On the Road to a Protest Event Ontology for Bulgarian: Conceptual Structures and Representation Design

Milena Slavcheva
Institute of Information and Communication Technologies
Bulgarian Academy of Sciences
milena@lml.bas.bg

Hristo Tanev
Joint Research Centre
European Commission
hristo.tanev@ec.europa.eu

Onur Uca
Mersin University
Sociology Department
onuruca@mersin.edu.tr

Abstract

The paper presents a semantic model of protest events, called *Semantic Interpretations of Protest Events* (SemInPE). The model is a practical application of the *Unified Eventity Representation* (UER) formalism, which is based on the *Unified Modeling Language* (UML), whose four-layer architecture (i.e., user objects, model, metamodel, and meta-metamodel) provides flexible means for building the semantic representations of the language units along a scale of generality and specificity. The analytical framework, inspired by the object-orientation paradigm in computer science and a cognitive approach to the linguistic analysis, provides suitable devices for capturing the continuously varying information in the social and political domain. The basic modeling elements of events are presented, which include modeling elements defining classes of participants in the events, types of relationship among the participants, as well as the participants behaviour. The acquisition of language objects that serve as instances of the various semantic classes contained in the model is also discussed.

1 Introduction

The paper presents a semantic model of events, which can be broadly defined as protest events. The model, which we call *Semantic Interpretations of Protest Events* (SemInPE) is a practical application of the *Unified Eventity Representation* (UER) - a cognitive theoretical approach and a graphical formalism (Schalley, 2004) based on the *Unified Modeling Language* (UML)[1] - an international standard for graphical representation and design of object-oriented systems in the field of Information Technologies (OMG, 2001).

The analytical framework used for building the semantic representations is inspired by the object-oriented paradigm in computer science and a cognitive approach to the linguistic analysis (Schalley, 2014). The application of this innovative formalism in our work is motivated by several merits of its, relevant to the task of building Language-Technologies-style ontologies utilisable in the social and political sciences.

The analytical framework we apply is based on the four-layer metamodel (i.e., user objects, model, metamodel, and meta-metamodel) of the *Unified Modeling Language* (UML). This multi-layered architecture provides flexible means for building the semantic representations of the language objects along a scale of generality and specificity. The inheritance mechanism of classes and objects provides a device for the definition of abstract, underlying semantic representations, which can be instantiated by specific descriptions corresponding to specific topics and specific languages. In our case, this particular conceptual modeling paves the way to building an ontology of protest events for Bulgarian, but it is utilisable in multilingual settings as well. The structuring devices of the applied model provide the possibility for a modular and dynamically extensible knowledge representation, which is of particular importance for capturing the continuously varying information in the social and political domain.

The cognitive approach to representing the linguistic units provides a conceptual modeling, which corresponds to the conceptualisation of the object-oriented modeling (Schalley, 2014). In this way a direct use of the handy object-orientation devices is ensured in the semantic representation of language entities. We can also point out the presence of ontological knowledge (i.e., relation to real-world knowledge) in the semantic descriptions via the reference to ontological categories (see Section 3).

The paper is structured as follows. In Section 2 we briefly refer to related work. In section 3 we present the model to be utilised in building an ontology in the domain of protest events. In section

[1] https://www.uml.org/

Proceedings of the 6th Workshop on Challenges and Applications of Automated Extraction of Socio-political Events from Text, pages 92–100, Varna, Bulgaria, Sep 7, 2023.
https://doi.org/10.26615/978-954-452-089-2_013

4 the extraction of language data necessary for our work is discussed. Section 5 provides concluding remarks and some hints on the future developments we envisage.

