

# From Light to Rich ERE: Annotation of Entities, Relations, and Events

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant and Xiaoyi Ma

Linguistic Data Consortium, University of Pennsylvania  
3600 Market Street, Suite 801  
Philadelphia, PA, 19104, USA

{zhiyi,bies,strassel,riese,jmott,joellis,jdwright,skulick,nryant,xma}@  
ldc.upenn.edu

## Abstract

We describe the evolution of the Entities, Relations and Events (ERE) annotation task, created to support research and technology development within the DARPA DEFT program. We begin by describing the specification for Light ERE annotation, including the motivation for the task within the context of DEFT. We discuss the transition from Light ERE to a more complex Rich ERE specification, enabling more comprehensive treatment of phenomena of interest to DEFT.

## 1 Introduction

DARPA's Deep Exploration and Filtering of Text (DEFT) program aims 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 (DARPA, 2012). Evaluations within the DEFT program focus on a variety of component technologies, united by a common focus on the problem of populating a knowledge base with information about entities and events and the relationships among them. Given the variety of approaches and evaluations within DEFT, we set

out to define an annotation task that would be supportive of multiple research directions and evaluations, and that would provide a useful foundation for more specialized annotation tasks like inference and anomaly. The resulting Entities, Relations and Events (ERE) annotation task has evolved over the course of the program, from a fairly lightweight treatment of entities, relations and events in text, to a richer representation of phenomena of interest to the program.

While previous approaches such as ACE (Dodington et al., 2004), LCTL (Simpson et al., 2008), OntoNotes (Pradhan et al., 2007), Machine Reading (Strassel et al., 2010), TimeML (Boguraev and Ando, 2005), Penn Discourse Treebank (Prasad et al., 2014), and Rhetorical Structure Theory (Mann and Thompson, 1988) laid some of the groundwork for this type of resource, the DEFT program requires annotation of complex and hierarchical event structures that go beyond any of the existing (and partially-overlapping) task definitions. Recognizing the effort required to define such an annotation task for multiple languages and genres, we decided to adopt a multi-phased approach, starting with a fairly lightweight implementation and introducing additional complexity over time.

In the first phase of the program, we defined Light ERE as a simplified form of ACE annota-

tion, with the goal of being able to rapidly produce consistently labeled data in multiple languages (Aguilar et al., 2014). In Phase 2, Rich ERE expands entity, relation and event ontologies and expands the notion of what is taggable. Rich ERE also introduces the notion of *Event Hopper* to address the pervasive challenge of event co-reference, particularly with respect to event mention and event argument granularity variation within and across documents, thus paving the way for the important goal of creating (hierarchical or nested) cross-document event representations.

In the remaining sections we describe the Light ERE annotation specification and the resources produced under this spec. We discuss the motivation for transitioning from Light ERE to Rich ERE, and present the Rich ERE specification in detail, along with developments in smart data selection and annotation consistency analysis. We conclude with a discussion of annotation challenges and future directions.

## 2 Related Annotation Efforts

A number of previous and current event annotation tasks have influenced the development of Rich ERE, including ACE and several tasks with the TAC KBP Evaluation series. We describe each in turn in the sections that follow.

### 2.1 ACE and Light ERE

At the start of the DEFT program it was necessary to scale up quickly to produce resources for system training and development, and so we looked to existing annotation tasks that were compatible with our desired approach. One such task was ACE (Automatic Content Extraction), designed to benchmark research in information extraction, focusing on entity detection and tracking, relation detection and characterization, as well as event detection and characterization (Doddington et al., 2004; Walker et al., 2006). ACE annotation labels mentions of people, organizations, locations, geopolitical entities, weapons, and vehicles, as well as subtypes for each entity type. ACE also annotates a target set of relations and events between and among those constructs. Multiple mentions of the same entity, relation or event within a document are coreferenced.

