
- vi. In situations where the facility of the protest action (i.e. the building or built environment which serves as its location) is at the same time its target (i.e. the entity against which the protest is taking place), both tags should be put accordingly. This situation usually occurs when the facility mention comprises premises of the entity (organization, government etc.) that the participants of the event protest against. Tag span rules of facility and target tags (see below their respective sections) will determine the tokens that are shared by two tags. For instance, the facility tags require their preceding prepositions to be included in the tag span, whereas target tags will never include prepositions. Hence, in the phrase “protesters pelted stones at the Office of Traffic Monitoring”, since the said “Office” is the named entity which is the antagonist of the protesters, it should have the target\_name tag. Also, since the expression “at the Office of Traffic Monitoring” comprises the facility information, the whole expression will be tagged with the facility\_name tag. In the final picture, the phrase “Office of Traffic Monitoring” will carry two overlapping tags, facility\_name and target\_name.

#### 4.3.4 Tag span rules:

##### 4.3.4.1 Tag span continuity

Even though the FLAT annotation environment allows non-continuous tag spans, all tag spans which cover any kind of event, entity or value must be continuous. In other words, there should be no tokens (including punctuation) that are untagged between the tokens of a tagged expression.

While annotating an expression that contains multiple tokens, make sure the tag span does not exceed the end of a FLAT-sentence as represented in FLAT. You can recognize these sentences as numbered on the left side of the document like rows in a table. If the phrase overflows to the next sentence, annotate the part of the phrase in the second sentence with a new tag, if that sentence is also an event sentence.

#### **4.3.4.2 Tag spans in complex expressions:**

As a rule of thumb, we must aim to include in the tag span as few words as possible. The more words are in a tag span, the less the likelihood of inter-annotator agreement, and thereby consistency. For many tags, coding single words for event references, entities and values is ideal. Adjective, adverbial, noun or prepositional qualifiers of the main entities are generally left out of the tag spans (exceptions are in tag descriptions below). Yet there are many instances where a single word will not indicate an entity in a meaningful way, and inclusion of more words will become necessary. While tagging such complex expressions, care should be taken to identify the smallest meaningful phrase which captures the entity in a meaningful way. The most general exception to the one word tag span principle is compounds. Compounds are single units of meaning that are formed by combining more than one word such as “someone”, “likewise”, “camcorder” etc. Some compounds, such as “floppy disk” and “ice cream”, however, are written separately and look and act like noun or adjective (or other) phrases. These might be difficult to discern from phrases which entail descriptions of qualified entities and are thus not single units of meaning as a whole. Consider “floppy disk” which designates a single entity, and is not a description (i.e. it does not mean “a disk that is floppy”). Common examples of compounds that we encounter often in protest annotation are phrasal verbs such as “flare up” and “walk out” or expressions such as “hunger strike”, “take to the streets”, “toddy tapping” and the like. Since compounds are single units of meaning, i.e. designate single entities, when they are tagged, the tag spans must cover all the words that make the compound (i.e. as if they were written in a conjoined way). On the contrary, modifier words in phrases must be left out of tag span as they do not change the nature of the concept that is tagged but merely describe it (e.g. school teachers, landmine blast etc.). To distinguish compounds from ordinary phrases you can keep in mind that the individual words which form them usually lose their original meanings once they are in a compound, that is, the compound means something different from each of its word-parts. You can also remember that they usually have their own definitions in dictionaries, unlike noun or adjective phrases. Below, under tag definitions, you will find tag span rules spelled out for certain complex expressions that require special treatment.

#### **4.3.4.3 Tagging punctuation and articles**

The punctuation marks, including quotation marks and parentheses, at the beginning and end of the expressions that are tagged should not be annotated. This principle does not apply to punctuation marks in the middle of the expressions tagged.

Indefinite articles (a, an) are not to be included in the tag when they occur before or after the expressions to be tagged. Definite articles (“the”) will only be tagged if it is part of the official name of an entity, e.g. “The Supreme Court”. Lowercase “the” will not be tagged.

## 4.4 Document Information Tags

### 4.4.1 Title of the document

The news article will have a title by default most probably at the top of the page. Titles can usually be distinguished by their linguistic characteristics—for instance, they are usually not complete sentences. Also, they tend to precede time and place the article published which are easier to distinguish. You should mark the document titles with the *document\_title* tag.

#### Tag rules

- i. If the document title is on top of the news article, mark it. This is the case most of the time.
- ii. Some articles may contain more than one news report independent of each other. Some of these reports may not include any protest events, in which case their titles will not be tagged. Mark only titles of the articles that include at least one event. You should insert a note in the comments section of the tag for the second and following titles in this case. The comments should start with “Event 2” for the second title and go on Event 3... if there are more titled reports.
- iii. In multiple report cases, the titles that follow the first one might not be as easily discernible as the first title, which is usually on the top. The flat environment does not contain font size or color variety that distinguishes the titles from the text body. Try to single out the titles from punctuation, inconsistent upper or lower case use etc.
- iv. We use this tag only to identify the news article titles as they were headlined by the newspapers. In case there is no title at the top of the article, do not search for it in between the lines. Simply notify the advisors about the missing title.
- v. The title generally contains key information about the event. You should first annotate the words or phrases that contain event information with relevant tags, and then annotate the entire title with the *document\_title* tag.

### 4.4.2 Document publication time

The publication date of the news article will be provided, commonly in the first flat sentence, as it is extracted from the HTML file of the news article. Please mark it with the *event\_time\_published* tag. Make sure that you mark the given expression in full, including the date, time and time zone -- if provided. Don’t forget to mark the punctuation marks (commas) within the time expression. e.g. “Sep 8, 2001, 23:34 IST”

### 4.4.3 Document publication place

The publication place of the article is also extracted from the HTML file and will commonly be located on the first flat sentence, near the document title and publication time. Publication places can be distinguished by virtue of the fact that they are written as stand-alone place names and do not form part of sentences or other phrases. The *event\_place\_published* tag will be used to label the publication place. In cases where there is more than one publication place, e.g. when the state is mentioned as well as the city, each place name must have its own *event\_place\_published* tag.

## 4.5 Event Reference Tags

### 4.5.1 Type of the Event

The event has a central role in our annotation efforts. You should first detect the event in a news article before you do any annotation. Events that can be included in the broad term contentious politics are in our scope (for the sake of simplicity, we use protest events to refer to contentious politics). Contentious politics refers to any form of grassroots political action, and actions of political and non-governmental organizations that are aimed at mobilizing the public in the name of political demands and grievances. Such actions are outside the institutional and legal forms of political participation such as voting, litigation etc. For details as to our event definition and various event types, refer to the DOLPAM.

Examples of protest events can be given as follows:

Demonstrations, rallies, marches, press declarations that take place outdoors in public, gatherings, meetings (when referring to public rally-like meetings), sit-ins, unofficial commemorations with mass participation (e.g. funerals or burial site visits), acts of civil disobedience and nonviolent resistance, collective petitions, collecting signatures, strikes and any other work related protest events (slow-downs pickets etc.), hunger strikes, death fasts, lynchings, lynch attempts, kidnappings, occupations, boycotts, cyber-attacks, hacktivism, hijacking, self-immolation, armed and unarmed violent attacks (by non-state actors), armed militancy actions against state security forces, unarmed clashes with security forces, clashes among groups of civilians caused by political issues, and the like.

Event types that are not among these examples should be evaluated based on their relation to definitions provided in the DOLPAM. Any doubt should be communicated to the experts.

The protest events that are detected in the news articles must be current or past events in order to be annotated. Events that are announced, threatened, reported to take place in the future in any other way will not be annotated. The only exception to this is when a threat of violence is issued by an actor, or a violent action attempt is thwarted, e.g. a death threat or a planted bomb that is diffused before going off. Self-harm threats and attempts are to be considered as violence attempt/threat, e.g. self-immolation attempt or threat. Such threats or attempts are thought to have the effect of events carried to completion and are thus regarded as events themselves. General or conceptual mentions to events, instead of specific happenings, will not be annotated either.