2 Related work

As pointed out in the Introduction, the work presented in this paper belongs to the analytical framework defined as object-oriented semantics. This relatively novel approach in linguistics so far has been demonstrated predominantly in the analysis of the meanings of verbs (e.g., (Schalley, 2004, 2014), (Benz, 2014), (Slavcheva, 2008, 2012). However, more recently, Morrissey and Schalley (2017) argue that the object-oriented approach is beneficial for the semantic representation of nominals as well, which is a useful development for large-scale conceptual modeling. The approach to the linguistic analysis in the work presented in this paper is a cognitive one. It is determined by principles relating perception, thinking and language. The basic assumption is that language reflects "patterns of thought", hence the study of language is connected to the exploration of "patterns of conceptualization" (Evans and Green, 2006). This makes it possible to relate conceptual structures of language to the conceptual base of object-oriented programming languages.

The graphical semantic formalism used in the application presented in this paper employs the *Unified Modeling Language* (UML) (OMG, 2001), which contains notation techniques for combining structural (that is, static) and behavioural (that is, dynamic) modeling. A long-term research work for developing ontological foundations for conceptual modeling has used UML in building the framework of the Unified Foundational Ontology (UFO) and especially in the development of OntoUML – "ontologically well-founded conceptual modeling version of UML" (Guizzardi et al., 2015, 2022).

The work on the conceptually grounded semantic descriptors provokes a comparison with ontologies like, for instance, the Suggested Upper Merged Ontology (SUMO)[2] (Niles and Pease, 2001), which has been mapped to WordNet (Niles and Pease, 2003). An open source knowledge engineering environment, Sigma, has been created (Pease and Benzmueller, 2013), which includes a full first order inference capability, as well as a natural language paraphrase capability for logical axioms.

[2]http://www.ontologyportal.org/

There are various taxonomies describing political events. The earliest event taxonomy for text analysis, which includes political events, is introduced in the context of the Automatic Content Extraction program (Ahn, 2006) and the following TAC initiative (Mitamura et al., 2017).

Other outstanding taxonomies in this domain include the Intrusion Detection Extensible Alert Taxonomy IDEA (Kácha, 2014), and the CAMEO taxonomy (Gerner et al., 2002). Several event data bases and systems such as GDELT and ICEWS (Ward et al., 2013) use CAMEO. Although CAMEO is sometimes referred to as ontology, the first fully fledged ontology in the domain of political events is PLOVER (Halterman et al., 2021), which includes protests and other political events as classes. An overview of the existing ontologies and taxonomies is presented in Balalia et al. (2021); they also introduce their own ontology, called COFEE.

The ontologies and taxonomies mentioned so far refer to the large domain of political event detection. In contrast, very little work is dedicated specifically to protest events: Danilova (2015) describes a model which includes arguments and classes similar to the ones we observed. Relevant to our work is also the multilingual NEXUS event detection system, which uses linguistic rules, lexicon-based event classification, and an ad-hoc taxonomy of event classes to detect protests, riots, and other conflict events (Piskorski et al., 2007).

Protest events have also been studied by the political and social sciences. Duruşan et al. (2022) defines the protest as "an action through which individuals, groups, or organizations voice their objections, oppositions, demands, or grievances to a person or institution of authority". According to Parry (2023), the value of protest consists in making a difference; the successful protest being the one that effects change in line with the protesters' goals.

Event databases, such as POLECAT (Halterman et al., 2023), the CAMEO dataset (Salam et al., 2020) and others represent a bridge between the world of ontologies and political sciences. They introduce means for qualitative political studies, trend analysis and conflict prediction.

3 The model

As pointed out above, the *Semantic Interpretations of Protest Events* (SemInPE) model we present

here is based on the *Unified Eventity Representation* (UER) theoretical approach and formalism (Schalley, 2004). In the subsections that follow we present basic modeling elements and the way they will be used in the ontology construction.

3.1 Eventity frames

A central modelling element is the EVENTITY FRAME, which represents the semantics of verbs as the key lexical encoders of events (or, eventities in the UER terminology) in texts. The EVENTITY FRAMES describe the eventity PARTICIPANTS, as well as their interaction and behaviour. The EVENTITY FRAMES incorporate modelling elements, each one of which can be specified to a different degree depending on the concrete task. Figure 1 contains an EVENTITY FRAME TEMPLATE, which, after binding its parameters, can describe verbs that typically occur in texts discussing protests like, for example, bg. *protestiram* ('protest') as used in sentence (1).

