AFRL-RI-RS-TR-2019-033

REORIENT: RESOURCES FOR OPERATIONALLY RELEVANT INFORMATION EXTRACTION FROM NON-EXPLICIT TEXT

UNIVERSITY OF PENNSYLVANIA

FEBRUARY 2019

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2019-033 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
EDWARD DEPALMA
Work Unit Manager

/ S /
TIMOTHY A. FARRELL
Deputy Chief, Information Intelligence
Systems and Analysis Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEBRUARY 2019	FINAL TECHNICAL REPORT	NOV 2012 – MAY 2018

4. TITLE AND SUBTITLE

REORIENT: RESOURCES FOR OPERATIONALLY RELEVANT INFORMATION EXTRACTION FROM NON-EXPLICIT TEXT

5a. CONTRACT NUMBER
N/A

5b. GRANT NUMBER
FA8750-13-2-0045

5c. PROGRAM ELEMENT NUMBER
62303E

6. AUTHOR(S)

Stephanie Strassel, Martha Palmer, Kevin Knight, Kira Griffitt, Jeremy Getman, Jennifer Tracey, Ann Bies, and Zhiyi Song

5d. PROJECT NUMBER
DEFT

5e. TASK NUMBER
12

5f. WORK UNIT NUMBER
08

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Pennsylvania
Linguistic Data Consortium
3600 Market Street, Suite 810
Philadelphia, PA 19104

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RIEA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2019-033

12. DISTRIBUTION AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The ReORIENT (Resources for Operationally Relevant Information Extraction from Non-explicit Text) research effort has developed a set of linguistic resources to support deep natural language understanding in the context of a diffuse, diverse and large community of researchers and stakeholders. To support Relational Analysis research we have developed data sets labeled for entities, relations, events, and AMR sembanking. To support Anomaly Analysis research we have developed resources labeled for sentiment and belief-based and event-based phenomena. To support Smart Filtering research we have created data sets labeled for textual entailment and inference. A total of 240 distinct data sets were developed under this effort and distributed to DEFT performers during the program. These resources have been consolidated into 34 corpora that have been or will soon be published in LDC's public catalog, making DEFT data available to the wider research community, thus amplifying the government's investment in linguistic data and stimulating relevant research outside of the program.

15. SUBJECT TERMS

Linguistic Resources, Annotation, Information Extraction, Corpora

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**EDWARD DEPALMA**
U	U	U	UU	33	19b. TELEPHONE NUMBER (*Include area code*) **NA**

TABLE OF CONTENTS

LIST OF TABLES

1.0 SUMMARY

Our research efforts have focused on developing a set of resources to support deep natural language understanding in the context of a diffuse, diverse and large community of researchers and stakeholders.

Primary activities included:

- Collection and creation of appropriate source data from naturally-occurring, real-world sources including both formal and informal/conversational data in multiple languages, domains and modalities. Collected data was tailored to be suitable for various kinds of human annotation tasks to support multiple research thrusts.
- Produce core resources with rich annotations to support algorithm development with an emphasis on whole document understanding for knowledge base population. To support Relational Analysis research, annotate entities, relations, events, and AMR sembanking. To support Anomaly Analysis research, develop resources labeled for sentiment and belief-based and event-based phenomena. To support Smart Filtering research, create data sets labeled for textual entailment and inference.
- Engage in annotation research to gain understanding of complex phenomena and ensure compatibility and efficiency across closely related tasks.
- Support DEFT evaluations by providing source data, annotations and assessment of system (and human) output.

A total of 240 distinct data sets were developed under this effort and distributed to DEFT performers during the program. These resources have been consolidated into 34 corpora that have been or will soon be published in LDC's public catalog, making DEFT data available to the wider research community thereby amplifying the government's investment in linguistic data and stimulating relevant research outside of the program. Our research goals have been substantially satisfied, with some reprioritization of effort after the first year to add greater emphasis on whole document understanding for knowledge base population, in keeping with DARPA's overall direction for the program.

2.0 INTRODUCTION

The overarching research goal has been to create a set of resources to support deep natural language understanding in the context of a diffuse, diverse and large community of researchers and stakeholders. Our primary research objectives were as follows:

(1) obtain or create informal, conversational real world data containing phenomena of interest in 3 languages and diverse range of genres including newswire, proxy reports, discussion forum and informal web text from a variety of existing sources of data and from new collection where required;
(2) produce core resources with rich annotations to support algorithm development with an emphasis on whole document understanding for knowledge base population
(3) create data processing/formatting standards and tools to support collection and processing of data
(4) develop methods and tools for annotation
(5) develop training and evaluation datasets to benchmark DEFT system performance
(6) conduct annotation research to gain understanding of complex phenomena and ensure compatibility and efficiency across closely related human annotation tasks
(7) support DEFT evaluations by providing source data, annotations and assessment of both system and human output

The project was designed to enable improved performance for DEFT systems by providing the volume, variety and quality of data required for system training, as well as data required to benchmark system performance and evaluate results. Additionally, the project was expected to yield a substantial number of new linguistic resources with durable value to the worldwide Human Language Technology research community (as well as the end users who benefit from HLT). Lasting linguistic resources amplify the value of the government's investment in data by enabling novel research at sites who are not funded under the program.

