

A Comparison of Event Representations in DEFT

Ann Bies^a, Zhiyi Song^a, Jeremy Getman^a, Joe Ellis^a, Justin Mott^a, Stephanie Strassel^a,
Martha Palmer^b, Teruko Mitamura^c, Marjorie Freedman^d, Heng Ji^e, Tim O’Gorman^b

^aLinguistic Data Consortium, University of Pennsylvania
3600 Market Street, Suite 810, Philadelphia, PA 19104, USA
{bies, zhiyi, jgetman, joellis, jmott, strassel}@ldc.upenn.edu

^bUniversity of Colorado at Boulder, Department of Linguistics
295 UCB, Boulder, CO 80309, USA
{Martha.Palmer, ogormant}@colorado.edu

^cLanguage Technologies Institute, Carnegie Mellon University
6711 Gates-Hillman Center, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA
teruko+@cs.cmu.edu

^dRaytheon BBN Technologies
10 Moulton St., Cambridge, MA 02138, USA
mfreedma@bbn.com

^eRensselaer Polytechnic Institute, Department of Computer Science
110 8th Street, Troy, NY 12180, USA
jih@rpi.edu

Abstract

This paper will discuss and compare event representations across a variety of types of event annotation: Rich Entities, Relations, and Events (Rich ERE), Light Entities, Relations, and Events (Light ERE), Event Nugget (EN), Event Argument Extraction (EAE), Richer Event Descriptions (RED), and Event-Event Relations (EER). Comparisons of event representations are presented, along with a comparison of data annotated according to each event representation. An event annotation experiment is also discussed, including annotation for all of these representations on the same set of sample data, with the purpose of being able to compare actual annotation across all of these approaches as directly as possible. We walk through a brief example to illustrate the various annotation approaches, and to show the intersections among the various annotated data sets.

1 Introduction

This paper will discuss and compare event representations across the various types of event annotation that are part of the Deep Exploration and Filtering of Text (DEFT) program: Rich Entities, Relations, and Events (Rich ERE), Light Entities, Relations, and Events (Light ERE), Event Nugget (EN), Event Argument Extraction (EAE), Richer Event Descriptions (RED), and Event-Event Relations (EER). The DEFT program seeks to improve state-of-the-art capabilities in automated deep natural language processing, with a particular focus on technologies dealing with inference, causal relationships, and anomaly detection across several languages (DARPA, 2012). The processing of events and event-event relations underlies this focus. The annotation of events and event-event relations is a crucial part of supporting work on these technologies, and a variety of approaches are currently underway.

This paper presents a comparison of event representations, the data annotated, and possible future plans for event annotation. An event annota-

tion experiment is also discussed, in which the same set of sample data was annotated in each of the above representations, with the purpose of being able to compare actual annotation across all of these approaches as directly as possible. The paper walks through a brief example to illustrate the various annotation approaches, and to show the intersections among the various annotated data sets.

2 Types of Event Annotation in the DEFT Program

2.1 Entities, Relations, and Events (ERE)

ERE was developed as an annotation task that would be supportive of multiple research directions and evaluations in the DEFT program, and that would provide a useful foundation for more specialized annotation tasks like inference and anomaly. The ERE tasks include the annotation of entities, relations, and events and their attributes, according to a specific taxonomy, as was also done in Automatic Content Extraction (ACE) (LDC, 2005; Walker et al., 2006).

Light ERE was designed as a lighter-weight version of ACE and a simple approach to entity, relation, and event annotation, with the goal of making annotation easier and more consistent. Light ERE captures a reduced inventory of entity and relation types, with fewer attributes (compared to ACE; for example, only specific entities and actual relations are taggable, and entity subtypes are not labeled). The event ontology of Light ERE is similar to ACE, with slight modification and reduction, and there is strict coreference of events within documents (Aguilar et al., 2014). As in ACE, the annotation of each event mention includes the identification of a trigger, the labeling of the event type, subtype, and participating event argument entities and time expressions. Simplifying from ACE, only attested actual events are annotated (no unrealis events or arguments).