(1) *Zsiteli na krivodolskoto selo Osen protestiraha sreshtu avtomobilniya trafik.*
(Eng.transl.) Inhabitants of the Osen village in the Krivodol region protested against the automobile traffic.

In the diagram in Figure 1, there is one prominent PARTICIPANT (the protester) represented by a PARTICIPANT CLASS stating that the PARTICIPANT ROLE is `Agent`, the PARTICIPANT ontological TYPE is `Individual`, and there is an ATTRIBUTE further characterising the participant as `human`. The prominent participant's behaviour is described in the dynamic core of the EVENTITY FRAME (denoted by the dashed-outline rectangle with rounded corners), which contains a STATE-machine, in this case consisting of an ACTIVE SIMPLE STATE (ASS) (depicted by the shape with straight top and bottom arcs, and convex arcs on the two sides). The ACTIVE SIMPLE STATE (ASS) denotes activities, actions performed by the prominent participant. The second participant is the reason, the motive, the stimulus[3] for the protest event.

As pointed out above, the EVENTITY FRAME in Figure 1 is a TEMPLATE, that is, it includes a parameter to be bound (indicated by the dash-outline rectangle in the upper right corner of the octagon). The parameter can be bound to names of the ACTIVE SIMPLE STATE, which refer to basic concepts like, for example:

ASS = {Protest, Strike, Demonstrate, . . . }

The specification of the STATES depends on the modeling granularity determined for a given representation and ranges from underspecification to different degrees of specification with the help of clusters of PROPERTIES. The PROPERTIES, which are part of the metamodel, have values of the ENUMERATION or *Boolean* data type. For example, the STATE-machine of the verb bg. *buntuvam se* ('riot') can be represented in the way shown in Figure 2.

It should be noted that the STATE-machines can provide conceptual structuring of different complexity. They can include modeling devices like SUBMACHINE STATES or SUBCORE STATES, which can reference reusable conceptual structures (or conceptual 'macros') specified elsewhere in the model. STATE-machines can make use of COMPOSITE STATES, which model processes and can describe the sequential and concurrent steps in those processes. The granulated structure of the dynamic core is beyond the scope of the current paper. Its development will be reported in follow-up works.

3.2 Participants

The PARTICIPANT CLASSES are selectors for sets of OBJECTS, which stand for participants appropriate for a given eventity. The metamodel provides the possibility for building taxonomies of participants whose modelling elements are at a different level of abstraction.

The participants belong to different ontological categories, which are referenced by PARTICIPANT TYPES merged into a very concise participant type ontology. This small ontological type hierarchy contains generalised categories, which roughly determine the kind of modeling elements that are used to specify the PARTICIPANT CLASSES and the PARTICIPANT OBJECTS as instances of those classes. The root node of the ontological type hierarchy is `Entity` and it encompasses the two top level categories of `Eventity` and `Ineventity`[4]. There is a small number of sub-levels further down the hierarchy, but what concerns directly our work here is that: 1) one of the subdivisions of `Ineventity`

[3] One would intuitively say the cause for the protest, but the word *cause* is deliberately avoided as it is reserved to name a central modeling element, the cause-SIGNAL

[4] Currently, we follow the naming and the definition of the high-level ontological types as set in the UER (Schalley, 2004). However, the ontological type hierarchy can be adjusted and complemented by a particular ontology designer. We envisage a further development and specification of the ontological types.