3.0 METHODS, ASSUMPTIONS AND PROCEDURES

Specific approaches included collection and/or creation of appropriate source data including spoken and written informal, conversational, naturally-occurring data sources containing phenomena of interest, suitable for annotation in multiple tasks. We proposed to produce "core resources" with rich annotations to support algorithm development for all three primary DEFT research areas. Relational Analysis core resources include labeling of entities, relations, events, and Abstract Meaning Representation (AMR) sembanking; Anomaly Analysis core resources label belief-based and event-based phenomena of interest in spoken and text data; and Smart Filtering core resources include textual entailment and inference annotation. We further proposed to undertake annotation research to gain understanding of complex (primarily event-based) phenomena and ensure compatibility and efficiency across closely related tasks.

4.0 RESULTS AND DISCUSSION

In the sections that follow we present results by major area of activity.

4.1 Entity, Relation and Event Annotation

The Entities, Relations and Events (ERE) annotation task was designed to support multiple research directions and evaluations, and to provide a useful foundation for more narrowly focused annotation tasks like inference and anomaly. ERE evolved over the course of the program from a lightweight treatment of entities, relations and events in text, to a richer ontology and representation particularly for events and event co-reference, including cross-document coreference in the final year of DEFT. We also defined the “event hopper” annotation construct that became central to supporting event coreference in the context of knowledge base population. ERE was also directly integrated into the TAC KBP evaluations as the baseline for gold standard references to support the Entity Detection and Linking, Event Argument, Event Nugget BeSt (belief and sentiment) evaluation tracks. In total we produced over 2Mw of ERE labeled data, spanning 45 corpora.

The Events Working Group was established as a group of event annotation experts from the DEFT community, led by Martha Palmer (Colorado). The goals were (1) biweekly working meetings to discuss events annotation issues encountered in real DEFT data, resulting in concrete recommendations, (2) collaborative annotation and analysis of events in a small corpus of DEFT data, and (3) identifying challenge areas for development and discussion at an ongoing series of Event/EventStory Workshops co-located at conferences in the community.

Richer Event Descriptions (RED) annotation marks all events in a document, as well as certain relations between those events. RED combines coreference and THYME Temporal Relations annotation 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 sub-event structure, cause-effect relations and reporting relations.

The Rich Event Ontology grew out of an effort to provide an independent conceptual backbone to unify existing semantic role labeling schemas and augment them with event-to-event causal and temporal relations. By unifying the FrameNet, VerbNet, Automatic Content Extraction, and Rich Entities, Relations and Events (Rich ERE) resources, the ontology serves as a shared hub for the disparate annotation schemas and therefore enables the combination of semantic role labeling training data into a larger, more diverse corpus.

4.2 Abstract Meaning Representation (AMR)

AMR utilizes Propbank frames and special AMR concepts to capture whole-sentence meanings by annotating connections between the more contentful and functional semantic elements of a sentence in a tree structure. During DEFT, AMR moved from more formal genres (e.g. newswire) to informal ones (e.g. discussion forums), added support for KB-level information by providing wikifications for a wide variety of named-entity types, and moved toward whole-document understanding by linking entities, events, and implicit arguments across a subset of AMRs. We produced a total of 61,547 AMRs, 8,156 of which contain multi-sentence links.

4.3 Text Entailment and Inference

We produced English and Chinese text entailment and inference annotations to support smart filtering research, resulting in over 20,000 English and 30,000 Chinese entailment/contradiction judgments and inference judgments for over 1200 English text-hypothesis pairs. However, these resources were not widely used given the program's emerging focus on whole document understanding for knowledge base population, and the task was eliminated after Phase 1 in favor of greater effort toward ERE, KBP and BeSt.

4.4 Anomaly Analysis

The BeSt task developed out of the Anomaly track and evolved over the course of the program to permit performers working on belief and sentiment to connect their task to the work being done by those working on entities, relations, and events. The task started with exhaustive committed belief annotation on every proposition, creating a dataset much larger than any other available annotation related to belief. From there, LDC and performers worked together to develop a new annotation task that added sentiment and limited the targets of belief and sentiment to entities, relations, and events annotated in ERE. This provided a connection between the BeSt track and other DEFT tracks, and allowed for evaluation of belief and sentiment detection using both gold standard ERE annotation for diagnostic purposes as well as system output from other DEFT systems for a realistic measure of performance within an end-to-end system. In all we produced nearly 534,000 belief/sentiment annotations across English, Spanish and Chinese data sets.

4.5 Narrative Text

While the first year of DEFT included creation of Narrative Text documents called Proxy Reports, which consisted of information from newswire articles transformed into something like an analyst report, these were dropped along with work on speech data, in favor of naturally-occurring formal and informal news-oriented text.

4.6 TAC KBP Data

Although not part of our original proposal, production of training and test data to support the annual evaluation of DEFT technology under the NIST TAC Knowledge Base Population evaluation series became a major focus of our work, in keeping with DARPA

priorities. KBP data added greater complexity over time, with more integration of distinct evaluation tracks and data sets, a greater emphasis on multilingual data for all tracks, an ontology more tailored to DEFT stakeholder interests, and an increasing focus on events (as well as entities and slots).

In 2013 we expanded Slot Filling to include data both Sentiment Slot Filling and Temporal Slot Filling. In 2014 we added data for two event-focused tracks Event Argument Linking and Event Nugget, and expanded Entity Linking annotation to cross-document entity extraction and clustering for English (Entity Discovery and Linking (EDL)); Chinese and Spanish EDL followed in 2015, which meant cross-document clustering was necessarily also cross-lingual. In 2016 we made the move to using the same source corpus for all KBP evaluation tracks, and we expanded Cold Start from monolingual English to cross-lingual English/Chinese/Spanish; we also added new information about entities, relations, and events with beliefs and sentiment. In 2017, Cold Start was further expanded to include event and sentiment slots, making Cold Start very nearly the sum total of all component KBP evaluations, testing extraction and clustering of entities, relations, events, and sentiment.