Light ERE was designed as a lighter-weight version of ACE (LDC, 2005; Walker et al., 2006) 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 (for example, only specific entities and actual relations are taggable, and entity subtypes are not labeled). Events are labeled following approaches developed in ACE and Machine Reading (Strassel et al., 2010), but adapted for informal genres such as Discussion Forums (DF). The event ontology of Light ERE is similar to ACE, with slight modification and reduction, and events are coreferenced 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. Simplifying from ACE, only attested actual events are annotated (no unrealis events or arguments).

Our Light ERE annotation effort also includes creating fully annotated resources in Chinese and Spanish in addition to English, with a portion of the annotation being cross-lingual. We developed a Chinese-English parallel Light ERE corpus which consists of approximately 100K words of Chinese data along with the corresponding English translation, both annotated in Light ERE. Portions of the parallel data have had other layers of annotation performed on it, particularly Chinese Treebank (CTB) on the Chinese side (Zhang and Xue, 2012) as well as English-Chinese Treebank (ECTB) on the English side (Bies et al., 2014). Light ERE annotation is in progress for Spanish on a dataset which is currently being annotated for Spanish Treebank as well. Multiple levels of annotation, such as ERE and treebank, that are keyed to the same dataset should together provide a resource that is expected to facilitate experimentation with machine learning methods that jointly manipulate the multiple levels.

### 2.2 TAC KBP Event Evaluations

The Text Analysis Conference (TAC) is a series of workshops organized by the National Institute of Standards and Technology (NIST) that was developed to encourage research in natural language processing (NLP) and related applications by providing a large test collection, common evalua-

tion procedures, and a forum for researchers to share their results. Through its various evaluations, the Knowledge Base Population (KBP) track of TAC encourages the development of systems that can match entities mentioned in natural texts with those appearing in a knowledge base and extract novel information about entities from a document collection and add it to a new or existing knowledge base.

In 2014, TAC KBP moved into the events domain with the addition of the Event Argument Extraction (EAE) evaluation, in which systems were required to extract mentions of entities from unstructured text and indicate the roles they played in events as supported by text (Ellis et al., 2014). Additionally, TAC KBP 2014 also 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 (Mitamura et al., 2015).

TAC KBP 2015 EAE and END evaluations both plan to expand the tasks such that event tuples would be grouped together or linked to one another to show event identity, either by linking event arguments that participate in the same event (EAE) or by grouping event nuggets that refer to the same event (END). Such expansion in both evaluations would require identification of event coreference, which is a challenging issue in both ACE and Light ERE. The transition from Light ERE to Rich ERE tackles this challenge with the addition of event hoppers.

### 3 Transition from Light ERE to Rich ERE

The simplified annotation in Light ERE allowed the annotation effort to scale up quickly. As the DEFT program moves toward more sophisticated algorithms and evaluations, the transition to a richer representation of events within the ERE framework becomes necessary. The development of Rich ERE lays the groundwork for upcoming expansion into the realm of event-event relations, as well as cross-document and even cross lingual event representation. Transitioning to Rich ERE requires both developing annotation guidelines for the expanded annotation of events and event arguments and also developing a new annotation tool to handle the new annotation task.

## 3.1 Development of Annotation Guidelines for Rich ERE

### 3.1.1 Expanded Entity Annotation

Rich entity annotation expands many areas of Light annotation starting with a general increase in taggability. Instead of restricting annotation to specific, asserted entities, we have added what ACE called underspecified and generic entities to the scope for Rich ERE annotation. Under the umbrella term “non-specific” (NonSPC), we now capture both underspecified and generic entities, in addition to the specific (SPC) entities that Light ERE already captured. We encountered many discussion forum documents that contained generic language while annotating Light ERE data. Previously, we would deprioritize such documents, but with the inclusion of NonSPC entity tagging in Rich ERE, our range of annotatable documents is much larger.