Rich ERE annotation expands on both the inventories and taggability of Light ERE (Song et al., 2015). Rich ERE Entity annotation adds non-specific entities and nominal head marking, in addition to adding a distinction between Location and Facility entity types. Rich ERE Relation annotation doubles the Light ERE ontology to twenty relation subtypes, and also adds future, hypothetical, and

conditional relations. A new category of argument fillers was added for Rich ERE, to allow arguments that are not taggable as entities to be used as fillers for specific relation and event subtypes. For each event mention, Rich ERE labels the event type and subtype, its realis attribute, any of its arguments or participants that are present, and a required “trigger” string in the text. Rich ERE Event annotation includes increased taggability in several areas, compared to Light ERE or ACE event annotation: a slightly expanded event ontology, the addition of generic and other (irrealis) event mentions, the addition of event argument fillers that are otherwise not captured as entities, such as weapon, vehicle, money etc., the addition of argumentless triggers for event mentions, additional attributes for contact and transaction events, double tagging of event mentions for multiple types/subtypes, and multiple tagging of event mentions for certain types of coordination.

In Rich ERE, the concept of Event Hopper was also introduced as a more inclusive, less strict notion of event coreference than that used in Light ERE or ACE. Event hoppers contain mentions of events that are intuitively coreferential to the annotator even if they do not meet the earlier strict event identity requirement, and therefore group events according to a more inclusive coreference specification, which will allow a wider range of event mentions to be coreferential (Song et al., 2015). Event mentions could be placed into the same event hoppers even if they differed in temporal or trigger granularity, their arguments were non-coreferential or conflicting, or if their realis mood differed, as long as they referred to the same event with the same type and subtype. For example, in the following two sentences:

- *The White House didn't confirm Obama's trip {Movement.TransportPerson, Other} to Paris for the Climate Summit last year.*
- *Obama went {Movement.TransportPerson, Actual} to Paris for the Climate Summit.*

Although the realis label differs for the two mentions (Other vs. Actual), if both mentions refer to the same trip to Paris in the document context, they still belong to the same event hopper.

The development of Rich ERE is intended to lay the groundwork for upcoming expansion into the realm of event-event relations, as well as cross-

document and even cross-lingual event representation.

2.2 Event Argument Extraction (EAE) and Event Argument Linking (EAL)

EAE was developed as a track within NIST’s 2014 TAC Knowledge Base Population (KBP) (<http://www.nist.gov/tac/>) evaluation (Freedman et al., 2014). Following the paradigm of Slot Filling and Cold Start, two other KBP evaluation tracks, EAE in 2014 was evaluated by assessment over a pool of participant submissions (Ellis et al., 2014).

A KBP EAE submission consisted of a table of event arguments. For each argument, systems returned the event type/subtype, the argument’s role in the event, a canonical mention of the argument entity (e.g., the most informative name string found for entity), textual justification of the extraction, a realis label (Actual, Generic, Other), and a confidence score. In the 2014 evaluation, the final pool of responses for scoring included a human-produced “manual run” developed by annotators along with up to 5 runs per participating team.

During assessment, annotators judged the correctness of the provided event type, the argument’s assigned role in the event, the (possibly normalized) mention string of the argument entity, and the mention string for the argument entity from the larger provenance connecting it to the event. For correct responses, assessors also provided a realis label (Actual, Generic, Other), a mention type label (Name, Nominal, Other), and created equivalence class clusters. Assessment was performed on responses taken from approximately 250 source documents in 2014, 50 of which were dually assessed.

For dually assessed documents, agreement and kappa were calculated for two conditions (a) over all appropriate tags, (b) collapsing CORRECT and INEXACT (this reflects what is done for the official score). The agreement/kappa numbers only reflect cases where both annotators made an assessment (in some cases, e.g., assessors disagreed on the presence of event type – one assessor would perform an assessment and the other would not).

Numbers before the ‘/’ are over all labels (requiring exact match); numbers after ‘/’ collapse CORRECT and INEXACT. In official scoring, an answer is added to the GS if Realis matches and all of the following are CORRECT/INEXACT: AET,

AER, BF, CAS. CAS Type is used for analyzing output, but is not a part of the scoring process.