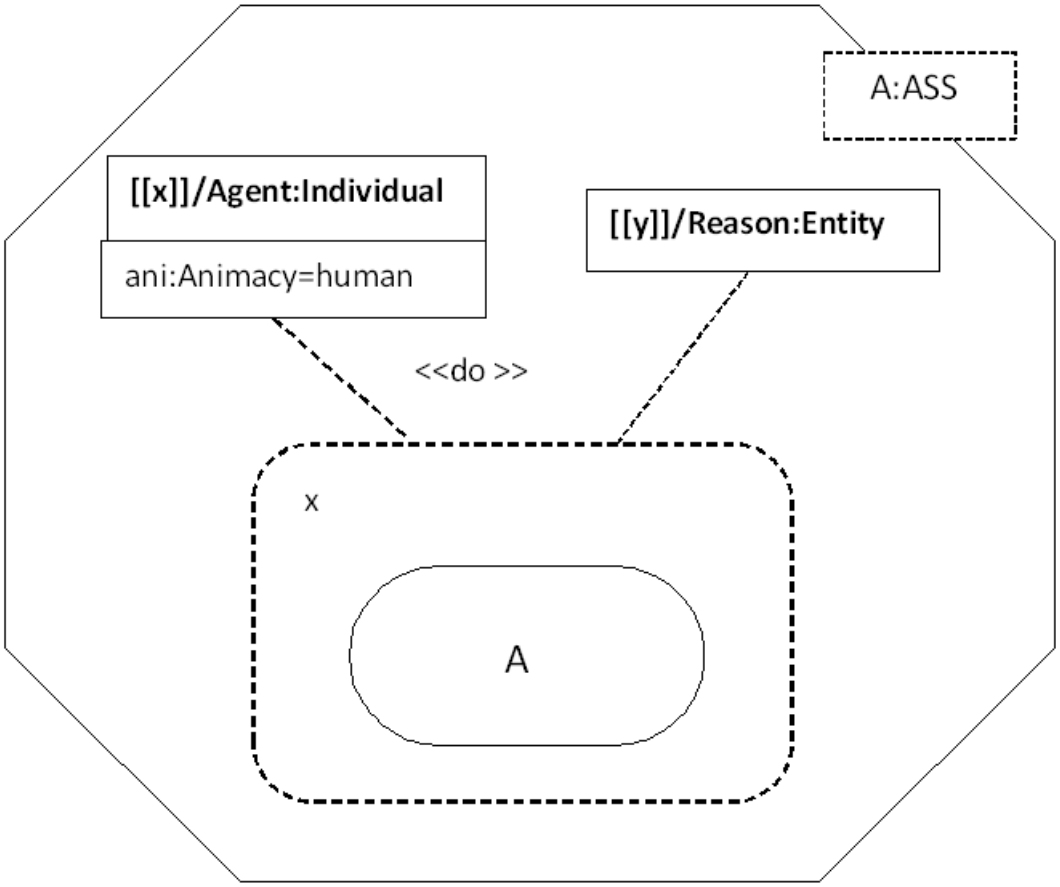

Figure 1: EVENTITY FRAME TEMPLATE modeling verbs typical for the topic of protests.

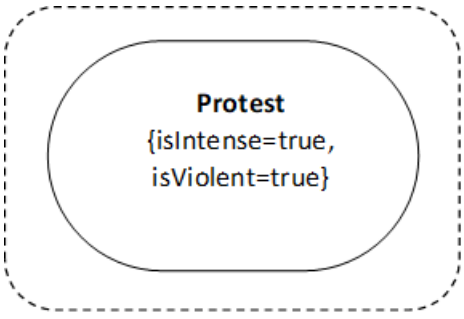

Figure 2: Dynamic core of the verb *buntuvam se* ('riot')

is the `Individual`[5] category, which is the typical category of protesters, and 2) it is possible for a participant to be of the `Eventity` category, which is a typical type for the reason, the motive for a protest.

The major discriminators of eventity PARTICIPANTS are the ATTRIBUTES that characterise them. The number and type of the ATTRIBUTES that specify a given PARTICIPANT CLASS can vary depending on the concrete task. In addition, given ATTRIBUTES can stay unspecified depending on the implementation.

For example, protesters can be characterised by clusters of ATTRIBUTES as shown in Figure 3.