Some other ACE features that we have revived are nominal head marking and distinguishing between Location and Facility entity types. Instead of marking heads for named and pronominal mentions as required in ACE, heads are manually marked only for nominal mentions in Rich ERE. Since named and pronominal heads are generally exactly the same string of text as the entity mention, their heads do not need to be manually marked separately. However, since the heads of nominal mentions are not trivially derivable, they are manually marked in Rich ERE. Furthermore, Light ERE lumped regions, landforms, buildings, and other structures into the Location entity type. Following ACE and to better align with TAC KBP evaluation tasks<sup>1</sup>, Rich ERE separates the Light ERE Location entity type into Facility as well as Location types. Man-made structures and infrastructure are considered Facilities, while regions, landforms, and other non-descript sites fall under Locations. Examples include (note that the heads of nominal mentions are indicated by underscoring):

- [Tourists]<sub>PER.NOM.NonSPC</sub> always end up at [Love Park]<sub>FAC.NAM.SPC</sub>
- [The last four tourists to show up]<sub>PER.NOM.SPC</sub> missed the bus

In addition, we created a new class called Argument Fillers, which are entity-like participants in

<sup>1</sup> NIST, 2015. <http://www.nist.gov/tac/2015/KBP/>

relations and events that are not annotated at the entity level. Argument fillers are annotated only when they fill argument roles in tagged relations or events. Examples of argument fillers are included in the discussion of relations and events below. Whereas ACE exhaustively tagged weapons and vehicles as entities, Rich ERE captures them as argument fillers. Rich ERE also adds the annotation of commodities as fillers.

Additionally, title entities from Light ERE have been reclassified as argument fillers, because they are only annotated when they can be connected to a named person entity in the relation phase. The full list of argument fillers is Title, Age, URL, Sentence, Crime, Money, Vehicle, Weapon, Commodity, and Time types. Each of these argument fillers corresponds to specific relation or event subtypes, meaning that they will only appear if the corresponding subtype lends itself to such information. For example, a person’s age will only be annotated as an argument filler of a generalaffiliation-personage relation, and a weapon will be annotated only in a limited number of event subtypes, including Conflict.Attack, Manufacture.Artifact, and Life.Injure.

### 3.1.2 Expanded Relation Annotation

Rich ERE relations looked to the TAC KBP Slot Filling Evaluation for inspiration by doubling the ontology from ten subtypes in Light ERE to twenty subtypes in Rich ERE. The KBP Slot Filling task asks annotators to look for textual information that is very similar in scope to ERE annotation. For example, both ERE and KBP Slot Filling annotate material that is based on a person’s employment or membership within an organization, familial relations, and nationality, as well as subsidiary-parent organization relationships and organization location. It was a natural step to expand the ERE relation ontology to incorporate more facets of KBP Slot Filling. Part of this cross-project sync up required the addition of brand new argument fillers for some relation types. Three new subtypes of relations use the argument fillers described above: personalsocial-role (Title), generalaffiliation-orgwebsite (URL) and generalaffiliation-personage (Age). Table 1 shows the newly added relation inventory in Rich ERE as compared with Light ERE.

Light		Rich	
Type	Subtype	Subtype	Type
physical	n/a	orgheadquarter	physical
	located	locatednear	
		resident	
	origin	orglocationorigin	generalaffiliation
member-origin-religion-ethnicity (more)			
	n/a	personage	generalaffiliation
	n/a	orgwebsite	
	n/a	org-political-religious-affiliation (opra)	
partwhole	subsidiary	subsidiary	partwhole
personalsocial	membership	membership	personalsocial
	business	business	
	family	family	
	unspecified	unspecified	
	role	role	
orgaffiliation	employment-membership	employmentmembership	orgaffiliation
	leadership	leadership	
	n/a	investshareholder	
	n/a	studentalum	
	n/a	ownership	
	n/a	founder	

Table 1: ERE Relation Taxonomy in Light and Rich

Finally, while Light ERE only annotated attested, asserted relations, Rich ERE annotates future, hypothetical, and conditional (but not negated) relations as well. All relations are assigned a realis attribute of “Asserted” vs. “Other” to mark this distinction. Examples of these additions and changes can be seen below:

- Now [53]<sub>AGE.ARG</sub>, [Barack Obama]<sub>PER.NAM.SPC</sub> signed important documents this morning. (General-Affiliation.PER-Age, Realis: Asserted)
- [[Spanish]<sub>GPE.NAM.SPC</sub> students]<sub>PER.NOM.SPC</sub> gathered to protest the growing cost of tuition. (General-Affiliation.MORE, Realis: Asserted)

- [She]<sub>PER.PRO.SPC</sub> has been living in [California]<sub>GPE.NAM.SPC</sub> for three years now. (Physical.Resident, Realis: Asserted)
- [He]<sub>PER.PRO.SPC</sub> may end up in [New York]<sub>GPE.NAM.SPC</sub>. (Physical.Located-Near, Realis: Other)

### 3.1.3 Expanded Event Annotation

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<sup>2</sup>, compared to Light ERE Event annotation: a slightly expanded event ontology, the addition of generic and other (irrealis) event mentions, 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.

#### A. Expansion of event ontology, and additional attributes for Contact and Transaction events

Rich ERE adds one new event type (Manufacture) to the Light ERE inventory of event types. The complete list of event types is as follows: Life, Movement, Business, Conflict, Contact, Personnel, Transaction, Justice, Manufacture. The Manufacture event type has only one subtype (Manufacture.Artifact), and can have the following arguments: agent, patient (weapon, facility, vehicle, commodity), time and location. For example,

- [China]<sub>AGENT</sub> is reportedly **constructing** [a second aircraft carrier]<sub>PATIENT.VEHICLE</sub>
- [the Imboulou hydroelectric power station]<sub>PATIENT.FACILITY</sub>, which was **constructed** by [Chinese technicians]<sub>AGENT</sub>

In addition to the new event type, Rich ERE adds several new event subtypes to already existing event types: Movement.Transport-Artifact, Contact.Broadcast, Contact.Contact, Transaction.Transaction.

The Movement.Transport-Artifact subtype can take weapon, vehicle, facility, or commodity as the patient. For example,

- [122 kilos of heroin hidden in a truck]<sub>ARTIFACT.COMMODITY</sub> which was set to **cross** into [Greece]<sub>DESTINATION.GPE</sub>
- [the cans of marijuana]<sub>ARTIFACT.COMMODITY</sub> were **launched** about 500 feet into the [U.S.]<sub>DESTINATION.GPE</sub> using [a pneumatic-powered cannon]<sub>INSTRUMENT.WEAPON</sub>

Contact event mentions are now labeled with attributes to describe Formality (Formal, Informal, Can’t Tell), Scheduling (Planned, Spontaneous, Can’t Tell), Medium (In-person, Not-in-person, Can’t Tell), and Audience (Two-way, One-way, Can’t Tell). Contact event subtypes are determined (automatically) based on the annotated attributes:

- Contact.Meet: Medium attribute must be “In-person” and audience attribute must be “Two-way”
- Contact.Correspondence<sup>3</sup>: Medium attribute must be “Not-in-person” and audience attribute must be “Two-way”
- Contact.Broadcast: Any Contact event mention where the audience attribute is “One-way”
- Contact.Contact: Used when no more specific subtype is available, and occurs when either the medium or audience attribute is “Can’t Tell”

Contact.Meet and Contact.Correspondence as subtypes are unchanged from Light ERE, but Contact.Broadcast and Contact.Contact are new subtypes in Rich ERE.

Note that that the Formality and Scheduling attributes are annotated for all Contact event mentions, but these attributes have no effect on the subtype determination.

Transaction.Transaction is a new subtype added to indicate cases where it is clear that a transaction event is mentioned, but it is not clear in context whether money or a commodity is being transferred. For example,

- I **received** a gift (Transaction.Transaction)

#### B. Addition of generic and other irrealis event mentions

In order to align ERE annotation more closely with the current EAE and END tasks, Rich ERE annotates a Realis attribute for each event mention.