AET (Event Presence)

Agreement: 83.92% / 87.54%

Agreement kappa: 64.41 / 69.55

AER (Event Role Presence)

Agreement: 89.75% / 91.16%

Agreement kappa: 58.42 / 61.18

BF (Base Filler)

Agreement: 79.67% / 91.08%

Agreement kappa: 65.54 / 74.52

CAS

Agreement: 76.89% / 84.26%

Agreement kappa: 64.28 / 67.19

CAS Type

Agreement: 80.50%

Agreement kappa: 57.26

Realis

Agreement: 87.18%

Agreement kappa: 61.90

The EAE task (and assessed results) differs in three important ways from the Rich ERE annotation. (1) The EAE task is assessed at the level of ‘entity’ and ‘document-level-event’ and not at the level of the event mention. As such, a system need only find an argument one time, even when an entity’s participation in some event is made explicit in several mentions throughout a document. (2) The EAE task treats as correct a broader set inference about arguments, for example inferring the date/location of an event through general reasoning over the document or inferring participation in an event through group membership. (3) The EAE task does not assess a sub-sentence linguistic trigger/justification of the event. Instead the event type (ET) assessment of EAE task assesses whether or not a sentence length unit justifies the event given a reasonable reader’s interpretation.

In 2015, EAE was extended to require systems to group together arguments participating in the same event. In the updated version of the task, named Event Argument Linking (EAL) and conducted as part of the 2015 TAC KBP evaluations, event arguments were extracted following the same guidelines as EAE, but then had to be grouped together into Event Hoppers, as defined by the Rich ERE task. Performance was measured by comparing system-developed argument clusters against

those created by annotators following assessment, which included all responses provided by systems for each document (Ellis et al., 2015).

To support the 2015 linking task, correct, non-generic assessments from the system submitted output were grouped into argument-sets of hopper-sized granularity. This annotation was performed on 50 sample documents as development data for the 2015 EAL evaluation and on 81 documents in the 2015 EAL test set. The development data overlaps with data for which Rich ERE annotation was performed. The overlapping data set could be used to explore how the differences in annotation procedure lead to differences in decisions about event granularity.

2.3 Event Nugget (EN)

An Event Nugget is a tuple of an event trigger, classification of event type and subtype, and realis attribute. It is similar to an event mention in ERE, but arguments are not labelled. EN annotation in 2014 focused on event nuggets (expanded triggers) only, and followed the same taxonomy of 33 event types and subtypes as Light ERE. However, instead of tagging minimal extent as the trigger, EN allowed multi-word event nuggets (Mitamura et al., 2015). Multi-word event nuggets can be either continuous or discontinuous, and are based on the goal of marking the maximal extent of a semantically meaningful unit to express the event in a sentence. EN also added a realis attribute for each event mention. The realis attribute labels each event as Actual, Generic, or Other. TAC KBP 2014 conducted a pilot evaluation on Event Nugget Detection (END), in which systems were required to detect event nugget tuples, consisting of an event trigger, the type and subtype classification, and the realis attribute.

Relation Type	Inheritance	Expansion	Contingency	Comparison	Temporality	
Subtype (sense)	Reemergence Sub-event Variation Coreference	Confirmation Conjunction Disjunction	Condition Causality	Competition Concession Negation Opposite	Before-After Vagueness Overlap Equality	Start Meet During Finish

Table 1: EER event relation types.

In 2015, TAC KBP ran an open evaluation on EN that was expanded to three evaluation tasks: Event Nugget Detection, Event Nugget Detection and Coreference, and Event Nugget Coreference. Full Event Nugget Coreference is identified when two or more Event Nuggets refer to the same event. EN annotation in 2015 followed the Rich ERE event taxonomy, which added 5 event types and subtypes to make a total of 38 event types and subtypes, and also followed the Rich ERE guidelines on trigger extents, which adopted the minimal extent rule and disallowed discontinuous event triggers (Song et al., 2016). Annotation of Event Nugget Coreference adopted the concept of Event Hopper as in Rich ERE.

2.4 Event-Event Relations (EER)

EER annotation focuses on relations between events in the ERE/ACE taxonomy, both within document and cross-document (Hong et al., 2016). Our general goal is to construct event-centric knowledge networks, where each node is an event and the edges effectively capture the relations between any two events. EER includes five main types of event relations – Inheritance, Expansion, Contingency, Comparison and Temporality – along with 21 sense-based subtypes (or relation senses), as shown in Table 1.

Events involved in a relation play certain roles. For example, an Attack event and an Injure event in a Contingency_Causality will play Cause and Result roles respectively. Figure 1 shows more information about types and roles.