We produced nearly 150 distinct KBP corpora, comprising close to 150,000 queries, 84,000 manual runs and over 310,000 assessments.

4.7 Other

To support improved performance for core NLP algorithms that serve as a precursor to DEFT technology (e.g. Spanish and Chinese parsing), we also produced a handful of narrowly defined DEFT data sets at the suggestion of performers and DARPA. We also produced additional data sets at DARPA's request to support cross-project research. These additional resources are described below.

DEFT Spanish Treebank was developed in collaboration with the Language and Computation Center (CLiC), University of Barcelona. The treebank consists of over 100Kw (54,394 tokens of international Spanish newswire text and 55,307 tokens of Latin American Spanish discussion forum text), and was part-of-speech tagged, lemmatized and syntactically annotated with constituents and functions. Newswire data was selected from Spanish Gigaword Third Edition while discussion forum data was selected from new Latin American Spanish discussion forum collection. Syntactic annotation followed annotation specifications for the AnCora treebank, with additional guidelines to support annotation of informal web text. The resulting corpus is the first Latin American Spanish corpus of non-standard language to be morphologically and syntactically annotated. It is a valuable linguistic resource that can be used for the training and evaluation of parsers and part-of-speech taggers.

Gender, number, and animacy are lexical indicators that can be useful in the detection of person mentions. LDC created the Chinese Lexical Resources for Gender, Number and Animacy Corpus by extracting information from newswire texts in the Chinese Gigaword Corpus. The corpus includes dictionaries of Chinese animate nominals and names; Chinese nominals and name with gender and number predicted; and other

dictionaries of Chinese nominals, names, verbs and pronouns; each dictionary contains frequency information as well as the features in question. The data was released to DEFT performers with an eye to improve performance of core NLP capabilities such as entity tagging.

We created additional richly annotated language resources to support development and evaluation of knowledge base population technology with a particular focus on data associated with a particular incident (e.g. a particular series of attacks or disasters). We completed incident-specific data collection, prepared a reference knowledge base based on existing resources augmented for incident-specific entities, and created labeled data for three languages (IL5/Tigrinya, IL9/Kinyarwanda and IL10/Sinhala). The resulting data has been used to evaluate EDL system capabilities in new languages.

Finally, we created additional speech data sets with challenging but important real-world characteristics, such as the presence of multiple speakers with varying dynamic ranges, background noise, singing and background conversations, and vocalized non-speech sounds like as laughter. This included collection/selection of source data with noise conditions including multiple speakers, background noise, background conversations and similar features, followed by transmission of the data over radio channels to introduce challenging channel noise. This effort resulted in the Novel Channel corpus distributed to performers on multiple DARPA programs including DEFT.

4.8 Summary of Data Produced in DEFT

Table 1. ERE Data (Words)

	Genre	English	Chinese	Spanish
Light ERE Training Data	OSC and Narrative	219,000	--	--
	NW	--	--	47,000
	DF	361,000	200,000	26,000
Rich ERE Training Data	NW	24,000	--	45,000
	DF	340,000	245,000	85,000
Rich ERE Data used in KBP 2016 Eval	NW	42,000	29,000	26,000
	DF	46,000	52,000	41,000
Rich ERE Data used in KBP 2017 Eval	NW	33,000	34,000	30,000
	DF	43,000	50,000	43,000

Table 2. TAC KBP Data

Cold Start	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English monolingual	99,316	3112	5199	44657
Spanish/English	30K/30K	2118	1238	818
English/Chinese/ Spanish	60K/60K/60K	2469	8234	52218
Entity Linking	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English	1820	2190	-	-
Chinese	5003	5408	-	-
Spanish	4039	4713	-	-
Entity Discovery & Linking	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English	973	57483	-	-
Chinese	652	43539	-	-
Spanish	635	24395	-	-
Regular Slot Filling	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English	2,099,319	200	4599	49611
Chinese	3,015,463	135	3825	2878
Sentiment Slot Filling	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English	2,099,319	723	2557	12688
Temporal Slot Filling	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English	2,099,319	278	1535	2035
Event Argument Linking	**Source Docs**	**Queries**	**Manual Responses**	**Assessed Responses**
English monolingual	3235	80	11952	145430
English/Chinese/ Spanish	335/334/334	-	44918	-
Belief and Sentiment	**Source Docs**		**Belief Annotations**	**Sentiment Annotations**
Eval English	332		41218	47728
Eval Chinese	328		31017	42743
Eval Spanish	334		28074	36975
Training English	246		17508	38664
Training Chinese	200		11540	27982
Training Spanish	95		7400	14299
Committed Belief only English	1217		161072	--
Committed Belief only Chinese	140		19638	--
Committed Belief only Spanish	72		8136	--

Table 3. AMR Data (Number of AMRs)

Discussion Forum	Newswire	Other genres	Weblog	TOTAL
43,353	10,446	6,682	1,066	61,547

Table 4. Text Entailment & Inference Data - Number of Entailments

English	Entailment	Contradiction	Unknown	TOTAL
Newswire	3,736	1,725	5,369	10,830
Discussion Forum	4,562	2,038	3,003	9,603
TOTAL	8,298	3,763	8,372	20,433
Chinese	**Entailment**	**Contradiction**	**Unknown**	**TOTAL**
Newswire	8,324	3,565	88	11,977
Discussion Forum	13,510	5,676	18	19,204
TOTAL	21,834	9,241	106	31,181

Table 5. Text Entailment & Inference Data - Number of Inference Sets

English	T/H Pairs	Inference Steps
All genres	1,227	6,946

Table 6. Narrative Text (Proxy Reports)

Source Docs	Words
1350	657,636

4.9 Significant Accomplishments by Year

2012-2013

- Defined Light ERE (Entities, Relations & Events) annotation task with the goal of being able to rapidly produce consistently labeled data in multiple languages
- Exceeded goals for narrative text creation (proxy reports)
- Defined annotation tasks for two new extensions to TAC KBP Slot Filling: Sentiment Slot Filling (SSF) and Temporal Slot Filling (TSF). Sentiment SF involved extraction of positive and negative sentiment held by entities toward other entities, including geopolitical entities (unlike regular SF). Temporal SF utilized the regular SF slots, and involved annotation of temporal information indicating when a given SF relation held true.