<sup>2</sup> Changes to coreference in Rich ERE are discussed below, in section 3.1.4.

<sup>3</sup> The Contact.Correspondence subtype is simply the new name for the subtype called Contact.Communication in Light ERE.

This is in sync with both EAE and END and is also compatible with ACE annotation.

The realis attributes are Actual (asserted), Generic (generic, habitual), and Other (future, hypothetical, negated, uncertain, etc.). Previously Light ERE annotation was restricted to Actual event mentions only.

- Actual: He **emailed** her about their plans
- Other: Saudi Arabia is scheduled to begin **building** the world’s tallest tower next week
- Generic: Turkey is a popular passageway for drug smugglers **trafficking** from south Asia to Europe

The realis of the relationship between each argument and the event mention will also be tagged, separately from the realis of the event mention itself. For example,

- [+irrealis] “Jon” as the agent for the asserted Conflict.Attack event: [Jon] denied [he] master-minded the **attack**

### C. Addition of argumentless triggers for event mentions

Unlike Light ERE, Rich ERE will allow the annotation of event mention triggers even when there are no arguments or participants of the event present in the text. This additional annotation will allow Rich ERE to align more closely with END (Mitamura et al., 2015).

### D. Double tagging of event mentions for multiple types/subtypes

Rich ERE will permit double tagging of event triggers to allow obligatory inferred events that are in the ERE event taxonomy to be tagged. For example, if both money and ownership are transferred in a Transaction event, then the event mention should be tagged twice, once for each subtype:

- I **paid** \$7 for the book (tagged as both Transaction.TRANSFER-OWNERSHIP, and Transaction.TRANSFER-MONEY)

The triggers that can be annotated this way are restricted to triggers that clearly indicate more than one event type or subtype in context. For example,

- Conflict.Attack and either Life.Injure or Life.Die: murder, victim, decapitate, kill
- Transaction.Transfer-Money and Transaction.Transfer-Ownership (money being exchanged for an item): buy, purchase, pick up

- Legal language that might trigger multiple Justice Events or other Event Types: guilty plea, execution (Life.Die / Justice. Execute), death penalty, testimony (Justice.Trial-Hearing, Contact.Meet)

In a change from Light ERE, event triggers may be the same string of text as an entity or the same string of the head of a NOM entity mention. Event triggers that are nested within an entity mention are also acceptable.

- The situation escalated and the [**murderer**] fled the scene. (This is an event trigger, even though “murderer” would already be a nominal PER entity.)
- The mayor agreed to meet with [angry **protestors**]. (This is a trigger, even though “protestors” would already be the head of a nominal PER entity.)
- [The one who **divorced** me] only thinks of himself. (Here “divorce” can be a trigger for a Life.DIVORCE event, even though it is nested within a longer PER entity and it is not the head noun.)

### E. Multiple tagging of event mentions for certain types of coordination

Rich ERE will also allow a single trigger to be tagged multiple times in cases where multiple events are indicated through coordination of arguments. The argument role that is coordinated determines whether a single event mention or multiple event mentions are tagged:

- If the TIME or PLACE role is coordinated or if there are separate times and places indicated, then multiple events are tagged.
- If any other argument role is coordinated, a single event is tagged. In this case, each of the coordinated arguments will be tagged separately as an argument of the event mention, and the result will be a single event with multiple arguments tagged for the coordinated argument role.

If the context or the language is too complicated to sort out the number of events, annotators are instructed to default to annotating a single event with multiple arguments.