The type of the event should be annotated as per in the examples below:

- Thousands of people **rioted** in Port-au-Prince, Haiti over the weekend.
- The union began its **strike** on Monday.
- Protesters **rallied** on the White House lawn.
- The **rioting** crowd **broke windows** and **overturned cars**.
- A crowd of 1 million **demonstrated** Saturday in the capital, San'a, protesting against Israel, the United States and Arab leaders regarded as too soft on Israel.
- In Ramallah, around 500 people **took to the town's streets, chanting slogans** denouncing the summit and calling on Palestinian Authority leader Yasser Arafat not to take part in it.
- For weeks Italian Jewish groups, World War II veterans and leftist political parties have staged **protests** against a meeting between the pope and Haider, arguing that a papal encounter would lend the Austrian politician legitimacy.
- More than 40,000 workers were back at their jobs Thursday following a 1-day **walkout** that closed social welfare offices and crippled public medical services. During the work **stoppage** Wednesday, local residents were unable to register marriages or get documents for real estate transactions.
- Shah 's supporters also had gone on the **rampage** outside the principal 's office on Tuesday.

## Tag rules

- Tag span rules:
  - Adjective, adverbial, prepositional, noun or any other qualifier word will not be included in the span. e.g. landmine **blast**, token **strike**, forcibly **stopped**, communal **flare-up**, indefinite **hunger strike**, fierce **encounters**, twin militant **attacks**, protest **demonstration**, protest **march**.
  - In verb event types, only the verb will be in the span. e.g. **boycotted** a meeting, **rioting** with weapons, forcibly **stopped** the auto, fight it with **bombs**, landmine **exploded**, hit by crude **bombs**, refused to work extra hours.
  - Auxiliary verbs will not be included, except when the verbs are in negative form e.g. have **blocked off**, will **protest**, are **demonstrating**, **did not start work**,
  - In noun event types, only the noun will be in the span e.g. laid an **ambush**, forced a **bandh**.
  - If the verb or noun itself is not meaningful or is dubious as a protest action, include the minimum required words. Since tag spans cannot be broken, this might necessitate inclusion of non event type words. e.g. **Violent reaction**, **violent culmination**, **took to**

**the town's streets, did not start work, refused to work, shouted slogans, written slogans, hung posters, carried banners, hurled stones.**

- ii. If the article contains multiple events, mark all. If there are multiple events in a news story, the events that have different times, places, facilities and participants from other events should be annotated separately. The rules about event separation are a bit complex so please refer to Event Separation section in this manual.
- iii. Demonstrations, gatherings, electoral rallies, etc. organized by political parties, organizations, or representatives of political parties should be included. However, events happening under the auspices of the parliament such as politicians leaving parliamentary sessions, disrupting them or having heated arguments with each other are accepted as usual proceedings within parliaments and thereby not considered as events. BUT if politicians engage in a protest action which is unusual and outside the formal parliamentary procedures and proceedings, (e.g. such as hunger strikes, sit-ins) these will be considered as protest events.
- iv. Spontaneous gatherings or demonstrations with no organizer or leader should be annotated as events.
- v. Be aware that Hindu rituals such as idol immersions and akhand paths are not protest events in themselves. However they can become means or scenes of protest events. In an expression like "**Clashes** broke out after idol immersion", only "clashes" should be annotated as event type but in a sentence like "An **akhand path** was organized to protest government's decision", this ritual is deployed as a protest itself.
- vi. Public interest litigations and writ petitions are to be tagged as event types in case they have political motivations, and are participated by multiple members of the public.
- vii. Expressions which sometimes refer to consequences of violent events, such as "kill" "injure" and the like, will be tagged only in cases when they refer to the action of killing or injuring, intentionally. Bear in mind that in English "Three people were killed in the attack" is identical to "Three people died in the attack" so in both sentences only "attack" is the event type.
- viii. The word "protest" is a very versatile word which can sometimes mean to be in or declare opposition to something, or a strongly worded refusal or rebuttal. Be mindful about such situations where the word protest does not stand for events that are in our scope.
- ix. Similarly, the word protest will not be tagged when it means intention or aim rather than the act of protest itself, as in expressions such as "They held a **rally** in protest against the management". In a sentence like "They **shouted slogans**, protesting against the tariffs", only "shouted slogans" will be tagged as event type.

## 4.5.2 Event Mention

Any word that refer to the event type which occurs after the first occurrence of the event\_type will be tagged with the event\_mention tag. Pronouns and token words which refer to the event type are included in the tag's scope. Token event words are generic expressions which denote the event (i.e. words like incident, event, protest and agitation).

### Tag rules

- i. For every separate event, there will only be one event\_type tag which will be the first occurrence of the event word, provided it is not a token event word (see below). All the following event words and pronouns that stand for the event type will be tagged event\_mention.
- ii. Event words in the document title will be tagged with event\_type tag. The first occurrence of the event word outside the document title will also be tagged with event\_type tag.
- iii. Token event words which stand for events, will be tagged with event\_mention tag even if they are the first occurring references to events (see the examples below). In principle, you can distinguish a token event word from a true event type word by the fact that the former, on its own, means a broad class of protest actions which can correspond to multiple categories of events defined in the DOLPAM, namely, demonstrations, industrial actions, group clashes and political violence and militancy. The words like incident, event, protest and agitation can stand for events from all (or more than one) of these categories. Such words will always be tagged event\_mention. The only exception to this is the case when token words are the only event words that are used in the article to refer to the events. In such a case, the general rule will hold and the first occurrence of the token word will be tagged event\_type and the following references will be tagged event\_mention.
- iv. Event separation logic will be the same as any other tag. That is, any event mention tag that is used for each separate event will be numbered in a comment.
- v. The sentences which contain the event\_mention will become event sentences, thus any relevant event information in these sentences will be tagged even if the event\_type tag occurs in other sentences.

### Examples

CPI(M) stages protest **rally** in Bhavnagar. The Bhavnagar unit of communist party of India CPI(m) on Friday staged a **demonstration** opposite the local post office here...

–The word “rally”, the first occurrence of the event word, is tagged event\_type, demonstration which refers to this event, is to be tagged with event\_mention instead of event\_type.

Commenting on the **strike** which was flagged off on Monday, the union secretary stated “**it** will continue as long as our demands are not met.”

–The word “it” stands for the event\_type “strike” and thus is tagged with event\_mention.



In a separate **incident** in the neighboring area of Awantipore, an air-force vehicle was damaged and two security force personnel were injured when a landmine **exploded**.

–Here, the main event word is “exploded” and incident

The students organized a **protest** by **marching** against the payment seat decision.

–Here the token event word “protest” is tagged event\_mention as any contentious politics event can be referred to as protest while the word “march” is tagged event\_type as it directly indicates what the protest is, and thus is a true event type.

A variety of NGO’s expressed their support for the **agitation** of teachers in the capital.

–Considering this sentence is the only one which mentions the event in the article, the token word agitation is tagged event\_type.

### 4.5.3 Semantic category of events

The “Semantic” focus in the flat environment contains the tags to be used when labeling the event triggers (types or mentions) according to the protest type that they indicate. The semantic tags will be placed on each event trigger in an overlapping manner. That is to say, the tag spans of the semantic category tags will be coterminous with the tag spans of event type and event mention tags. The tags and their definitions are as follows:

#### 4.5.3.1 Demonstrations

A demonstration is defined as a form of political action in which a demand and/or grievance is raised outside the given institutionalised forms of political participation in a country. The aim of the event is to draw attention of politicians and/or general public to said goal by making it visible in public sphere which implies any space which is open to members of the public. Violent or peaceful forms of social protest that take place in open or closed public spaces such as streets, plazas, vicinity of prominent buildings fall in the category of demonstrations. Notable examples of demonstrations are as follows: Protest marches, silent marches, rallies, demonstrations, outdoor press declarations, gatherings, sit-ins, acts of civil disobedience, dharnas, bandhs, demonstrators clashes with security forces, commemorations and religious rituals (when they become means of protest, e.g. akhand paths), collective petitions (collecting signatures), self-immolation (and threats of self-immolation), hunger strikes, fasts, barricading, picketing (road blocks), burning of vehicles, occupations of public buildings that are not included in the industrial action category and the like.,

#### 4.5.3.2 Industrial actions

Industrial actions are types of protest events that take place within workplaces and/or involve the production process in the protest. Notable examples include any kind of strike (comprising slowdown strikes, wildcat strikes, sympathy strikes and green bans), workplace occupations, boycotts, picket lines and gheraos. Note that actions which take place in public streets, roads which are otherwise become scenes of demonstrations, the previous category defined above, might become industrial actions when these spaces are actual workplaces of certain groups of

workers which carry out the protest. An example to this would be public transportation workers carrying out a protest by blocking roads. While a road block would otherwise fall in the demonstration category, here it becomes a group of workers disrupting the work process by withdrawing their services and somehow “occupying” their workplace. On the contrary, marching of a group of workers on the streets in order to draw attention to their work related demands must be considered a “demonstration” rather than an industrial action. Similarly, occupations or disruptions of public institutions that are not worker protests, e.g. when university students occupy education facilities, are to be considered demonstrations, and not industrial actions.