- Released first increment of English text-hypothesis pairs with manual entailment judgments.
- Increased AMR corpus size to over 13K AMRs. Began increasing genre diversity by adding informal genre (DF data) to annotation pipeline (~20% DF, ~80% NW)
- Events Working Group organized the Workshop on Events: Definition, Detection, Coreference, and Representation at NAACL HLT 2013.

2014

- Expanded Light ERE annotation to Chinese and Spanish.
- Defind Rich ERE annotation which expanded entity, relation and event ontologies as well as the notion of what is taggable. Rich ERE also introduced 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.
- Shifted focus in Anomaly from speech annotation to annotation of committed belief in text in response to changing program goals; defined the task and completed annotation on training and test data to support pilot evaluation of committed belief detection on English text.
- Expanded the scope of entity extraction in TAC KBP through a new variant of the English Entity Linking task, named Entity Discovery & Linking (EDL), the goal of which was full entity extraction from a collection of documents, followed by linking entities to a knowledge base (KB) and clustering any entities not in the KB.
- Added event extraction to TAC KBP through the Event Argument Extraction task, which required annotators to extract event arguments (entities or attributes playing a role in an event) and indicate their role in an event of a given type.
- Released second increment of English text-hypothesis pairs and manual entailment judgments, plus a pilot corpus of Inference annotations on a subset of the English textual entailment data.
- Increased AMR corpus size to over 18K AMRs, with increasing genre diversity (~43% DF, 51% NW, and 6% other informal genres).
- Events Working Group analyzed the Richer Event Descriptions (RED) event annotation pilot results and provided recommendations for clarifying the guidelines and annotation of temporal relation categories.
- Events Working Group ran the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation at ACL 2014.

2015

- Rich ERE annotation expanded to include Chinese and Spanish.
- Completed basic committed belief annotation for Chinese and Spanish.
- Defined belief-and-sentiment (BeSt) task to add sentiment in addition to belief, to mark the source of belief and sentiment, and to establish annotated entities, relations, and events from ERE as the allowable targets of belief and sentiment to create greater cohesion with other DEFT annotation tasks and to guide the scope of annotation toward concepts of interest to the sponsor.
- Further expanded the extraction of events in TAC KBP with the Event Nugget task, which sought to evaluate system performance in detection and coreference of event references in text.
- Event Argument annotation was extended to include argument linking, which involved grouping event arguments involved in the same event.
- Spanish and Chinese EDL tracks were added to TAC KBP, as was a post hoc tri-lingual NIL coreference pass.
- Completed Spanish Treebank annotation of 50Kw of international Spanish newswire data and 50Kw of Latin American Spanish discussion forum data.
- Released corpus of Chinese text-hypothesis pairs and with manual entailment judgments.
- Performed extensive updates to AMR corpus to better capture concepts and predicates in informal genres.
- Increased AMR corpus size to nearly 20K AMRs, with increasing genre diversity (~42% DF, 49% NW, and 9% other genres), and improved treatment of constructions.
- Colorado produced an updated, unified set of set of Propbank frame files (in which the same concept was expressed by the same frame across different parts of speech) which was then retrofitted to existing AMR annotations.
- Released a set of AMR-sentence pairs to serve as test data in the SemEval-2016 - Task 8 (Meaning Representation Parsing task).
- Events Working Group conducted a multi-site event annotation experiment for the purpose of comparing results across annotation schemas, including participation from LDC, Colorado, RPI, CMU and BBN. The group compared the annotation of Rich ERE, Event Nugget, RED as well as Event Argument extraction as part of the event experiment to understand similarities and differences among different annotation schema and to improve Rich ERE annotation based on what was learned in the experiment.
- Colorado completed Richer Event Descriptions (RED) event annotation, and LDC distributed it to DEFT performers.
- The Events Working Group ran The 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation at NAACL HLT 2015.

2016

- Integrated Rich ERE annotation into the TAC KBP evaluation as an upstream task in the overall KBP data creation pipeline, providing input annotation to downstream annotation tasks supporting ED&L, EA, EN and BeSt.
- Completed BeSt annotation of training and test data on English, Chinese, and Spanish to support the first evaluation of this task as part of KBP.
- Spanish and Chinese annotation were added to the TAC KBP Cold Start task (in addition to English), necessitating the development of new tri-lingual query, manual annotation, and assessment pipelines.
- A new topic-driven source data scouting approach was developed in support of TAC KBP such that a single source corpus could meet the needs of every evaluation under TAC KBP 2016, unlike previous years' KBP evaluations, in which separate tracks used separate source corpora.
- Increased AMR corpus size substantially, to nearly 40K AMRs, with informal data now comprising the majority of the annotated source data (71% DF, 24% NW, and 6% other genres). Also improved AMR’s treatment of discourse phenomena found in discussion forum data.
- Released a set of AMR-sentence pairs to serve as test data in the SemEval-2017 - Task 9 (Abstract Meaning Representation Generation task).
- Colorado supported the development and production and of multi-sentence AMR annotation, by developing task guidelines, providing an annotation tool, and training other AMR sites on multi-sentence AMR annotation.
- Colorado prepared a set of 95 full corpus documents with RED annotation for public release.
- The Events Working Group ran The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation at NAACL HLT 2016. The workshop included a shared annotation task.
- A subcommittee from the Events Working Group (EWG) began work on developing a Reference Event Ontology, covering ERE event types and subtypes and also mapping ERE as well as other resources to a reference ontology, and completed basic reorganization of the ontology to a more modular framework in Protege.