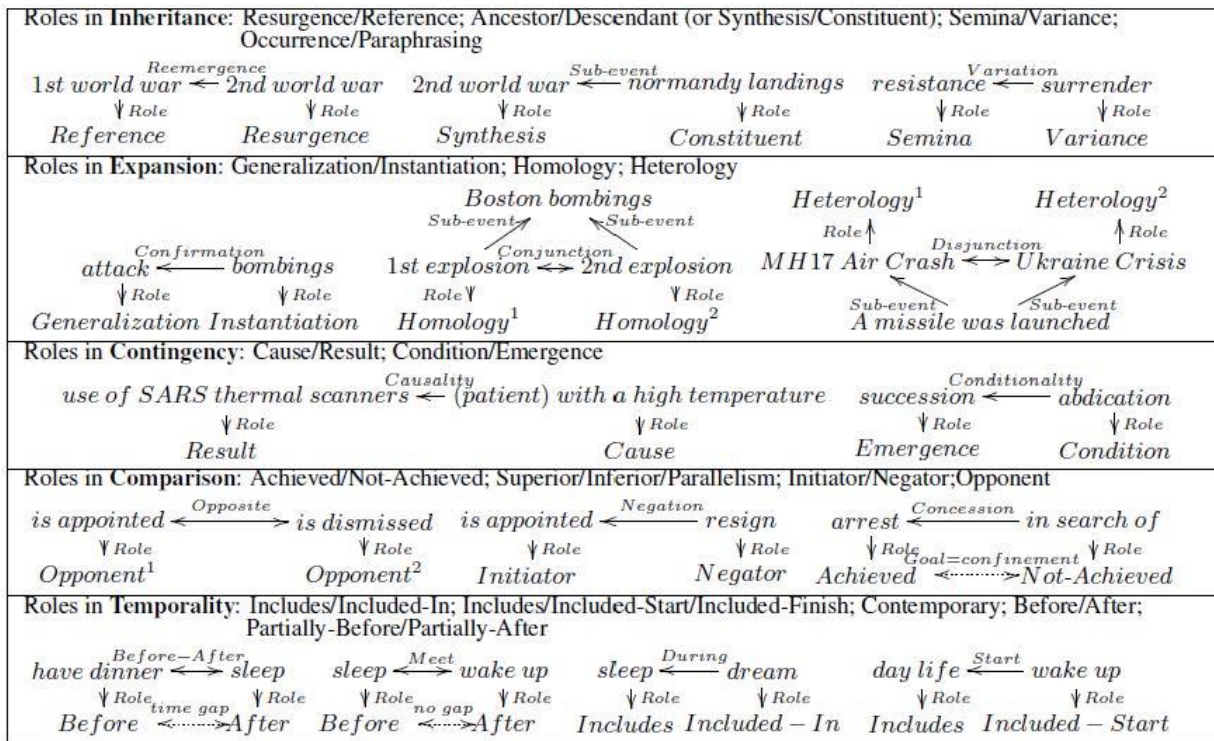


Figure 1: Roles and examples specific to fine-grained event-event relation subtypes.

2.5 Richer Event Descriptions (RED)

RED annotation (Ikuta et al., 2014) marks all events in a document, as well as certain relations between those events. RED combines coreference (Pradhan et al., 2007; Lee et al., 2012) and THYME Temporal Relations annotation (Styler et al., 2014) to provide a thorough representation of entities, events and their relations. The RED schema also goes beyond prior annotations of coreference or temporal relations by also annotating subevent structure, cause-effect relations and reporting relations. Guidelines for RED annotation can be found at <https://github.com/timjogorman/RicherEventDescription/blob/master/guidelines.md>.

3 Annotation Features and Data Annotated

The representation of events and the scope of annotation vary across the different annotation approaches. Table 2 compares the definition of events in each annotation. Table 3 compares how event coreference and event relations are annotated. Table 4 shows the annotated data volume for each annotation schema, completed as of 2015. Annotated data has been distributed to DEFT performers, and will be made available to the wider community as part of the Linguistic Data Consortium (LDC) catalog.