The values of the ATTRIBUTES in Figure 3 are of data type ENUMERATION as exemplified in Figure 4.

Looking at the data extracted from news texts in Bulgarian (see Section 4), we can find several general semantic dimensions of reasons for protesting. Some of the protests belong to a single dimension (e.g., a political protest against a new government), others are characterised by features stemming from more than one dimension (e.g., demands for increase of the salaries of medical personnel, which concern the social, economic, and health dimensions).

The semantic dimensions are defined as follows:

[5]It should be noted that the category `Individual` does not stand for the concept of Person but for any entity that conforms to the ATTRIBUTE {inherentlyBounded = true}

Figure 3: Attributes describing a protester participant.

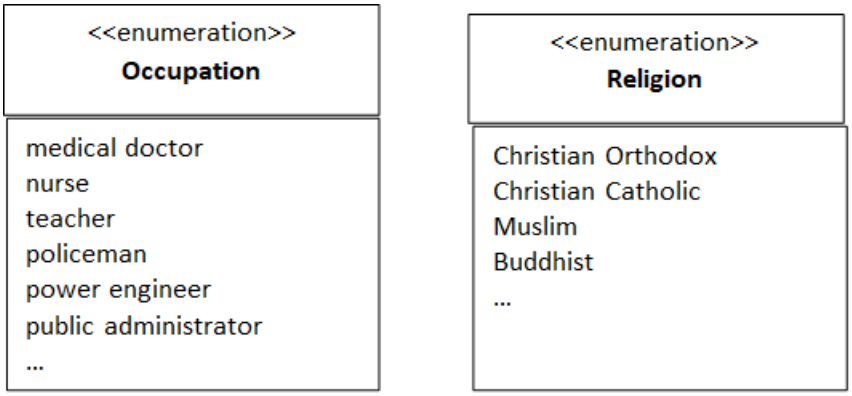

Figure 4: Enumerations of attribute values characterising protest participants.

- *Political dimension*. It is related to the political realm of the Parliament, the Presidency, the government and the governmental administration, the administration of the regions, etc.
- *Social dimension*. It concerns the rights and welfare of different social groups, for example, the status of old people, the education, etc.
- *Criminal dimension*. Criminality, mafia, corruption, and the rule of law in general are the protest triggers in this dimension.
- *Ideological dimension*. Here the concerns are related to ultra right or ultra left movements, political figures, and similar concerns of the protesting people related to various ideologies, which are not acceptable according to them.
- *Religious dimension*. It includes protests stemming from religious convictions, for example, protests for the rights of the Islamic population in China.
- *Legislation dimension*. Here the demands are directed in favour or against a new legislation or an old legislation, which is in conflict with certain social realities.
- *Health dimension*. Health is an important concern in society, especially during and after the COVID pandemic. Various protests target vaccines, health insurances, health legislation and the health system in general. This dimension is related to the social dimension.

On the basis of the above summarising, for the `Reason` participant, we can define PARTICIPANT CLASSES as exemplified in Figure 5. Examples of lexical items represented by such a participant class are: "the President", "the opposition", "the mafia", "new law", "the COVID masks", etc.

Typically, the `Reason` participant is of the ontological type `Eventity`. It can be assigned a sub-division category of `Eventity` like `Proposition` (which encompasses abstract eventities), `State`, `Process`. Except for the `Animacy` attribute, which is irrelevant for an `Eventity` type of `Reason`, the attributes in Figure 5 are valid also for the `Eventity` type of

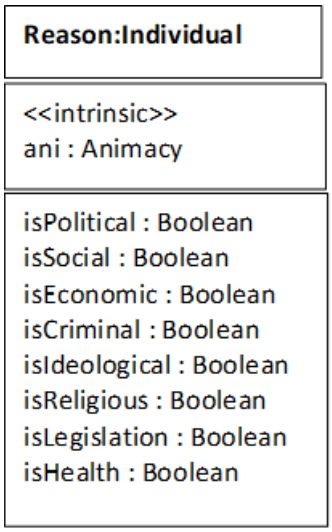

Figure 5: Participant kind of reason

`Reason`. In addition, the `Reason Eventity` can be represented, in its own right, in a structured way. Examples of `Reason` participants of the `Eventity` type are: "increase of prices", "murder", "firing of workers", "construction", "animal rights violation", "the lack of treatment of mosquitoes", etc.