2017

- Cross-document event coreference annotation was added as part of ERE annotation.
- BeSt annotation was enhanced with more detail about holders of belief and sentiment.
- Provided data for 2017 BeSt evaluation as part of TAC KBP.
- The Cold Start track was further expanded in 2017 to include slots focused on extraction of sentiment (as in the earlier Sentiment SF task) and events (a new set of slots based on the event types annotated in the Event Argument and Event Nugget tracks). As such the Cold Start track in 2017 involved extraction of all entities, relations, sentiments, and events from the TAC KBP evaluation document collection, in all three DEFT languages (English, Chinese, Spanish).
- Increased AMR corpus size to over 55K AMRs, with informal data making up the vast majority of the annotated source data.
- Produced multi-sentence AMR annotations on a subset of the existing AMR data, using the guidelines and tool provided by Colorado.
- Colorado updated and developed over 125 multi-word predicate frames for use in AMR annotation, and completed retrofitting of those frames onto existing AMR data.
- The Events Working Group, along with organizers of the previous Computing News Storylines, ran the newly combined Events and Stories in the News workshop at NAACL HLT 2017.

2018

- Completed EDL data sets for 3 incident languages including new data collection plus annotation for named entities as well as entity linking.
- Completed preparation of DEFT corpora for publication in LDC catalog

4.10 Technology Transfer

Technology transfer is achieved by distributing labeled data sets to DEFT performers to enable DEFT algorithm research and the development and testing of DEFT technology. Table 7 below summarizes the data sets released to DEFT, and the consolidated versions of those data sets to be published (or already published) in LDC's public catalog. Table 8 lists the complete set of DEFT corpora produced for the program.

Table 7. Technology Transfer Summary

Task	Data Sets Released to DEFT	Planned LDC Catalog Corpus Publications (Consolidating DEFT Program Releases)
KBP	147	14
ERE	45	5
BeSt	14	6
AMR	16	5
T/E & Inference	3	2
Other	8	2
Total	**240**	**34**

Table 8. Complete List of Corpora Produced For DEFT Program

Catalog ID	Title	Task
LDC2013R07	DEFT Phase 1 AMR Annotation Source Data R1	AMR
LDC2013R13	DEFT Phase 1 AMR Annotation Source Data R2 V1.1	AMR
LDC2014R46	DEFT Phase 2 AMR Exploratory Source Data	AMR
LDC2015R11	DEFT Phase 2 AMR Selected Segmented DF Source Data V2.0	AMR
LDC2015R33	DEFT Phase 2 AMR Selected SemEval Source Data	AMR
LDC2015R36	DEFT Phase 2 AMR Selected SemEval Test Data	AMR
LDC2016R33	DEFT Phase 3 AMR Selected SemEval 2017 Test Data	AMR
LDC2013E117	DEFT Phase 1 AMR Annotation R3	AMR
LDC2013E30	DEFT Phase 1 AMR Annotation Sample	AMR
LDC2013E47	DEFT Phase 1 AMR Annotation R1	AMR
LDC2013E66	DEFT Phase 1 AMR Annotation R2	AMR
LDC2014E41	DEFT Phase 1 AMR Annotation R4	AMR
LDC2015E86	DEFT Phase 2 AMR Annotation R1	AMR
LDC2016E25	DEFT Phase 2 AMR Annotation R2	AMR
LDC2016E33	SemEval-2016 Task 8 - Meaning Representation Parsing - Gold Standard AMRs	AMR
LDC2017E01	SemEval-2017 Task 9 Abstract Meaning Representation - Generation Subtask - AMRs Only	AMR
LDC2014R31	DEFT Sample Committed Belief Annotation	BeSt
LDC2013E93	DEFT - Source Data for Speech and Belief Anomaly Pilot Annotation Exercises	BeSt
LDC2014E106	DEFT Committed Belief Annotation R2	BeSt
LDC2014E125	DEFT Committed Belief Annotation Self-Evaluation Package	BeSt
LDC2014E55	DEFT Committed Belief Annotation R1 V1.1	BeSt
LDC2015E109	DEFT Spanish Committed Belief Sample Annotation	BeSt
LDC2015E110	DEFT English Belief and Sentiment Sample Annotation	BeSt
LDC2015E85	DEFT Chinese Committed Belief Sample Annotation	BeSt
LDC2015E99	DEFT Chinese Committed Belief Annotation	BeSt
LDC2016E27	DEFT English Belief and Sentiment Annotation V2	BeSt
LDC2016E28	DEFT Chinese Belief and Sentiment Sample Annotation	BeSt
LDC2016E40	DEFT Spanish Committed Belief Annotation	BeSt
LDC2016E61	DEFT Chinese Belief and Sentiment Annotation	BeSt
LDC2016E62	DEFT Spanish Belief and Sentiment Annotation	BeSt
LDC2013R25	NIST PREVIEW of DEFT - Source Data for Event Anomaly Pilot Annotation Exercise	ERE