#### **4.5.3.3 Group clashes**

Group clash events are instances of confrontation that stems from politicized conflicts (e.g. identity or economic interest based or ideological conflicts) between social groups such as fighting, lynching, ransacking, arson, any armed or unarmed clash between civilians of a political nature. Actions that target religious, ethnic or similar minorities fall within this category. Said actions can be unidirectional, that is, the target of the event might or might not retaliate during the event. Group clashes that take place in open public spaces might involve or escalate from actions which are defined in the category demonstration. The key to identifying group clashes would then be the actors involved. The distinguishing feature of group clashes is that both the participants (or organizers) and the targets of these events are non-state actors. While in violent demonstrations, the targets would be state institutions, state officials and the like, group clashes target private individuals, businesses, religious buildings etc. Also note that group clashes can be violent events which might involve the use of firearms, making these events somewhat hard to distinguish from armed militancy events described in the next category. Again, both participants and targets being non-state actors is a clue to identify group clashes.

#### **4.5.3.4 Armed militancy**

Politically motivated violent actions that fall within our event definition are included in this category. These actions are usually carried out by political and/or militant organizations which systematically resort to violence as a means of reaching their goals. Their targets can be officials or civilians. Notable examples of this type of events are kidnapping, assassination, bombing, suicide bombing, hijacking and the like. A feature of these events that distinguish them from group clashes is the nature of perpetrators. Violent actions that target civilians will be included in this category if the perpetrators are armed militant organizations. Also, even though such actions are generally carried by organizations, there might be cases in which such acts are carried by individuals not necessarily affiliated to organizations or groups. Hence, if the act which intentionally resorts to violence carrying political goals, they are included in our definition even when they are carried out by a proverbial “lone wolf”.

#### **4.5.3.5 Electoral politics**

Electoral politics events are rallies, marches or any similar mass mobilization event that are organized within the scope of election campaigns of political parties or leaders. The aim of these

mobilizations are garnering support or votes for the political entity competing in the upcoming elections. Note that demonstrations or protests that take place during the elections, such as boycotts, clashes, sabotaging voting etc. will **not** be considered as electoral politics and will be tagged with the corresponding semantic tag other than electoral politics.

#### 4.5.3.6 Other

Other tag will be used for any event which does not fit in one of the categories listed above.

## 4.6 Event's Time and Location Arguments

### 4.6.1 Time of the event

Any information which gives information about the time the event takes place will be annotated with the *event\_time* tag. The date information such as the year, month, or day; information about the time of the day as in "in the morning", at 3 p.m., "late at night"; less definite expressions such as "three weeks ago", "this year", "recent" are all valid event time expressions to be tagged. Keep in mind that the time of the event might be expressed more than once in different ways. All such expressions must be labeled as long as they denote the time of the event.

#### Tag rules

- i. Tag span rules:
  - a) Time expressions will be tagged in their entirety, including the prepositions which indicate the time information as a whole. E.g. "at 3 pm", "before noon", "at the early hours of the day", "three weeks later", "in last September" etc.
  - b) Complex time expressions that contain multiple values but indicate a single time information will not be divided. e.g. "at 6 p.m. yesterday" "June last year". In this vein, expressions which indicate the day of the month will be included in the tag span. e.g. "Monday (15)"
- ii. Indirect and relative time values will **not** be tagged as event time. A case of this is the expressions regarding the frequency of recurrent events such as "75<sup>th</sup> anniversary" which are not be tagged as event time. Time expressions which are related with other events or developments such as "after the meeting", "before the prime ministers visit" are also not to be considered as time information and will not be tagged as event time.
- iii. Duration expressions will **not** be tagged as event time information, e.g. "after 102 hours of hunger strike", "the strike ended in its fourteenth day".
- iv. The expressions which give information about the beginning and the end of the events which take place over long periods, such as strikes or hunger strikes, will be tagged as event time. e.g. "The strike ended on Monday"

## 4.6.2 Place of the event

Event places are official geographical or administrative divisions in which protest events take place. Each country will have its own system of categorizing its administrative and geographical divisions. States, provinces, cities, districts, neighborhoods, villages, taluks, and/or panchayats can be counted as examples of these divisions. The given official place names of these divisions will be tagged with the event\_place tag. A division might have its own local administration such as municipalities in cities or governors in states. Some divisions however will not correspond to an administrative unit but will simply indicate a geographical definition, such as districts, neighborhoods or regions in some countries. The requirement for the label of event\_place is official recognition and/or naming. Names of built environments such as streets and plazas, buildings etc. are not to be considered as place names; they have their own tag defined below.

### Tag rules

- i. Tag span rules:
  - a) Only the Place name will be tagged. Generic settlement categories (village, city etc.) and prepositions will **not** be included e.g. Chapadu mandal, near Charar-e-Sharief, in Bihar's impoverished Aurangabad district, Gecek köyü. Nouns such as town, village etc. in these expressions will be tagged as location identifiers (See the section "Centrality").
  - b) Direction words like east west will be included only when they are parts of the official names e.g. Imphal west district, Salem West district, West Bengal.
  - c) The conjunction "and" will only be included if it is part of the official name of the place, if it is used to connect different places, it will not be a part of the span and places it connects will have separate tags (this rule is valid for all the tags) e.g. Jammu and Kashmir, Karnataka and Gujarat states.
  - d) Prepositions that are preceding a place name should **not** be included in event\_place annotation. e.g. "Tens of thousands of workers went on strike in Hong Kong"
  - e) Expressions like "city" or "center" might be included in official place names; in which case they must be included in the event place tag span. The way to discern whether such expressions are part of official place names is looking at whether they are written capitalized or not. We will assume that capitalized expressions like the "Central district of Hong Kong" or "Ciudad Autonoma de Buenos Aires" (commonly abbreviated as CABA, sometimes referred simply as "la Ciudad") are official event place names, and will tag them accordingly. Other instances where the expressions "city" and "center" are not capitalized, they will be excluded from tag spans.
- ii. If the article mentions more than one place that the event takes place in, all place names must be tagged. This happens in two situations. The different scales of locality which have different names might be mentioned, as in "Salem city in Tamil Nadu state". Also, events which start at one place and finish at another **without interruption** such as "Teachers

marched through **Kherwadi**, **Bandra West** and **Pali Hill** neighborhoods.” Keep in mind that in other cases, events that take place at multiple, different places are to be treated as different events (see the next rule).

- iii. If the article mentions separate events from multiple places, mark all events and the event related places and note the relevant event number in the comments section (see the Event Separation Section).
- iv. If a protest news report about a country which is in our focus contains protest events in another country, all the information about them will also be labelled. However, country names will **never** be tagged as event places.

### 4.6.3 The facility in which the event takes place

Facilities are human made structures or built environments that serve as the location the events take place in. These facilities can be streets, factories, airports, universities, vicinity of prominent buildings or sites (e.g. in front of a party office or an organization) etc. Nouns which indicate the class of facility that the event takes place in will be tagged with the `facility_type` tag, e.g. “**on the streets**, **in front of the court**, **towards the building**” etc. The unique, named facility entities will be tagged with `facility_name` tag, as in “**Safdarjung Hospital**”, “**The Supreme Court**”, “**INTUC headquarters**” etc.. Note that all facility expressions will be tagged with the corresponding prepositions and qualifiers which indicate the location information in full. An event taking place “at” a certain place has a different facility from another one which takes place “in” it. E.g. “the strike had crippled all services **in the hospital**”, “the strikers gathered **in front of the hospital**”.