Needless to say, the `Reason` ontological types are characterised by plasticity, that is, the `Individual` and the `Eventity` types are interchangeable. For example, a protest against the President, in one case, can be viewed to be against the personality of the president, in other case, against actions of the president.

The semantic representation of protest events can be enriched by modeling the relations among the participants, which is the subject of discussion of the next section.

3.3 Relationships among participants

An EVENTITY FRAME describing a protest event can incorporate different PARTICIPANTS, which are in various relationships with one another. The different aspects of those relationships can be described by the ASSOCIATION modeling element, as well as the ASSOCIATION CLASS, which displays properties of the ASSOCIATION. The modeling elements of this kind are a useful device for providing rich semantic descriptions of the relations among the different types of participants, which we illustrate by the examples below.

For a given semantic representation, it would be necessary to point out the relation of employee and employer between the participants in a protest event as displayed in Figure 6.

The ASSOCIATION CLASS connects the PARTICIPANT CLASSES and defines a set of features that describe the relationship itself as exemplified in Figure 7.

4 Data

The first step in preparing our semantic model was to acquire language objects that serve as instances of the various semantic classes contained in the model:

1. We extracted nouns in Bulgarian, whose very close equivalents in English are "protest", "demonstration", "riot", "strike", etc.

2. We identified the verbs that are morphologically and semantically related to those nouns. For example, the correlative of the noun bg.*protest* ('protest') is the verb bg.*protestiram* ('to protest'). Bulgarian is a language of very rich verb morphology, hence, specific members of the verb form paradigm are of interest, in this case, bg.*protestirat* (present tense, plural), and bg.*protestiraha* (past tense, plural). These verb forms are frequently used to denote the focus of news articles describing protest events and convey meanings related to "actions happening at the moment", "actions that happened in not distant past", and "actions performed by a number of people".

3. We searched for relevant terms in a corpus of approximately 100,000 news articles in Bul-

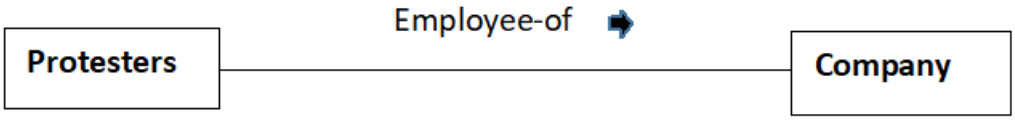

Figure 6: ASSOCIATION between two PARTICIPANTS.

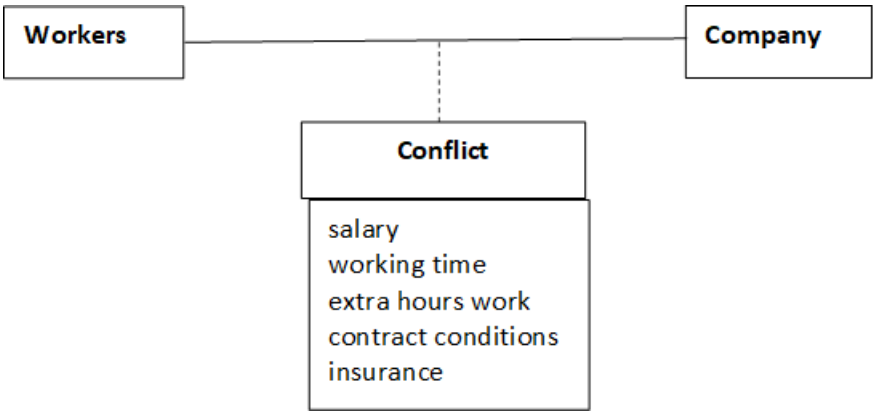

Figure 7: ASSOCIATION CLASS describing the relation of conflict between the participants.

garian gathered by scraping various Bulgarian news websites in the period 2021-2022.