LDC2014R13	DEFT Events Working Group ACE and ERE Sample Source Data	ERE
LDC2014R19	ACE and ERE Sample Source Data for DEFT Event Mention Pilot Annotation	ERE
LDC2014R50	DEFT Event Mention Detection Pilot Annotation Source Data	ERE
LDC2014R55	DEFT Event Mention Detection Pilot Tokenized Data	ERE
LDC2014R56	DEFT Event Mention Detection Training Annotation Source Data	ERE
LDC2014R61	DEFT Event Mention Annotator Training Test Annotation	ERE
LDC2014R69	DEFT Event Annotation Experiments Source Data V2	ERE
LDC2015R04	DEFT 2014 Event Nugget Evaluation Annotation Data	ERE
LDC2015R16	DEFT Rich ERE English Training Annotation Tab Delimited Format	ERE
LDC2016R14	Rich ERE English Training Data R1 with Augmented Events	ERE
LDC2013E18	DEFT Phase 1 Events Workshop Data	ERE
LDC2013E28	DEFT Phase 1 ERE Annotation Sample	ERE
LDC2013E42	DEFT Phase 1 ERE Annotation Events Workshop Documents	ERE
LDC2013E44	DEFT Phase 1 ERE Annotation R1	ERE
LDC2013E48	DEFT Phase 1 ERE Annotation R2	ERE
LDC2013E64	DEFT Phase 1 ERE Annotation R3 V2	ERE
LDC2013E88	DEFT - Source Data for Event Anomaly Pilot Annotation Exercise	ERE
LDC2014E105	DEFT 2014 Event Nugget Evaluation Pilot V2	ERE
LDC2014E113	DEFT ERE Chinese Discussion Forum Annotation V2	ERE
LDC2014E114	DEFT ERE Chinese and English Parallel Annotation V2	ERE
LDC2014E121	DEFT Event Nugget Evaluation Training Data	ERE
LDC2014E129	Richer Event Description Annotation Data	ERE
LDC2014E31	DEFT ERE English Discussion Forum Annotation V3	ERE
LDC2014E38	DEFT - Phase 1 Inference Annotation Pilot	ERE
LDC2015E03	DEFT 2014 Event Nugget Evaluation Source Data	ERE
LDC2015E105	DEFT Rich ERE Chinese Training Annotation	ERE
LDC2015E107	DEFT Rich ERE Spanish Annotation V2	ERE
LDC2015E112	DEFT Rich ERE Chinese Training Annotation R2	ERE
LDC2015E23	DEFT Rich ERE English Sample Release	ERE
LDC2015E29	DEFT Rich ERE English Training Annotation V2	ERE
LDC2015E40	DEFT Event Annotation Comparison Experiment V2	ERE
LDC2015E68	DEFT Rich ERE English Training Annotation R2 V2	ERE
LDC2015E69	DEFT 2014 Event Nugget Evaluation Annotation Data	ERE
LDC2015E71	DEFT Spanish Light ERE Training Data V2	ERE

LDC2015E78	DEFT Rich ERE Chinese and English Parallel Annotation V2	ERE
LDC2016E117	DEFT Event Sequencing After-Link Annotation Training Data	ERE
LDC2016E130	DEFT Event Sequencing After-Link and Parent-Child Annotation Training Data V4	ERE
LDC2016E31	DEFT Rich ERE English Training Annotation R3	ERE
LDC2016E32	Events Workshop Shared Annotation Data	ERE
LDC2016E34	DEFT Rich ERE Spanish Annotation R2	ERE
LDC2016E47	IC Domain Event Annotation from CMU	ERE
LDC2017E08	DEFT Event Sequencing Pilot Evaluation Source Data	ERE
LDC2017E20	DEFT ERE Cross-Doc Event Coreference Training Data Annotation Sample	ERE
LDC2017E24	DEFT ERE Cross-Doc Event Coreference Training Data Annotation V2	ERE
LDC2014E15	TAC 2014 KBP English Entity Linking Training AMR Queries and KB Links V1.1	KBP
LDC2013R14	TAC 2013 KBP English Regular Slot Filling Evaluation Queries	KBP
LDC2013R16	TAC 2013 KBP English Temporal Slot Filling Evaluation Queries	KBP
LDC2013R23	TAC 2013 KBP English Regular Slot Filling Evaluation Assessment Results PRELIMINARY V1.1	KBP
LDC2013R24	TAC 2013 KBP English Sentiment Slot Filling Evaluation Queries	KBP
LDC2013R30	TAC 2013 KBP English Entity Linking Evaluation Queries	KBP
LDC2013R31	TAC 2013 KBP Chinese Entity Linking Evaluation Queries	KBP
LDC2013R32	TAC 2013 KBP Spanish Entity Linking Evaluation Queries	KBP
LDC2013R33	TAC 2013 KBP English Cold Start Evaluation Assessment Results PRELIMINARY	KBP
LDC2013R37	TAC 2013 KBP English Temporal Slot Filling Assessment Results PRELIMINARY	KBP
LDC2013R38	TAC 2013 KBP English Sentiment Slot Filling Assessment Results	KBP
LDC2014R04	TAC 2014 KBP Events Mention Exercise Sample Documents	KBP
LDC2014R33	TAC 2014 KBP Source Documents for LDC2014E15	KBP
LDC2014R38	TAC 2014 KBP English Regular Slot Filling Evaluation Queries	KBP
LDC2014R39	TAC 2014 KBP English Sentiment Slot Filling Evaluation Queries V1.1	KBP
LDC2014R41	TAC 2014 KBP English Cold Start Evaluation Queries	KBP