In this example, there are two Conflict.Attack events, and two Life.Die events triggered by “murder”, because the TIME argument is different:

- Cipriani was sentenced to life in prison for the **murder** of Renault chief George Besse in 1986 and the head of government arms sales Rene Audran a year earlier
  - Conflict.Attack: Trigger = murder, agent = Cipriani, victim = George Besse, time = 1986
  - Conflict.Attack: Trigger = murder, agent = Cipriani, victim = Rene Audran, time = a year earlier
  - Life.Die: Trigger = murder, argument = George Besse, agent = Cipriani, time = 1986
  - Life.Die: Trigger = murder, argument = Rene Audran, agent = Cipriani, time = a year earlier

In the following example, only one event is tagged, with multiple giver arguments and multiple recipient arguments:

- China and the US are the biggest **lenders** to Brazil and India
  - Transaction.Transfer-Money: Trigger = lenders, giver = China, giver = US, recipient = Brazil, recipient = India

### 3.1.4 Event Hoppers and Event Coreference

In Light ERE as well as ACE, event coreference was limited to strict event identity. Following component judgments, annotators marked two events as coreferential in Light ERE if they had the same agent(s), patient(s), time, and location. However, there are many event mentions that annotators intuitively feel are the same that do not meet the strict event identity standard and therefore would not be coreferential in Light ERE or ACE. Some events might have been inconsistently marked as coreferential because of the conflict between the annotators' intuitive judgment and the strict identity coreference standard.

In Rich ERE, we instead introduce the concept of *Event Hopper* as a more inclusive, less strict notion of event coreference. Event hoppers contain mentions of events that “feel” coreferential to the annotator even if they do not meet the earlier strict event identity requirement. More specifically, features of event mentions that go into the same hopper are

- They have the same event type and subtype (exceptions to this are Contact.Contact and Transaction.Transaction mentions, which

can be added to any Contact or Transaction hopper, respectively)

- They have the same temporal and location scope, though not necessarily the same temporal expression or specifically the same date (*Attack in Baghdad on Thursday* vs. *Bombing in the Green Zone last week*)
- Trigger granularity can be different (*assaulting 32 people* vs. *wielded a knife*)
- Event arguments may be non-coreferential or conflicting (*18 killed* vs. *dozens killed*)
- Realis status may be different (*will travel [OTHER] to Europe next week* vs. *is on a 5-day trip [ACTUAL]*)

Every tagged event mention will be put into an event hopper in Rich ERE, and all tagged event mentions that refer to the same event occurrence will be grouped into the same event hopper.

Event hoppers will allow annotators to group together more event mentions and therefore also label more event arguments in Rich ERE. This richer annotation will lead to a more complete knowledge base and better support for the Event Argument Linking and END evaluations in 2015, when one of the goals is to evaluate event identity.

### 3.2 Development of an Annotation GUI for Rich ERE

The Rich ERE annotation tool was developed following the framework described in Wright et al. (2012), allowing for rapid development of a new interface for Rich ERE. Numerous features were included “for free” in that they were developed for previous interfaces, and therefore required no additional development time. One important example of this is the representation of annotated text extents with underlines that can overlap arbitrarily, be color coded based on other annotations (e.g., entity type), and allow the user to click to navigate among the annotations. An important feature developed specifically for the Rich ERE tool is a “reference annotation”, which is essentially one widget pointing to another. Once a complete set of annotations for a mention or entity has been done, a single annotation can be used to plug them as a whole into relation or event arguments, but referentially, allowing the original annotations to be safely changed. In addition, annotation managers had an important role in development of the tool beyond specification, as there is an editor that

grants direct access to the database where the interface is defined. Managers can add widgets, change them (e.g., add menu choices), and even specify logical constraints between the annotations (e.g., a “resident” relation must take a “person” argument).