#### Examples

- Owners of contract buses, auto-rickshaws and taxis stayed **off the roads**.
- Shah 's supporters also had gone on the rampage **outside the principal 's office** on Tuesday .
- The strike had crippled all services **in the hospital** and most patients had been discharged.
- Two individuals died in the bomb explosions organized by radicals **in places of worship**
- Services **in government hospitals** were paralyzed due to the strike.

#### Tag rules

- i. Tag span rule: The prepositions and other qualifiers which designate the location of the event in full, will be included in the tag span e.g. **on the Osmania University campus**, **at Cusrow Wadia Institute of Technology**, **into the university campus**.
- ii. Like the event place, an event might take place in multiple facilities. As long as the event is continuous or the multiple facility references are of different scales, as in “**in front of the bus stop**, **on the main street**”, all off them will be annotated separately with facility tags. In

other cases, different facility references indicates different events which must be separated as per described in the Event Separation section.

- iii. Take care in distinguishing between the targets and facilities of events. In cases where buildings or built environments are attacked, damaged or targeted by the protesters in any way, they will be annotated with facility tags. Refer to the Target Characteristics section for a more detailed elaboration on this.
- iv. In industrial action events such as strikes, the facilities will be the workplaces in which the strike is taking place.
- v. Capitalization might indicate unique, named entities that are to be tagged with facility\_name tag, but do not rely solely on this. There are times when different news sources do not apply this principle consistently, also text extraction processes might change character cases. Instead, try to decide whether to use the name or type tag on the basis of the facility that the article mentions being unique or not. If you cannot decide for certain, use facility\_type tag. Facilities such as municipality administration buildings, governors offices and the like are commonly unique.

#### 4.6.4 Centrality of the event location (Urban or rural)

The linguistic information that relays information about the centrality of the place in which the event takes place will be labeled with urban and rural location identifier tags. The expressions such as “town”, “city”, “village”, “panchayat”, “hamlet”, or “township” (South Africa), which directly indicate this will be tagged. In addition, expressions which identify built environments (human made surroundings as settings for human activity) such as plaza, avenue, slum, (village) well, that are invariably associated with urban or rural settings will be tagged with the corresponding location identifier tag.

##### Tag rules

- i. The place name itself is not an urban or rural identifier. Nouns such as village, city, town which follow place names will be tagged with urban or rural location identifier tags, as in Matahalli **village**, Vadadora **city**, Rocinha **favela**, Gecek **Köyü** etc.
- ii. Plaza, avenue, slum, favela are typical expressions where urban location identifier tags can be used. Be careful to add more to this list and do not tag other entities with location identifier tags without consulting with the annotation supervisors. Types of buildings, institutions or types of economic activity, such as hospitals, farms, universities etc. will **not** be tagged with location identifier tags.
- iii. If expressions such as avenue or plaza are facilities of protest events and thus tagged as facilities, they will not be tagged with location identifier tags. In other words, there will not be overlapping location identifier tags and facility tags will have priority.

## 4.7 Event Actor Arguments

### 4.7.1 Participant Arguments

#### 4.7.1.1 Types of participants

The participants are the main actors of a protest event. They can be individuals or groups that engage in a protest. Groups may be referred to as masses (common people, public), armed or violent political actors (attackers, militants, militias, guerrillas), unarmed non-violent political actors (activists), blue-collar workers (factory workers, low skilled workers) white-collar workers/professionals (lawyers, doctors, nurses, corporate employees, teachers, engineers), the unemployed, students, women, LGBTI Individuals, peasants (including landless peasants and sharecroppers), villagers, war veterans and their relatives, politicians (members of the parliament, political leaders, party members), academicians, journalists, public figures, artists, intellectuals, artisans and shopkeepers, the disabled (physically challenged), prisoners and their relatives, refugees/immigrants, consumers, hackers or hacktivists, ultras, sports spectators, supporters, relatives of political violence victims, prisoners, kidnapped, disappeared or assassinated individuals, and the like.

#### Examples

Teaching and non-teaching staff went on strike on Tuesday.

Hundreds of karamcharis staged a dharna in Delhi.

Skirmishes occurred between the Bharatiya Kisan Union (BKU)-EKTA activists and the police at various places in the district, on Monday.

Senior Congress leaders including Kamal Nath, Oscar Fernandes, Narendra Nath, DPCC chief Subhash Chopra and party legislators participated in the demonstration.

More than 800 agitators participated in the protest.

#### Tag rules

- i. Tag span rules
  - a) Fewest possible words which indicate the actual participants meaningfully. No qualifiers will be included in the tag span e.g. rival party supporters, resident doctors, Muslim women, jobless diamond workers, secondary school teachers, inter-city bus drivers, angry karamcharis.
  - b) As an exception to the previous rule, some qualifiers which modify certain occupational groups change the nature of those occupations so significantly that they become a different category of employees with different socioeconomic status and/or organizational characteristics. In these instances the qualifiers should be included in the tag. It is not straightforward to determine these occupations as they require certain unique country characteristics to be taken into account. When a likely case is

encountered, don't hesitate to ask the project experts about the coverage of the tag spans. Prominent examples are as follows: **auto-rickshaw drivers**, **college teachers**, **safai karamcharis**. Note that the words, drivers, teachers and karamcharis are the only words to be included in the tag spans in other instances, see the previous note.

- ii. Words like mob, group, crowd, will be tagged only if they are the only words standing for participants. When they are parts of phrases with more definite participant words, they will not be included e.g. a group of about 100 **people**, group of **locals**, group of **militants**, group of **students**, the angry **mob**, the **crowds** gathered in front of the supreme court.
- iii. Named entities or expressions referring to a specific group of people may be expressed with different expressions after their first occurrence. The different occurrences should be annotated with the tag(s) used in their first occurrence if the alternative expression also denotes the same class of entity (event type, participant type, sector etc.). E.g. If a tagged expression such as "teachers' strike", where the word "teachers" is tagged with participant\_type and participant\_SES, is followed by a phrase such as "agitators want a higher salary" where "agitators" refers to teachers, the word "agitators" is to be tagged as participant\_type only as the word "agitator" denotes a type of participant but denotes neither the sector nor the SES.
- iv. Groups of people, regardless of whether they are organizers, participants, or targets may be expressed with pejorative terms, e.g., "miscreants". This kind of terms should be annotated with the tag corresponding to their role in the event, that is, organizers, participants or targets.

#### 4.7.1.2 Names of participants

The name of the participant(s) may be mentioned. This can be a first name and/or last name of an individual who is engaging in the protest. Please use the tag participant\_name for annotating these names. Titles such as Mr., Ms. or Dr. will not be included in the tag span.

##### Example

- Senior congress leaders including **Kamal Nath**, **Oscar Fernandes**, **Narendra Nath**, DPCC chief **Subhash Chopra** and party legislators participated in the demonstration.

#### 4.7.1.3 Participant Count

Any expression that indicates the total quantity of individuals who participated in the event should be annotated with the tag participant\_count.

##### Examples

- **a large number** of people, many people
- **hundreds** of BJP workers, four agitators
- as many as **50** employees gathered at the DC's office



- more than 800 agitators participated in the protest
- about 700 members of the BKU are sitting on a dharna.

## Tag rules

- i. Tag span rules:
  - a) In Expressions, such as “more than a hundred demonstrators”, “nearly fifty assailants”, “as many as five thousand workers”, and “about twenty militants”, that give an estimation of the number of participants, only the number expressions will be tagged; that is the estimation qualifiers will not be included in the tag span.
  - b) Complex participant count expressions which contain multiple values, but indicate a single participant quantity information will not be divided. e.g. “between 800 and 1000 people marched in the streets”
- ii. Sometimes quantity expressions refer to a part of all participants (e.g. the police arrested 11 of the protesters). These expressions should not be tagged with `participant_count` as they do not give information on the total quantity participants.

### 4.7.1.4 Ideology of the Participants

In case an explicit mention of participant ideology is mentioned, this is tagged as `participant_ideology`. The participant ideology can be communist, socialist, anarchist, social democrat, left-wing, right-wing, feminists, liberal, nationalist, religious fundamentalist, conservative, far right, green or environmentalist, animal rights, pro-LGBTI, and any expression similar which identifies the ideology or world view of participants. Please tag any word/phrase that indicates the ideology of the participants.