LDC2014R42	TAC 2014 KBP English Cold Start Evaluation Source Corpus	KBP
LDC2014R43	TAC 2014 KBP English Event Argument Extraction Evaluation Source Corpus V1.1	KBP
LDC2014R51	TAC 2014 KBP English Entity Discovery and Linking Evaluation Queries V1.1	KBP
LDC2014R52	TAC 2014 KBP Chinese Entity Linking Evaluation Queries V2.0	KBP
LDC2014R53	TAC 2014 KBP Spanish Entity Linking Evaluation Queries	KBP
LDC2015R01	TAC KBP 2014 Chinese Regular Slot Filling Evaluation Queries	KBP
LDC2015R02	TAC KBP 2014 Spanish Regular Slot Filling Evaluation Queries	KBP
LDC2015R21	TAC KBP 2015 Tri-lingual Entity Discovery & Linking Training V2.0	KBP
LDC2015R26	TAC KBP 2015 Event Nugget and Event Coreference Linking Evaluation Gold Standard Annotation Corpus	KBP
LDC2015R29	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Evaluation Source Corpus for Producing SERIF	KBP
LDC2013E100	TAC 2013 KBP English Sentiment Slot Filling Assessment Results V1.1	KBP
LDC2013E101	TAC 2013 KBP English Cold Start Assessment Results	KBP
LDC2013E39	TAC 2012 KBP Cold Start Automated Queries Assessment Results	KBP
LDC2013E45	TAC 2013 KBP Source Corpus	KBP
LDC2013E60	TAC 2013 KBP English Regular Slot Filling Training Data	KBP
LDC2013E61	TAC 2013 KBP English Sentiment Slot Filling Sample Queries and Annotations V1.1	KBP
LDC2013E65	TAC 2011 KBP English Temporal Slot Filling Assessment Results	KBP
LDC2013E77	TAC 2013 KBP English Regular Slot Filling Evaluation Queries and Annotations V1.1	KBP
LDC2013E78	TAC 2013 KBP English Sentiment Slot Filling Training Queries and Annotations	KBP
LDC2013E82	TAC 2013 KBP English Temporal Slot Filling Training Queries and Annotations	KBP
LDC2013E86	TAC 2013 KBP English Temporal Slot Filling Evaluation Queries and Annotations V1.1	KBP
LDC2013E87	TAC 2013 KBP English Cold Start Evaluation Queries and Annotations	KBP
LDC2013E89	TAC 2013 KBP English Sentiment Slot Filling Evaluation Queries and Annotations	KBP

LDC2013E90	TAC 2013 KBP English Entity Linking Evaluation Queries and Knowledge Base Links V1.1	KBP
LDC2013E91	TAC 2013 KBP English Regular Slot Filling Assessment Results	KBP
LDC2013E96	TAC 2013 KBP Chinese Entity Linking Evaluation Queries and Knowledge Base Links V1.2	KBP
LDC2013E97	TAC 2013 KBP Spanish Entity Linking Evaluation Queries and Knowledge Base Links	KBP
LDC2013E99	TAC 2013 KBP English Temporal Slot Filling Assessment Results	KBP
LDC2014E123	TAC KBP 2014 Chinese Regular Slot Filling Training Data	KBP
LDC2014E124	TAC KBP 2014 Spanish Regular Slot Filling Training Data	KBP
LDC2014E13	TAC 2014 KBP English Source Corpus	KBP
LDC2014E20	TAC 2014 KBP Event Argument Extraction Pilot Source Corpus V1.1	KBP
LDC2014E29	TAC 2014 KBP Chinese Source Corpus	KBP
LDC2014E30	TAC 2014 KBP Spanish Source Corpus	KBP
LDC2014E40	TAC 2014 KBP Event Argument Extraction Pilot Assessment Results V1.1	KBP
LDC2014E46	TAC 2014 KBP Spanish Entity Linking Discussion Forum Training Data V1.1	KBP
LDC2014E47	TAC 2014 KBP Chinese Entity Linking Discussion Forum Training Data	KBP
LDC2014E54	TAC 2014 KBP English Entity Discovery and Linking Training Data V1.3	KBP
LDC2014E66	TAC 2014 KBP English Regular Slot Filling Evaluation Queries and Annotations V1.1	KBP
LDC2014E72	TAC 2014 KBP English Sentiment Slot Filling Evaluation Queries and Annotations V1.1	KBP
LDC2014E73	TAC 2014 KBP English Cold Start Evaluation Queries and Annotations V1.1	KBP
LDC2014E74	TAC 2014 KBP English Event Argument Extraction Evaluation Annotations V1.1	KBP
LDC2014E75	TAC 2014 KBP English Regular Slot Filling Evaluation Assessment Results V2.0	KBP
LDC2014E81	TAC 2014 KBP English Entity Discovery and Linking Evaluation Queries and Knowledge Base Links V2.0	KBP
LDC2014E82	TAC 2014 KBP English Cold Start Evaluation Assessment Results V2.0	KBP
LDC2014E83	TAC 2014 KBP Chinese Entity Linking Evaluation Queries and Knowledge Base Links V2.0	KBP
LDC2014E84	TAC 2014 KBP Spanish Entity Linking Evaluation Queries and Knowledge Base Links	KBP