#### 4 Linguistic Resources Labeled for ERE

To date we have released approximately 570,000 words of English Light ERE data, including both NW and DF, plus 200,000 words of Chinese DF. Another 100,000 words of Spanish Light ERE data is currently in progress and is expected to be completed in the coming weeks. Rich ERE annotation in English is also currently underway, with 32,420 words (91 documents) completed to date. We expect to complete another 170,000 words of English and 100,000 words in each of Chinese and Spanish within the next several weeks. A portion of the Rich ERE data is new, while the remainder has previously been annotated for Light ERE. Details for each language, genre and task are provided in Table 2 below. The ERE data is currently available to DEFT and TAC KBP performers and will also be published in LDC’s catalog in future, making it available to the research community at large.

	Genre	English	Chinese	Spanish
<b>Light ERE</b>	<i>NW</i>	220Kw	--	50Kw
	<i>DF</i>	350Kw	200Kw	50Kw
<b>Rich ERE</b>	<i>NW</i>	24Kw	--	50Kw
	<i>DF</i>	175Kw	100Kw	50Kw

Table 2: Existing and Planned ERE Resources

The overall target for this phase of DEFT is to complete 400Kw of Rich ERE annotation per language on English, Chinese and Spanish data. 100Kw each from Spanish and Chinese will be parallel to Rich ERE annotation on English translations of the same data. We expect the annotation goal to be met by the end of this year.

#### 4.1 Smart Data Selection

In an attempt to minimize annotator effort on documents with insufficient content, documents were fed into the annotation pipeline in descending order of event trigger density, defined as the number of event triggers per 1,000 tokens. Triggers were automatically tagged using a deep neural network based tagger trained on the ACE 2005 annotations (Walker et al., 2006) with orthographic and word

embedding features. The word embeddings were trained using word2vec (Mikolov et al., 2013) on several billion words of newswire and discussion forum data. Preliminary results using this selection process have been very encouraging, with annotators reporting much richer documents on average, compared to the prior approach in which no ranking was imposed.

#### 4.2 Rich ERE Challenges and Next Steps

One of the challenges in event annotation is to determine the level of granularity that will be distinguished as sub-event vs. event hopper. We observed this issue in our pilot Rich ERE annotation, and the goal is to have sub-event annotation be a relationship between event hoppers in the future. In order to represent the relations between event hoppers, we are planning the addition of a notion such as Narrative Container (Pustejovsky and Stubbs, 2011) to capture non-identity event-event relations such as causality, part-whole, precedence, enablement, etc. Event hoppers will serve as a level between individual event mentions and Narrative Containers. Event hoppers will be grouped into Narrative Containers, and so relations will be between event hoppers, instead of between individual event mentions. More specific relations between individual event mentions can then be derived from the event-event relations between the event hoppers within narrative containers or from relations between narrative containers.

#### 4.3 Inter-Annotator Agreement

Work on inter-annotator agreement (IAA) will be based on the method outlined in Kulick et al. (2014), which described a matching algorithm used at each level of the annotation hierarchy, from entity mentions to events. This work focused on the evaluation for entity, relation, and event mentions, as well as for entities overall. The algorithm for entity mention mapping is based on the span for an entity mention, while the mapping for relation and event mentions is more complex, based on the mapping of the arguments, which in turn depends on the entity mention mapping. IAA work will be conducted on dual annotation for Rich ERE. Analysis will be reported in the future.



## 5 Conclusion

Rich ERE annotation includes a more comprehensive annotation of entities, relations and events, including expanded taggability, expanded categories, annotation for realis and specificity, and expanded coreference with the event hopper level. The expansion and change will populate more information to a knowledge base. Looking to the future, the additions to Rich ERE, particularly expanded taggability and the looser coreference of the event hopper level, are expected to improve support of within-document event-event relations and eventually cross-document and cross-lingual annotation.

Event Hoppers group events according to a more inclusive coreference specification, which will allow a wider range of event mentions to be coreferential. This is closer to the real world situation in which the same event is often referred to in a variety of ways that cannot meet a strict identity standard as was used in ACE and Light ERE. This kind of more inclusive event coreference will be increasingly necessary as work on informal genres, cross-document, and cross-lingual data is desired. In addition, event hopper annotation will allow knowledge base population to draw from a broader grouping of coreferenced event mentions, allowing for a more complete representation of event slots.