#### Examples

- Left-wing militants blocked the road in Caracas.
- Maoists staged a rally in the city center.
- The city was the scene of massive anti-government demonstrations.

### 4.7.1.5 Ethnic identity of participants

Expressions which identify the ethnic and racial identity of participants are marked with `participant_ethnicity`.

#### Examples

- Catalan protesters took to the streets of Barcelona.
- Two of the activists who participated in the event were identified as Tamil militants

### 4.7.1.6 Religious Identity of Participants

Any expression that identifies the religious affiliation of participants are tagged with `participant_religion`.

## Examples

- Some **Catholics** are protesting what they regard as a blasphemous stage production that disrespects God and the Catholic faith.
- Nationalistically aroused **Sikhs** are leading demonstrations to ban the sale of cigarettes since the use of tobacco is forbidden to Sikhs.
- **Muslim** workers protested against the ban of Friday praying in factories in Mumbai.

### 4.7.1.7 Caste of the Participants

Expressions which designate the caste of participants will be annotated with participant\_caste tag. This tag is specific for annotation of news from India.

#### Example

They also explained the aggression of the nayaks on harijans on every trivial issue.

### 4.7.1.8 Socioeconomic Status of the Participants

The expressions which give information as to the social standing or class of participants will be tagged with participant\_SES. Any word or phrase that indicate the class position, socioeconomic status, or income or education level of participants should be annotated.

#### Tag rules

- i. The expressions which designate, if indirectly, the socioeconomic status of participants, they will be tagged. Examples can be “starvation”, “malnutrition”, “slums”, gated community etc.
- ii. If the participant type is a category implying participant SES (like it is the case for teachers, doctors, workers) then mark them with both participant type and participant SES tags. The words “politicians”, “activists”, “students”, “housewives” do not imply any SES information and hence will not be tagged participant\_SES.

## 4.7.2 Semantic Categories of Participants

The research agenda of the EMW project involves examining social groups and classes which engage in contentious politics actions, as how governments respond to protest events has important determinants in who engages in protest action. Thus, sociological characteristics of participants of contentious politics events are as important as characteristics of those events themselves. Just like protest events, the GLOCON database aims to classify protest participants into sub-types. The semantic categorization of participant actor expressions serves this task. Since participant names give little linguistic information that can be classified, semantic category tags will be put on expressions that are labeled with participant\_type tag.

#### Tag rules

- i. Every expression that is tagged with participant\_type, must also have a participant semantic category tag.

- ii. Semantic label tag spans will be coterminous with those of the participant\_type tags that they are used in an overlapping manner.
- iii. Participant semantic tags will be consistent for each expression that refers to the same group in different event sentences throughout the text. In case, for instance, a group of “teachers” who go on strike is referred to as “strikers”, “protesters”, or “agitators” in different sentences, every such expression will be labeled with the same semantic tag (“workers” in this example).

The tags for participant semantic categories are as follows:

#### **4.7.2.1 Peasant**

People who work in agriculture and/or live in rural areas. The category includes small holders, that is subsistence level farmers who own small plots and use mainly domestic labor.

#### **4.7.2.2 Worker**

Any kind of blue or white collar employee in private sector companies or public institutions. This category should exclude those who are classified under “professionals”.

#### **4.7.2.3 Small producer**

Owners of small shops, small traders, artisans, or any small business which employs family labor (includes transport owners like 4 owner taxi drivers).

#### **4.7.2.4 Employer/executive**

Owners and managers of medium and large size businesses which employ waged labor.

#### **4.7.2.5 Professional**

University educated professionals who are self employed, or work in private or public sector (physicians, lawyers, academics, journalists etc.).

#### **4.7.2.6 Student**

Students from all levels of education.

#### **4.7.2.7 Politician**

Members of political parties who hold office in legislative organs and/or executive branches of government. (There may be rare cases in which politicians are independent, i.e. not affiliated with political parties. Party membership, although very common, is not mandatory.)

#### **4.7.2.8 Activist**

Ordinary members of political parties, and members of grassroots organizations and NGOs. In distinguishing activists from ordinary people who engage in protest, the key here is to look at whether in the news report, participants are organized under a political organization. In most cases activists will be mentioned together with the NGO or party that they belong to. In the absence of

such an explicit mention of an organization, expressions that denote activism will be used (e.g. environmentalist, activist, feminist, human rights advocate etc.). Note also that political party members who hold offices in governments and parliaments should be labeled as “politician” rather than “activist”.

#### **4.7.2.9 Militant**

Members of political organizations which are commonly illegal and resort to armed violence (e.g. Islamic fundamentalist militants, members of armed revolutionary organizations etc.). In some cases distinguishing militants and activists will be difficult. There will be events where certain members of ordinary, legal political parties will resort to armed violence. On the other hand, in certain events, you will see members of militant organizations engage in peaceful forms of protest. The way to decide in such situations is to consider the character of the political organization to which participants of these events are affiliated to. Thus, we will decide on the basis of how the organization is characterized in press, rather than the type of the event.

#### **4.7.2.10 People**

General categories which refer to citizens without organizational or employment ties. Examples include women, residents, religious or ethnic community members, etc. Non-specific expressions which are labeled as participant type\_such as mob, crowds etc. will also be labeled as people.

#### **4.7.2.11 Other**

Any category of participants which cannot be placed into categories above. This can be due to the fact that the given participant expression is not covered in any of the categories above, or it is unclear to which type of protester is the expression referring.

### **4.7.3 Organizer Arguments**

Organizer is another actor category which engages in contentious politics. Organizers organize and/or lead the event and participants. Some events do not have organizers because they are spontaneous events, in which participants take action without being led by anyone. In other cases, an organization, institution, a group of people or single individuals might organize or lead the protest. Any organization which takes part in the protest and individuals who lead or organize the event will be tagged with organizer tags. Individuals can be organizer only in cases in which it is explicitly stated that they lead or organize the event.

#### **4.7.3.1 Distinguishing between organizers and participants**

The difference between organizer and participant of an event is very important. First, the same participant group can join two events organized by two different organizations and the result may be totally different. For example, workers can join a left-wing rally organized by unions or a right-wing lynching against migrants, organized by a racist organization. Second, some events are organized by some organizations or it can happen spontaneously without any organizing effort. Theoretically, events have one of the following organizer-based characteristics:

- ➔ There is not any explicit organizer, spontaneous events
- ➔ There are organizers and participants, coordinated events. Unions leading a strike, for example.
- ➔ There are organizers, but there are not any participants: This is not very common. We see some parties/organizations coming together to organize an event, in which no one participated.
- ➔ There are no organizers and participants mentioned. This could be an explosion, for example, where no one was identified as responsible.

Some examples that clarifies this distinction are as follows:

- In case there is a boycott at a university campus, the organizer is the student union and the participants are the students.
- The potential organizer of a strike is a labor union and the participants are the employees.
- Some events are reported as led by a particular person, in such a case the person is the organizer, e.g. “the protesters led by Shelia Dikshit”, where “Sheila Dikshit” is the organizer\_name. If the person is not clearly specified as leading the event, it should be annotated with the appropriate participant tags. For instance, an assassin is still a participant of an assassination event.
- A phrase like “association members” should be tagged as “association”: organizer\_type and “members”: participant\_type.
- Some cases mention names of an organization and its leader, such as “BJP leader Madan Lal Khurana , architect of many anti-government protests , lent his voice , and a broad smile .”. These cases, in case they are in the scope of an event, should be annotated as “BJP”: organizer name, “Madan Lal Khrana”: participant\_name.
- Some organizations or people express their support to an event. These entities should not be tagged with any tag in case there is not any indication of organization or participation effort. So, sole support is not enough to be annotated.

#### **4.7.3.2 The event organizer type**

The organizer(s) of an event can be Political party, Non-governmental organization, associations, charities, Labor union, Non-party political organizations (student organization, women’s organizations etc.), Anti-state armed groups such as terrorist organizations, resistance movements, guerrilla groups, pro-state armed groups such as paramilitary organizations, Social communities (tribes, religious groups), Local public authority (municipalities, governorates etc...) or Multiple organizations (more than one party, NGO and other type of organization coordinated the event). Please tag any type of the organizer with organizer\_type tag.