4. Then we gathered all the uni- and bi-grams that appeared in immediate proximity to the search terms, where only one non-stop word was allowed between the search term and the n-gram.

5. We calculated the TF.IDF for each n-gram extracted in this way, and picked out the 500 with the highest TF.IDF.

6. Then we manually identified the terms for the respective target semantic class in this list of 500 terms adjacent to "protest", its synonyms and hyponyms in the Bulgarian language.

As an additional data source we used the Bulgarian section of Google News[6] and downloaded 100 news articles from 2022 and 2023 related to protests, riots and strikes, and manually extracted from them additional relevant terms for each semantic class under consideration.

It should be noted that all semantic classes were extracted from the aforementioned set of terms and the Google news corpus.

[6]http://news.google.com

In this way we extracted the terms for the semantic class PROTEST REASONS (here the English translations of the Bulgarian lexical items are given): "construction", "new law", "new order", "increased prices", "the President", "the opposition", "the conditions", "the mafia", "the ambassador", "the COVID masks", "murder", "working conditions", "animal rights", etc. These are protest reasons, typical for the Bulgarian society. Similarly, we can deal with the other semantic classes in the model like CONFLICT, OCCUPATION, RELIGION, etc.

5 Conclusion

We presented a semantic model, which contains flexible devices for representing the underlying conceptual structures of protest events. They include modeling elements defining classes of participants in the events, types of relationship among the participants, as well as the participants behaviour. The modeling framework of object-orientation proves to be a convenient tool for building information structures in language semantics, which can be adjusted to serve specific tasks and user demands. This assertion has been demonstrated by modeling elements of different degree of abstraction, which constitute a dynamic system of interrelated seman-

tic classes.

The presented *Semantic Interpretations of Protest Events* (SemInPE) model underlies the construction of the protest event ontology for Bulgarian, which is the next step on the way of providing resources enhancing the text processing in the social and political domain.

References

David Ahn. 2006. The stages of event extraction. In *Proceedings of the Workshop on Annotating and Reasoning about Time and Events*, pages 1–8.

Ali Balalia, Masoud Asadpoura, and Seyed Hossein Jafaria. 2021. COfEE: A Comprehensive Ontology for Event Extraction from text, with an online annotation tool. *CoRR*.

Anton Benz. 2014. Ergativity and the object-oriented representation of verb meaning. In Klaus Robering, editor, *Events, Arguments, and Aspects – Topics in the Semantics of Verbs*, pages 65–87. John Benjamins, Amsterdam / Philadelphia.

Vera Danilova. 2015. A pipeline for multilingual protest event selection and annotation. In *2015 26th International Workshop on Database and Expert Systems Applications (DEXA)*, pages 309–313. IEEE.

Fırat Duruşan, Ali Hürriyetoğlu, Erdem Yörük, Osman Mutlu, Çağrı Yoltar, Burak Gürel, and Alvaro Comin. 2022. Global contentious politics database (GLOCON) annotation manuals.

Vyvyan Evans and Melanie Green. 2006. *Cognitive Linguistics: an Introduction*. Edinburgh University Press, UK.

Deborah J Gerner, Philip A Schrodt, Omür Yilmaz, and Rajaa Abu-Jabr. 2002. Conflict and mediation event observations (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association, New Orleans*.

Giancarlo Guizzardi, Alessander Botti Benevides, Claudenir M. Fonseca, Danielle Porello, Joao Paulo A. Almeida, and Tiago Prince Sales. 2022. UFO: Unified Foundational Ontology. *Applied Ontology*, 17(1):167–210.