	Event Type and Subtype	Modality/Realis	Arguments	Trigger
Light ERE	8 types 33 subtypes	Actual	Labelled, must include at least one	Minimal span
Rich ERE	9 types 38 subtypes	Actual, Generic, Other	Labelled, but events with no arguments are possible	Minimal span
Event Argument 2014 – 2015	9 types 31 subtypes	Actual, Generic, Other	At least one	Event mentions are not tagged
Event Nugget 2014	8 types 33 subtypes	Actual, Generic, Other	No	Maximal semantic unit
Event Nugget 2015	9 types 38 subtypes	Actual, Generic, Other	No	Minimal span
Event-event relation	8 types 33 subtypes	Actual, Generic, Other	No	Minimal span
RED	Untyped, all predicating events	Actual, Generic, Hypothetical, Uncertain/hedged	No	Single word

Table 2: What is a taggable event mention in each representation.

	Identity coreference	Event-event relations
Light ERE	Strict identity	n/a
Rich ERE	Hopper	n/a
Event Argument 2014	n/a	n/a
Event Argument 2015	Hopper	n/a
Event Nugget 2015	Hopper	n/a
Event-event relation	Strict identity	5 types 21 subtypes
RED	Strict identity, Narrative Containers, Subevent, Set/Member	Aspectual, Causation, Reporting, Temporal

Table 3: Event-event relations in each representation.

	Event Experiment	2014	2015
Rich ERE	52 docs	--	English: 375Kw Chinese:245Kw Spanish: 139Kw
Event Argument	55 docs	519 docs	300 docs
Event Nugget	55 docs	English: 470Kw	English: 210K words
Event-event relation	11 docs	60 docs	79 documents (343 events, 20,235 pairs, 594 related pairs)
RED	55 docs	20 docs	100 docs

Table 4: Volume of data manually annotated for the Event Experiment, for 2014, and for 2015.

4 Comparison: The Event Annotation Experiment

As a result of this diversity in event annotation within DEFT, the Event Working Group which is focusing on events within DEFT has initiated an experiment in which participating teams perform event annotations on the same set of data using all of the different annotation schemas. The goal of the experiment is to make the differences and similarities between the approaches more apparent and to facilitate comparison.

The 50 documents that had been dually assessed for EAE in 2014 were chosen as the data set for the experiment, as both system output and human assessment already exists for the data. Additionally, five newswire documents from the EAE pilot data pool were identified as particularly challenging for event annotation, and so these five documents were also included in the experiment set.

The annotation for this experiment has been completed and released to the DEFT community (LDC, 2015) and will be subsequently published in LDC's catalog, making it available to the broader research community. EN annotation in this experiment adopted the 2014 annotation schema, instead of the 2015 schema. Additionally EAE annotation in the experiment did not group arguments of the same event.

To illustrate the differences of the annotation tasks, below are the annotations in each representation for the following two sentences:

Media tycoon Barry Diller on Wednesday quit as chief of Vivendi Universal Entertainment. Parent company chairman Jean-Rene Fourtou will replace Diller as chief executive of US unit.

Rich ERE, Light ERE, EN, and EAE: Shown in Table 5.

1. Hopper	2. Trigger	3. Type.Subtype	4. Realis	5. Arguments
Hopper 1	<i>quit</i>	Personnel.EndPosition	Actual	<i>Barry Diller chief Wednesday</i>
	<i>replace</i>		Actual	<i>Diller Chief</i>
Hopper 2	<i>replace</i>	Personnel.StartPosition	Other	<i>Jean-Rene Fourtou Chief</i>

Table 5: Event annotation experiment example annotation for Rich ERE (the full table), Light ERE (Hopper 1 only, columns 1-3, 5), EN (columns 1-4), and EAE (columns 1, 3-5).

RED: There are two events in the sentence according to RED. The first predicate, *quit*, is tagged as an Actual Event that occurs BEFORE Document Time. The second, *replace*, is tagged as an Actual Event that is expected to occur AFTER Document Time. RED also labels a causal and temporal relation between the two events, "BEFORE/PRECONDITIONS", showing that the quitting event leads to, but does not directly cause, the replacement, and a temporal CONTAINS relation linking *quit* to Wednesday.