## Examples

- Left parties organized a silent march.
- A prominent banned religious outfit, Deendar Anjuman, exploded a mine on Sunday

## Tag Rules

- i. Tag span rule: Like any other tag span, the shortest possible type expression will be tagged, i.e. words like association, party, union, employees association. If the word association stands for a workers union, which is generally the case in India, the word before the association will be included, i.e. employees association.
- ii. Business entities, companies and firms, are civilian actors which can engage in conflict with other civilian actors, be it their workers or residents of the localities where their plants are located. Most often, companies are actors that are at the receiving end of protest events, i.e. they are the targets of protest events engaged in by their adversaries mentioned above. However, in certain conflictual situations, their security or other personnel might engage in clashes or similar confrontations with other civilian actors. When the news report refers to “company” “firm” or similar entities as acting subjects of protest events, they will be labeled as organizers. Expressions which refer to individuals that are associated with companies, such as security personnel, will be labeled with participant tags. e.g. In the sentence “The firm retaliated to the protesters by burning their tents”, “firm” will be labeled as organizer\_type tag.

### 4.7.3.3 The Name of the Event Organizer

In case the name of the organizer of the event is mentioned, please tag it/them with organizer\_name tag.

## Examples

- “Frankie Boyle has started a hunger strike in solidarity with Guantánamo Bay prisoner Shaker Aamer”,
- “The Labour Party organized a rally on Sunday”,
- “A prominent banned religious outfit, Deendar Anjuman, exploded a mine on Sunday”
- The University Teachers’ Association joined the campaign and signed the anti- austerity statement.
- “...when the protesters, led by senior youth congress leader Ish Kohli, held a demonstration just before the beginning of the council meeting, Firdous came down and received a memorandum demanding not to handover the property to the private organisation”. youth congress: organizer\_name, leader: organizer\_type, protesters: participant\_type, Ish Kohli: organizer\_name.

## Tag Rules

- i. Tag span rule: Full organizer name will be included. If the abbreviation of the name is present within brackets, it will be tagged separately without the brackets e.g. Datta Gule's auto union, All India Bank Officers' Association, (PUCL), Indian National Trade Union Congress (INTUC), Joint Front of Trade Unions of the GIC and the LIC.
- ii. As it was mentioned in the Facility Name section, capitalization might indicate a unique entity that must be tagged with an organizer\_name tag but due to possible errors in syntax or tokenization, this should not be relied upon on its own.

### 4.7.3.4 The Ideology of the Event Organizer

The organizer of a protest may have an explicit ideology, which can be at least one of Communist/Socialist, Anarchist, Social democrat, Mixed left (including feminists, environmentalists etc...), Other left-wing ideologies, Conservative, Liberal, Nationalist, Religious/Conservative, Far right, Other right-wing ideologies, Mixed right, Feminist, Green movement/Environmentalist, Animal rights, pro-LGBTI, Mixed right/left together, feminists, socialists, or anarchists and environmentalists etc. Mixed right includes right-wing groups that organize together: nationalists and conservatives. Please examine the country specific ideologies list and annotate the occurrence of an ideology with organizer\_ideology.

#### Example

- Civil rights organizations and Left parties on Friday staged a protest rally

### 4.7.3.5 The Ethnicity of the Event Organizer

If an organizer is described using its ethnic or racial identity, mark the identity with organizer\_ethnicity.

### 4.7.3.6 The Religion of the Event Organizer

Organizer's religious identity is annotated with participant\_religion.

### 4.7.3.7 Caste of the Event Organizer

In case caste of the organizer is provided, it should be annotated with organizer\_caste tag. This tag is specific for annotation of news from India.

### 4.7.3.8 Socioeconomic Status of the Organizer

If the news article contains information about whether the organizers of the event are poor, rich, or middle-class people annotate this with the tag organizer\_SES. Please annotate all mentions.

## Tag rules

- i. If there are SES indicative phrases are related to organizers, mark them with this tag. These can be "starvation", "malnutrition".

## 4.7.4 Semantic Categories of Organizers

Like participants, we are also interested in classifying organizations that engage in protest into different sub-types. Every expression which is labeled with the `organizer_type` and `organizer_name` tags will thus be labeled with organizer semantic category tags as well. Note that different from participant names, organizer named entities still contain generic organizer type information, which is the reason why we label organizer names as well as organizer types.

### Tag Rules

- i. Every expression labeled with `organizer_type` and `organizer_name` tags will also be labeled with organizer semantic tags (see overlapping annotation section).
- ii. Organizer semantic tag spans will be coterminous with those of the `participant_type` tags that they are used in an overlapping manner.
- iii. Participant semantic tags will be consistent for each expression that refers to the same organizer in different event sentences throughout the text.

The tags for participant semantic categories are as follows:

#### 4.7.4.1 Political party

As the quintessential and ubiquitous political organization, political parties are the most common organizers of contentious politics events. Every organization which refers to itself as a political party to itself will be labeled with this organizer semantic tag regardless of it competing elections, take part in government, and even being legally recognized as such. The only exception to this are armed/militant organizations that refer to themselves as political parties, which will be labeled with militant/armed organization tag (see below).

#### 4.7.4.2 NGO

Any non-governmental organization, grass roots activist organization, or charity organization will be labeled with the NGO semantic tag.

#### 4.7.4.3 Union

Trade unions, i.e. organizations, associations of blue or white collar workers which aim at protecting their rights, and improving the conditions under which they work. Unions are usually organized on the basis of economic sector or activity. They are usually legally recognized and protected organizations, but there will be instances when such recognition is not present and workers would be acting together under the banner of a union even if it is informal. This category is closely associated with the “industrial action” event category, i.e. in most cases unions will be organizers of protest actions that are considered under that heading. Therefore, mentions of “social movement unions” should be considered under the NGO category, rather than union.



#### 4.7.4.4 Militant/armed organization

Political organizations that are illegal and routinely resort to armed violence in achieving their goals will be labeled with this tag. In most cases, the use of armed violence is built into the founding principles of these organizations, which commonly identify themselves within an armed conflict with states and/or other political organizations. That is to say, when political organizations that are not established on the basis of using armed violence (e.g. ordinary political parties) engage in occasional violent (even armed) protest actions, this is not a sufficient basis to classify them as militant organizations. Similarly, reference to non-violent political organizations as “terrorist” etc. by governments or news reports is not enough on its own to classify them as militant organizations.

#### 4.7.4.5 Chamber of professional

This is a special category of economic organizations which bring together practitioners of certain legally protected professions, such as medical, or legal profession. This is different to unions in that chambers have legally recognized regulatory functions in the conduct of their professions, such as licensing practitioners, determining fees etc. While there will be variations in countries as to how these organizations are established and operate, a good example of this organization that is found in many countries is bars, i.e. institutions of legal profession. In some countries medical associations and associations of engineers will have a similar form. While these organizations are not usually politicized, in some countries, they can become prominent organizers of protest events for their members.

#### 4.7.4.6 Person

When individuals are labeled as organizers of protest events (see the `organizer_type` section for the cases that this is so), they will be labeled with the “person” organizer semantic tag. Note that this tag will be used with both with `organizer_type` expressions (such as “leader”, “president” etc.) and named person organizers (such as “Lula”, “Kirchner” etc.).

#### 4.7.4.7 Other

Any expression that is labeled with `organizer_type` or `name` tag which cannot be classified into categories defined above will be labeled with “other” organizer semantic tag.

### 4.7.5 Target Arguments

The organization, institution, group(s) or individual(s) that is the antagonist of the protest event is annotated with target tags. Targets are the entities that are opposed by those who organize or participate in the protest. Certain protest actions can aim at the targets directly for instance when they are attacked by the protesters. In other instances, the entities towards which the actions are directed are not the targets but nevertheless become the objects of the protests. In the latter case, care should be taken not to tag the objects of the protest as targets. Consider the examples where “small businesses closing their *shops* to protest some **government** regulation,” “workers blocking *roads* to protest their **employer**.” “a majority caste members attack the *settlements* where

members of a **minority caste** live. The entities italicized are the objects of protest actions but the true targets are those expressions that are highlighted. In brief, target implies the entity that is opposed by the protesters in the abstract. Most of the time, the objects of the protest events are entities annotated with facility tags. Facilities, as defined above, are human made structures or spaces which the protests take place in or even target at times and should not be confused with targets. In the examples already mentioned, shops, roads and settlements are facilities even though the actions are directed at them; whereas the government, the employer and the minority caste are targets, i.e. the antagonists of the protest.