Giancarlo Guizzardi, Gerd Wagner, Joao Paulo A. Almeida, and Renata S. S. Guizzardi. 2015. Towards ontological foundations for conceptual modeling: The Unified Foundational Ontology (UFO) story. *Applied Ontology*, 10(3):259–271.

Andrew Halterman, Benjamin Bagozzi, Andreas Beger, Phil Schrodt, and Grace Scraborough. 2023. PLOVER and POLECAT: A new political event ontology and dataset.

Andrew Halterman, Katherine Keith, Sheikh Sarwar, and Brendan O'Connor. 2021. Corpus-level evaluation for event QA: The IndiaPoliceEvents corpus covering the 2002 Gujarat violence. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4240–4253, Online. Association for Computational Linguistics.

Pavel Kácha. 2014. Idea: security event taxonomy mapping. In *18th International Conference on Circuits, Systems, Communications and Computers*.

Teruko Mitamura, Zhengzhong Liu, and Eduard H Hovy. 2017. Events detection, coreference and sequencing: What's next? overview of the TAC KBP 2017 event track. In *TAC*.

Lochlan Morrissey and Andrea C. Schalley. 2017. A lexical semantics for refugee, asylum seeker and boat people in Australian English. *Australian Journal of Linguistics*, 37(4):389–423.

Ian Niles and Adam Pease. 2001. Towards a Standard Upper Ontology. In *Proceedings of the 2nd international conference on Formal Ontology in Information Systems (FOIS-2001)*.

Ian Niles and Adam Pease. 2003. Linking lexicons and ontologies: Mapping WordNet to the Suggested Upper Merged Ontology. In *Proceedings of the IEEE International Conference on Information and Knowledge Engineering*.

OMG. 2001. *OMG Unified Modeling Language Specification, Version 1.4*. Object Management Group (OMG), http://www.omg.org/.

Jonathan Parry. 2023. What's the point of protest? https://www.lse.ac.uk/philosophy/blog/2023/02/15/whats-the-point-of-protest/.

Adam Pease and Christoph Benzmueller. 2013. Sigma: An integrated development environment for logical theories. *AI Communications*.

Jakub Piskorski, Hristo Tanev, and Pinar Oezden Wennerberg. 2007. Extracting violent events from on-line news for ontology population. In *Business Information Systems: 10th International Conference, BIS 2007, Poznan, Poland, April 25-27, 2007. Proceedings 10*, pages 287–300. Springer.

Sayeed Salam, Patrick Brandt, Vito D'Orazio, Jennifer Holmes, Javiar Osorio, and Latifur Khan. 2020. An online structured political event dataset based on CAMEO ontology.

Andrea C. Schalley. 2004. *Cognitive Modeling and Verbal Semantics. A Representational Framework Based on UML*. Trends in Linguistics. Studies and Monographs 154. Mouton de Gruyter, Berlin – New York.

Andrea C. Schalley. 2014. Object-orientation and the semantics of verbs. In Klaus Robering, editor, *Events, Arguments, and Aspects – Topics in the Semantics of Verbs*, pages 159–186. John Benjamins, Amsterdam / Philadelphia.

Milena Slavcheva. 2008. Thinking in objects: Towards an infrastructure for semantic representation of verb-centred structures. In Petr Sojka, Aleš Horák, Ivan Kopeček, and Karel Pala, editors, *Proceedings of 11th International Conference on Text, Speech and Dialogue (TSD 2008)*, Lecture Notes in Artificial Intelligence, Vol. 5246, pages 193–200. Springer-Verlag, Berlin Heidelberg.

Milena Slavcheva. 2012. Semantic descriptors of the reflexive forms of verb structures in contemporary Bulgarian, French and Hungarian. *Abstracts of Dissertations, Bulgarian Academy of Sciences, Institute of Information and Communication Technologies*.

Michael D Ward, Andreas Beger, Josh Cutler, Matthew Dickenson, Cassy Dorff, and Ben Radford. 2013. Comparing gdelt and icews event data. *Analysis*, 21(1):267–297.