- Event 1: *quit* - BEFORE DOCTIME, Actual Modality
- Event 2: *replace* - AFTER DOCTIME, Actual Modality
- Relation 1: *quit* BEFORE/ PRECONDITIONS *replace*
- Relation 2: *Wednesday* CONTAINS *quit*

Although RED does not annotate the arguments of events, it is intended to be combined with semantic role annotations such as PropBank (Bonial et al., 2014) or AMR (Banarescu et al., 2013), which would provide the argument information. For this example, the *quit* and *replace* events would also be given the predicate argument structures below:

quit.01

Arg0: Media Tycoon Barry Diller
 Arg1: as chief of Vivendi Universal Entertainment
 ArgM-TMP: on Wednesday

replace.01

Arg2: Parent company chairman Jean-Rene Fourtou
 Arg1: Diller
 ArgM-MOD: will
 ArgM-PRD: as chief executive of US unit.

EER: The following events are connected by *Condition* and *Temporality* relations:

- Event 1 (Personnel.EndPosition): *quit*
- Event 2 (Personnel.StartPosition): *replace*

A preliminary analysis of the Rich ERE and Event Argument annotations shows that EAE includes more annotated events, since both non-inferred and inferred events are targeted by EAE, while Rich ERE currently does not annotate inferred events. Initial analysis of Rich ERE and Event Nugget annotation shows that there is considerable overlap in event mentions annotated in both Rich ERE and EN 2014. The main differences are as expected due to the differences in the annotation tasks, namely (1) annotated extents (EN 2014 tags maximal extents of event nuggets to capture the complete semantic unit of the nugget, whereas Rich ERE captures the minimal extent of the event trigger), and (2) double tagging (Rich ERE allows annotation of the same event trigger more than once when the trigger instantiates more than one event type/subtype or in case of conjunction, while EN in 2014 annotated each nugget for a single event type/subtype). To better align with the EAE evaluation as well as the existing Rich ERE annotated data, EN 2015 adopted the Rich ERE guidelines so these differences between Rich ERE and EN 2014 were eliminated.

5 Future Work

In future and on-going work, several of these approaches are coming together in certain aspects for the purpose of conducting event evaluations.

To avoid producing different annotation datasets to support Event Argument and Event Nugget evaluation (as was the case in both 2014 and 2015) and to bring Event Argument and Event Nugget evaluation together to address the goals of extraction and clustering of events for populating a knowledge base, both Event Argument and Event Nugget will use Event annotation in Rich ERE as evaluation data in 2016. Event Argument will be switching to a gold-standard based evaluation for measuring performance in extracting event arguments and within-document grouping of arguments when they participate in the same event. Event Nugget will evaluate event nugget detection and coreference, as in 2015. Both Event Argument and Event Nugget evaluations will be expanded to multiple languages (English, Chinese and Spanish). For both, the set of valid event types will be reduced to the 18 listed below:

Conflict.Attack
 Conflict.Demonstrate
 Contact.Broadcast
 Contact.Contact
 Contact.Correspondence
 Contact.Meet
 Justice.Arrest-Jail
 Life.Die
 Life.Injure
 Manufacture.Artifact
 Movement.TransportArtifact
 Movement.TransportPerson
 Personnel.Elect
 Personnel.EndPosition
 Personnel.StartPosition
 Transaction.Transaction
 Transaction.TransferMoney
 Transaction.TransferOwnership

Additionally, a query- and assessment-based evaluation will also be conducted as a task in Event Argument evaluation to test capabilities in cross-document clustering of events. The size of the evaluation source corpus for Event Argument will increase dramatically (from 500 documents in 2015 to 90,000 documents in 2016).

Finally, the EN team is working on a possible pilot evaluation for ‘event sequencing’ for 2016.

6 Conclusion

The variety of event representations and annotations in the DEFT program seek to tackle the issues of event annotation from multiple directions, and each type of representation brings an aspect of the larger event picture into focus. The comparisons across these representations that will be possible using the annotation from the Event Annotation Experiment will provide on-going useful information as event annotation evolves to meet the goals of the DEFT program and the research community.

Acknowledgments

This material is based on research sponsored by Air Force Research Laboratory and Defense Advanced Research Projects Agency under agreement number FA8750-13-2-0045. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclu-

sions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory and Defense Advanced Research Projects Agency or the U.S. Government.