#### 4.7.5.1 Target Type

The type of entities that are the targets of the protest events, such as organizations, individuals, groups countries etc. should be identified and annotated with `target_type`. The examples of the organizations are Government (governing party and its politicians, ministries), state, the army, police, employer & business, mayor or governor (including local governments, municipalities and the governorate), Other political parties and politicians (except for the governing party), judiciary, civilians and non-political social groups (students, women, workers etc...), minority groups (ethnic, religious, racial, caste etc.), non-party political organizations (including non-state armed groups, activist organizations), political activists, politically affiliated individuals, intellectuals and alike, public institutions, bureaucracy and bureaucrats (regulatory agencies, public schools and hospitals), the parliament, foreign countries, NGOs, transnational organizations, etc..., unions, armed paramilitary organizations, criminal organizations (drug trafficking, mafia etc...). In case the event did not occur against a specific individual or institution or it is not clear what the target is, there will not be any annotation.

#### Examples

- During the protest, activists accused the **minister** of violating the international law.
- More than 2000 persons staged a protest near Vijay Theatre ground against the district **police**.
- Contract Bus Association **chief** Harish Sabharwal was booed by the protesters.
- **Priests** were among the 7 that were shot dead in Kashmir.

#### 4.7.5.2 Target Name

The name of the target should be annotated with the `target_name` tag.

#### Examples

- NATO, World Bank, Prime Minister Theresa May
- Contract Bus Association chief **Harish Sabharwal** was booed by the protesters
- After 102 hours of hunger strike, there has been no effort from the **ADCB** chairman to resolve the issue.

## 4.8 Event Separation

### 4.8.1 Introduction

Sometimes news articles include information about multiple events at once. Pieces of information about one event might not be applicable to the other event. In this case, we need to distinguish different events within the article. This is referred to as event separation and is subject to a number of rules to ensure coherence in coding. Note that in separating events we need to think of the news text rather than the actual reality that the text recounts. That is to say, we are more interested in the separate event references in the text than whether the said events are actually separate from each other in real life. As will be clearer in the examples below, sometimes it is not possible to know or show for certain whether separate event references correspond to separate real life events.

In order to mark separate events in the flat environment, we use comments in the tag window. Each event will be numbered from 1 and on in accordance with the ***order in which they appear in the article***. The comment style will be as in “Event 2” “Event 2, Event 3” or “Event 3, Event 4, Event 6” etc. The first event will not have a number comment. If the tags do not contain an event number comment they will be assumed to be related to event 1.

Every tag, meaning all tags from all annotation foci (event, participant, organizer etc.) which belongs to a separate event will contain the same event number in the comment.

#### Example

- “At noon, BJP workers gathered in the square and shouted slogans, condemning the failure of the Union Government in delivering justice to the victims of last year’s terror attack at the train station where armed militants killed 25 people.”

There are two separate events in this sentence. BJP workers’ demonstration is the first event, the attack at the train station is the second event (in the order in which they appear in the document). The tags of the first event will not include an event number comment and are as follows:

At noon: e\_time; BJP: org\_name; workers: part\_type; gathered: e\_type, demonstration; in the square: f\_type; shouted slogans: e\_mention ,demonstration; Union Government: target\_type.

The second event’s tags will all contain the comment “Event 2” and are as follows:

last year’s: e\_time; attack: e\_type, armed militancy; at the train station: f\_type; militants: part\_type; killed: e\_mention, armed militancy.

### 4.8.2 Situations in which events will be separated

The separation of event references will be based on difference in at least one of the following: event time, event place, facility (name or type), or participants.

### Time difference

Events that occur at different times will be separated from each other. The time difference necessary for separation is 24 hours. Events that continue throughout the same day will not be separated even if they are reported to occur at different times of the day.

### Location difference

Events which are reported to take place in different locations will be separated as different events. Locations can be **event places** or **facilities**. An event that has started at some place and continued at another, e.g. a march that started at a location and proceeded at somewhere else, is not going to be separated. However, if an event is happening

- a) simultaneously at multiple locations,
- b) or at multiple locations at different times but not in continuum,

it will be separated such that every location reference will count as a separate event. Likewise, a multiple event reference (e.g. protests, attacks etc.) will be separated into the multiple mention **and** each individual event reference that comprise it (see below Plural Event Mentions).

### Participant or organizer difference

Events which are carried out by different participants or organizers with separate goals and motivations will be separated. This separation will take place even in cases where different protests occur at the same time and location. The separation is based on event motivations or goals but since motivation info is not something that we annotate and might at times be elusive, we try to distinguish events based on participants and organizers. The most frequent cases which exemplify this situation are that of counter-protests where two groups of participants or organizers demonstrating against each other and/or with conflicting agendas. Note that in cases where there are multiple types of participants and/or organizers that protest **together**, the event will not be separated.

### Semantic event category difference

Events which occur at the same time, place and facility, and organized and participated by the same participants but have a different semantic event category will be separated as different events. In other words, as a result of separation by semantic event category, the triggers of each event in a document that is separated by its respective event number will have only one semantic category tag. Although this case is rare, it can be encountered when rallies, marches or other types of demonstrations occur during industrial strikes.

## 4.8.3 Rules of event separation

- i. No annotation outside event sentence rule applies to the separated events. In a multiple event situation, a given sentence that contains information about an event whose event type or mention is not present in the same sentence will not be annotated even if it contains event type or mention of another event. Consider a slightly different version of the example above:

- “At noon, BJP workers gathered at the train station and shouted slogans, condemning the failure of the Union Government in delivering justice to the residents of Mumbai.

The incident which had taken place last year at the train station where armed militants killed 25 people and the manner in which the government handled the situation has been a chief agenda item for BJP ever since.”

In this example, “at the train station” expression is the location of both the location where BJP “gathered and militants killed. Yet, in each sentence, it will only contain the event number of the event trigger in that sentence. Also, note that the expressions “government” and “BJP” in the second sentence are not tagged due to the no tagging outside event sentence rule as they are not related to the “killed” event in that sentence.

- ii. Plural Event References: In cases where the articles contain multiple events, there might be event triggers (types or mentions) that are plural, i.e. refer to more than one event that should be separated. We have a unique procedure for separating these events. In a nutshell, if an article contains a plural event reference, such as “protests” which refers to 2 different events, each of which are reported in the article, that article will have 3 separate event numbers.

Consider the example of the first three sentences of a document below:

- Karnataka State Government Employees Association engaged in a wave of demonstrations across Karnataka yesterday, urging the government not to go ahead with the new retirement scheme.

In Bangalore, hundreds of workers participated in the rally in front of the collectorate.

Mysore was also the scene of protest as around 3000 employees took to the streets.

The three sentences above will be treated as containing three separate event references, even though in real life the first reference “demonstrations” comprises the events in the following sentences. The reason for this is that the protests that take place in Karnataka, Bangalore and Mysore all have different event place information attached to them and it is not possible for a reader uninformed beyond what is present in the actual news article, to decide for certain whether these three events coincide or not based solely on the place names. This also applies to cases where the plural events happen at distinct times and facilities (i.e. the case where demonstrations in the above example happen at two distinct times “within the last month”, say yesterday and 2 weeks ago. Again, there will be 3 event annotations one, plural, happening “within the last month”, the second “2 weeks ago” and third “yesterday”.)

If the different event times and/or locations are mentioned in the same sentence, both the plural event trigger and other pieces of information belonging to the separate events will be numbered in the same sentence.

- “Karnataka State Government Employees Association organized demonstrations in Bangalore and Mysore yesterday, urging the government not to go ahead with the new retirement scheme.”

Here, all tagged expressions except event places will contain “Event 1, Event 2” comment as all of them give the same information about two events that are separated according to the different place, different event rule. Bangalore will not contain an event number tag (because it is event 1) and Mysore will contain the tag “Event 2”.