## 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 conclusions 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*.

Ann Bies, Justin Mott, Seth Kulick, Jennifer Garland, Colin Warner. 2014. Incorporating Alternate Translations into English Translation Treebank. In *Proceedings of the 9th Edition of the Language Resources and Evaluation Conference (LREC 2014)*, Reykjavik, May 26-31.

Branimir Boguraev and Rie Kubota Ando. 2005. TimeML-Compliant Text Analysis for Temporal Reasoning. In *Proceedings of International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 997--1003.

DARPA. 2012. *Broad Agency Announcement: Deep Exploration and Filtering of Text (DEFT)*. Defense Advanced Research Projects Agency, DARPA-BAA-12-47.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, Ralph Weischedel. 2004. The Automatic Content Extraction (ACE) program- tasks, data, and evaluation. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, May 24-30.

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*.

Seth Kulick, Ann Bies, Justin Mott. 2014. Inter-annotator Agreement for ERE Annotation. ACL 2014: 52nd Annual Meeting of the Association for Computational Linguistics, Baltimore, June 22-27. *2nd Workshop on EVENTS: Definition, Detection, Coreference, and Representation*.

Linguistic Data Consortium. 2005. *ACE (Automatic Content Extraction) English Annotation Guidelines for Events Version 5.4.3*.

William C. Mann and Sandra Thompson. 1988. Rhetorical Structure Theory: Toward a Functional Theory of Text Organization. *Text* 8(3) (243-281).

Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. In *ICLR*, 2013.

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*.
- Sameer S. Pradhan, Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, Ralph Weischedel. 2007. OntoNotes: A Unified Relational Semantic Representation. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC-07)*. Irvine, CA.
- Rashmi Prasad, Bonnie Webber, Aravind Joshi. 2014. Reflections on the Penn Discourse TreeBank, Comparable Corpora and Complementary Annotation. *Computational Linguistics, Vol. 40, No. 4*, pp. 921-950, doi:10.1162/COLI\_a\_00204.
- James Pustejovsky, and Amber Stubbs. 2011. Increasing Informativeness in Temporal Annotation. *Linguistic Annotation Workshop V (LAW-V)*, Portland, Oregon. ACL, June 2011.
- Heather Simpson, Christopher Cieri, Kazuaki Maeda, Kathryn Baker, Boyan Onyshkevych. 2008. Human Language Technology Resources for Less Commonly Taught Languages: Lessons Learned Toward Creation of Basic Language Resources. In *Proceedings of SALTMIL Workshop: Free/Open-Source Language Resources for the Machine Translation of Less-Resourced Languages*, at LREC 2008: 7th International Conference on Language Resources and Evaluation, Marrakech, May 28-30.
- Stephanie Strassel, Dan Adams, Henry Goldberg, Jonathan Herr, Ron Keesing, Daniel Oblinger, Heather Simpson, Robert Schrag, Jonathan Wright. 2010. The DARPA Machine Reading Program - Encouraging Linguistic and Reasoning Research with a Series of Reading Tasks. In *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC 2010)*, Valletta, May 17-23.
- Christopher Walker, Stephanie Strassel, Julie Medero, Kazuaki Maeda. 2006. *ACE 2005 Multilingual Training Corpus*. Linguistic Data Consortium, LDC Catalog No.: LDC2006T06.
- Jonathan Wright, Kira Griffitt, Joe Ellis, Stephanie Strassel, Brendan Callahan. 2012. Annotation Trees: LDC's Customizable, Extensible, Scalable Annotation Infrastructure. In *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)*, Istanbul, May 21-27.
- Xiuhong Zhang and Nianwen Xue. 2012. Extending and Scaling up the Chinese Treebank Annotation. In *Proceedings of the 2nd CIPS-SIGHAN Joint Conference on Chinese Language Processing (CLP-2012)*, Tianjin, China.