References

- Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, Joe Ellis. 2014. A Comparison of the Events and Relations Across ACE, ERE, TAC-KBP, and FrameNet Annotation Standards. *ACL 2014: 52nd Annual Meeting of the Association for Computational Linguistics*, Baltimore, June 22-27. *2nd Workshop on Events: Definition, Detection, Coreference, and Representation*.
- L. Banarescu, C. Bonial, S. Cai, M. Georgescu, K. Griffitt, U. Hermjakob, K. Knight, P. Koehn, M. Palmer, and N. Schneider. 2013. Abstract Meaning Representation for Sembanking. *Proc. Linguistic Annotation Workshop*.
- Claire Bonial, Julia Bonn, Kathryn Conger, Jena Hwang and Martha Palmer. 2014. PropBank: Semantics of New Predicate Types. *The 9th edition of the Language Resources and Evaluation Conference*. Reykjavik, Iceland.
- DARPA. 2012. *Broad Agency Announcement: Deep Exploration and Filtering of Text (DEFT)*. Defense Advanced Research Projects Agency, DARPA-BAA-12-47.
- Joe Ellis, Jeremy Getman, Dana Fore, Neil Kuster, Zhiyi Song, Ann Bies, Stephanie Strassel. 2015. Overview of Linguistic Resources for the TAC KBP 2015 Evaluations: Methodologies and Results. In *Proceedings of TAC KBP 2015 Workshop, National Institute of Standards and Technology*, Gaithersburg, Maryland, USA, November 16-17, 2015.
- Joe Ellis, Jeremy Getman, Stephanie M. Strassel. 2014. Overview of Linguistic Resources for the TAC KBP 2014 Evaluations: Planning, Execution, and Results. In *Proceedings of TAC KBP 2014 Workshop, National Institute of Standards and Technology*, Gaithersburg, Maryland, USA, November 17-18, 2014.
- Marjorie Freedman and Ryan Gabbard. 2014. Overview of the Event Argument Evaluation. In *Proceedings of TAC KBP 2014 Workshop, National Institute of Standards and Technology*, Gaithersburg, Maryland, USA, November 17-18, 2014.
- Yu Hong, Tongtao Zhang, Sharone Horowitz-Hendler, Heng Ji, Tim O’Gorman and Martha Palmer. 2016. Building a Cross-document Event-Event Relation Corpus. Submitted.

- Rei Ikuta, William F. Styler IV, Mariah Hamang, Tim O’Gorman, and Martha Palmer, 2014. Challenges of Adding Causation to Richer Event Descriptions. In *Proceedings of the 2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 12–20, Baltimore, Maryland, USA, June 22–27, 2014. © 2014 Association for Computational Linguistics.
- H. Lee, M. Recasens, A. Chang, M. Surdeanu, and D. Jurafsky. 2012. Joint entity and event coreference resolution across documents. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL)*, Jeju Island, 489–500.
- Linguistic Data Consortium. 2005. *ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3*.
- Linguistic Data Consortium. 2015. *DEFT Event Annotation Comparison Experiment VI*. Catalog ID: LDC2015E40, Linguistic Data Consortium.
- Teruko Mitamura, Yukari Yamakawa, Sue Holm, Zhiyi Song, Ann Bies, Seth Kulick, Stephanie Strassel. 2015. Event Nugget Annotation: Processes and Issues. The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015). *3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*.
- S. Pradhan, L. Ramshaw, R. Weischedel, J. MacBride, and L. Micciulla. 2007. Unrestricted Coreference: Identifying Entities and Events in OntoNotes. In *Proceedings of the IEEE International Conference on Semantic Computing (ICSC)*, September 17–19.
- Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant and Xiaoyi Ma. 2015. From Light to Rich ERE: Annotation of Entities, Relations, and Events. The 2015 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2015). *3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*.
- Zhiyi Song, Ann Bies, Stephanie Strassel, Joe Ellis, Teruko Mitamura, Hoa Dong, Yukari Yamakawa, Sue Holm. 2016. Event Nugget and Event Coreference Annotation. The 2016 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL HLT 2016). *4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation*.
- W. F. Styler, S. Bethard, S. Finan, M. Palmer, S. Pradhan, P. C. de Groen, B. Erickson, T. Miller, C. Lin, G. Savova and J. Pustejovsky. 2014. Temporal Annotation in the Clinical Domain, *Transactions of the Association of Computational Linguistics*, 2:143–154.
- Christopher Walker, Stephanie Strassel, Julie Medero, Kazuaki Maeda. 2006. *ACE 2005 Multilingual Training Corpus*. Linguistic Data Consortium, LDC Catalog No.: LDC2006T06.