## 5 Violent Protest Events Annotation Manual

This task will follow the token level annotation task, although it is a classification task that is performed on the basis of the whole document. The articles which contain at least one violent protest events, as per defined below, will be labeled as **violent**. On the contrary, the documents which do not contain any violent events will be labeled **non-violent**. The documents selected for labeling in this task are those that have already been identified as containing at least one contentious politics event, as defined in DOLPAM. Only those events which are included in our event definition will be evaluated on the basis of violence. That is to say, events which do not conform with the event definition laid out in DOLPAM will not be evaluated and labeled.

### 5.1 Identifying violent protest events

Violent protest events are contentious politics events whose participants deliberately resort to violence any time during the event. Violence refers to causing harm and/or damage to individuals or property in any way. The harm or damage in question must be inflicted deliberately, that is, accidents or any harm caused unintentionally will not be enough on their own to label the event as a violent one. Violence must be understood as any action which has as its goals:

- a) Infliction of bodily harm to any individual(s). Possible targets of the violent act include: state officials (including security personnel), any civilian or group of civilians that are the antagonists of the protest (i.e. targets of the protest action), participants of the event themselves (in cases of internal clashes among participants, and acts of self-harm such as self-immolation, hunger strike and the like), or individuals which have neither any role nor relevance to the event such as bystanders etc.
- b) Infliction of damage to property and/or environment: Government offices, public built environments (e.g. roads, parks), buildings belonging to any non-official entity (party or union offices, shops or homes belonging to a particular minority), vehicles, religious sites or places of worship, natural environment (e.g. in cases of forest burning) etc.

#### 5.1.1 Some protest events that are violent

- i. Forceful removal or destruction of movable property such as ransacking, looting are to be considered violent events. Theft actions serving political goals and/or carried out by political organizations which are otherwise petty crimes are also to be marked violent events. Examples to this are bank robberies, looting warehouses, hijacking public transport, extortion, mugging and the like.
- ii. Nude parades is a lynch-like form of violent action where the victim is stripped naked and paraded in public spaces for degradation and insult. It is thus a violent event.

- iii. Violence threats and attempts that are included in our event definition as per DOLPAM are to be considered violent events even if they are not carried out. Instances of hijacking commonly involve threat of violence and thus will be considered as violent events.
- iv. Protests which involve deliberate self-harm by participants are considered violent events. Self-immolation and hunger strikes are among examples of self-harming protests.
- v. References to events that are by their nature likely to involve violence must be considered carefully. Riots, civil unrest events and the like might be reported without explicit mention of the damages caused by violence. An event mention such as civil unrest or disobedience might not, on its own, be enough to label the event as violent. In such a case other details in the article, for instance expressions such as “victims”, “violence”, “damage” can give away the occurrence of violence. Riots on the other hand can be assumed violent on their own, unless the article explicitly states that it is non-violent.

### **5.1.2 Some protest events that are not violent**

- i. Violence must be resorted to by the participants of the protest. In cases where there is violence directed at the participants by security forces but there’s not an explicit mention of participants resorting to violence in retaliation etc., these documents must be labeled non-violent. On the other hand, a protest event that started as non-violent, met with state violence and then involved participants’ violent retaliation is to be labeled a violent event.
- ii. Sometimes the news articles might use the expression “violent” while describing protest events without giving specific details as to the nature of the violence (e.g. “violent reaction”, “violent opposition”, “violently objected” etc.). In such cases, the qualifier violent might refer to the intensity of the protest and thus not indicate actual violence. Take care in such cases to detect other expressions which leave no doubt as to the use of physical violence by the participants. If such statements do not exist, do not label the event as violent.
- iii. Actions which can be characterized as symbolic violence which aims at mockery, insult, ridicule, defamation or desecration are not to be labeled violent events. Such actions might be directed at sacred religious symbols, artifacts of sentimental value such as flags or people. Examples include flag burning, offensive slogans, paint bombs, throwing eggs, tomatoes, shoes etc.



## 6 Protest Event Demands Annotation Manual

In the Emerging Markets Welfare project, we are interested in determining the effect of contentious political movements and events in shaping welfare policies of governments that respond to them. Protest events have various demands and grievances that their participants enunciate at their core. In researching the effects of protest events and movements on welfare policies, a key question to address is whether the types of demands and grievances that give rise to protest events make a difference in terms of governments responding with welfare policies. In this task, we will classify news articles about protest events in accordance with the demand and/or grievance which motivates the protest action. To underline, similar to the violent events annotation task, demand annotation task is also a classification task performed on the document level. For this classification, we define three broad categories of demand/grievance types in terms of their economic and welfare related nature. These categories are as follows:

### 6.1 Non-Economic Demands

These are demands and grievances which stem from social and political issues that are not related to social conflicts based on distribution of resources or production processes. Demands and grievances that stem from social and political conflicts based on identity and territory are to be considered within this category. Protest events that are motivated by issues regarding political representation, participation, human rights and liberties, and environment are also within this category. To enumerate, all protest event demands and grievances that are based on ethnic, racial, gender, religious and caste divisions; ecological concerns; political issues of territorial and/or sovereignty issues, representation, accountability of officials, corruption, state and/or social violence, political rights and liberties (such as freedom of expression), are to be regarded in the non-economic category.

### 6.2 Economic Non-Welfare Demands

These are demands and grievances that relate to social conflicts that are based on distribution and production relations with the exclusion of those that relate to social welfare, which is identified below. All protest events that are motivated by demands and grievances based on labor issues and other economic interest based conflicts, poverty and economic inequality, public infrastructural investment, and land ownership are to be categorized within economic-non-welfare category. To enumerate, issues such as cost of living, inflation, wage levels, labor protection and employment, taxation, infrastructure and municipal services, housing, and land ownership are to be regarded in the economic-non-welfare category.

### 6.3 Economic Welfare Demands

These are demands and grievances that stem from issue of economic distribution and relate directly to social welfare. By social welfare, we refer to politics of social policy in its following

dimensions: social security (old age pension, unemployment and sickness insurance, non-wage benefits of waged employment), social assistance, education, healthcare, social care services (child care, care for the sick and elderly). Protest events whose participants demand welfare provision and/or higher quality welfare provision in any of the fields enumerated above are to be regarded in the economic-welfare category.

## 6.4 Rules of Demand Annotation

- i. There will be news articles which will not report event demands explicitly. In these cases try to infer the demands from implicit information, for instance from the event types or background information present in the text. However, try not to base your decisions on information external to the news text.
- ii. Strikes (industrial actions) will be labeled economic-non-welfare unless a welfare related demand is mentioned in the article.
- iii. Demand annotation must be exclusive. If a protest event enunciates multiple demands, chose the first demand that the report mentions as the demand of the protest event.
- iv. Similar to the rule above, documents that contain multiple separate events will be evaluated in terms of the demands of the first appearing event.

## 7 Publications Based on the Annotation Manual

Hürriyetoglu, A., et. al. (2019). A task set proposal for automatic protest information collection across multiple countries. In: Leif Azzopardi, et al. (eds) *Advances in Information Retrieval*, pp. 316–323. Springer International Publishing, Cham.

Hürriyetoglu, A., et al. (2019). Overview of clef 2019 lab protestnews: Extracting protests from news in a cross-context setting. In: Fabio Crestani, et al. (eds) *Experimental IR Meets Multilinguality, Multimodality, and Interaction*, pp. 425–432. Springer International Publishing, Cham.

Hürriyetoglu, A., et al. (2020) Automated extraction of socio-political events from news (AESPEN): Workshop and shared task report. In: *Proceedings of the Workshop on Automated Extraction of Socio-political Events from News*, pp. 1–6. European Language Resources Association (ELRA), Luxembourg.

Hürriyetoglu, A., Yörük, E., Yüret, D., Mutlu, O., Yoltar, Ç., Duruşan, F., & Gürel, B. (2020). Cross-context news corpus for protest events related knowledge base construction. *Automated Knowledge Base Construction*.

Hürriyetoglu, A., Yörük, E., Mutlu, O., Duruşan, F., Yoltar, Ç., Yüret, D. and Gürel, B.: Cross-context news corpus for protest event-related knowledge base construction. *Data Intelligence* 3(2), 2021. doi: 10.1162/dint\_a\_00092.

Yörük, E., et al. (2021) Random sampling in corpus design: Cross-context generalizability in automated cross-national protest event collection. To be published in *American Behavioral Scientist*.