LDC2014E85	TAC 2014 KBP English Sentiment Slot Filling Evaluation Assessment Results	KBP
LDC2014E87	TAC 2014 KBP English Entity Discovery and Linking Evaluation Source Corpus	KBP
LDC2014E88	TAC 2014 KBP English Event Argument Extraction Evaluation V2.0	KBP
LDC2015E01	TAC KBP 2014 Chinese Regular Slot Filling Evaluation Queries and Manual Run	KBP
LDC2015E02	TAC KBP 2014 Spanish Regular Slot Filling Evaluation Queries and Manual Run	KBP
LDC2015E100	TAC KBP 2015 English Cold Start Evaluation Assessment Results V3.1	KBP
LDC2015E101	TAC KBP 2015 English Event Argument Linking Evaluation Assessment Results V2.0	KBP
LDC2015E102	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Evaluation Queries V1.2	KBP
LDC2015E103	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Evaluation Gold Standard Entity Mentions & Knowledge Base Links	KBP
LDC2015E17	TAC KBP Chinese Entity Linking Comprehensive Training and Evaluation Data 2011-2014	KBP
LDC2015E18	TAC KBP Spanish Entity Linking Comprehensive Training and Evaluation Data 2012-2014	KBP
LDC2015E19	TAC KBP English Entity Linking	KBP
LDC2015E20	TAC KBP English Entity Discovery and Linking Comprehensive Training and Evaluation Data 2014	KBP
LDC2015E22	TAC KBP English Event Argument Extraction Comprehensive Pilot and Evaluation Data 2014	KBP
LDC2015E41	TAC KBP 2015 English Event Argument Linking Training Data	KBP
LDC2015E42	TAC KBP Knowledge Base II - BaseKB	KBP
LDC2015E43	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Knowledge Base Entries Creation Algorithm	KBP
LDC2015E44	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Pilot Gold Standard Knowledge Base Links V1.1	KBP
LDC2015E45	TAC KBP Comprehensive English Source Corpora 2009-2014	KBP
LDC2015E46	TAC KBP English Regular Slot Filling - Comprehensive Training and Evaluation Data 2009-2014	KBP
LDC2015E47	TAC KBP English Sentiment Slot Filling - Comprehensive Training and Evaluation Data 2013-2014	KBP
LDC2015E48	TAC KBP English Cold Start - Comprehensive Evaluation Data 2012-2014	KBP

LDC2015E49	TAC KBP English Surprise Slot Filling - Comprehensive Training and Evaluation Data 2010	KBP
LDC2015E50	TAC KBP English Temporal Slot Filling - Comprehensive Training and Evaluation Data 2011 and 2013	KBP
LDC2015E61	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Pilot Source Corpus	KBP
LDC2015E62	TAC KBP 2014 Chinese Regular Slot Filling Evaluation Queries	KBP
LDC2015E67	TAC KBP 2014 Chinese Regular Slot Filling Evaluation Assessment Results V2.0	KBP
LDC2015E72	TAC KBP 2015 Cold Start - Entity Discovery Sample Data	KBP
LDC2015E73	TAC KBP 2015 Event Nugget Training Data Annotation V2	KBP
LDC2015E75	TAC KBP 2015 Tri-lingual Entity Discovery & Linking Training Data V2.1	KBP
LDC2015E76	TAC KBP 2015 English Cold Start Evaluation Queries V2.0	KBP
LDC2015E77	TAC KBP 2015 English Cold Start Evaluation Source Corpus V2.0	KBP
LDC2015E79	TAC KBP 2015 English Event Argument Linking Evaluation Source Corpus	KBP
LDC2015E80	TAC KBP 2015 English Cold Start Evaluation Queries and Manual Run	KBP
LDC2015E81	TAC KBP 2015 English Cold Start Entity Discovery Evaluation Gold Standard Entity Mentions V1.2	KBP
LDC2015E92	TAC KBP 2015 English Event Argument Linking Evaluation Manual Run	KBP
LDC2015E93	TAC KBP 2015 Tri-lingual Entity Discovery and Linking Evaluation Source Corpus V2.0	KBP
LDC2015E94	TAC KBP 2015 Event Nugget and Event Coreference Linking Evaluation Source Corpus	KBP
LDC2015E95	TAC KBP 2015 Event Nugget Annotation for Event Coreference Linking Evaluation	KBP
LDC2016E106	TAC KBP 2016 Cold Start Evaluation Assessment Results V4.0	KBP
LDC2016E107	TAC KBP 2016 English Event Argument Linking Evaluation Assessment Results V2.0	KBP
LDC2016E114	TAC KBP 2016 Belief and Sentiment Evaluation Gold Standard Annotation V2	KBP
LDC2016E35	TAC KBP Chinese Regular Slot Filling Comprehensive Training and Evaluation Data 2014	KBP
LDC2016E36	TAC KBP English Event Nugget Detection and Coreference Comprehensive Training and Evaluation Data 2014-2015	KBP

LDC2016E37	TAC KBP English Event Argument Comprehensive Training and Evaluation Data 2014-2015	KBP
LDC2016E38	TAC KBP Entity Discovery and Linking Comprehensive Training and Evaluation Data 2014-2015	KBP
LDC2016E39	TAC KBP English Cold Start Comprehensive Evaluation Data 2012-2015	KBP
LDC2016E41	TAC KBP 2016 Bilingual Spanish-English Cold Start Pilot - Training Data V1.1	KBP
LDC2016E42	TAC KBP 2016 Bilingual Spanish-English Cold Start Pilot - Source Corpus	KBP
LDC2016E43	TAC KBP 2016 Bilingual Spanish-English Cold Start Pilot - Queries	KBP
LDC2016E44	TAC KBP 2016 Bilingual Spanish-English Cold Start Pilot - Queries and Manual Run	KBP
LDC2016E49	TAC KBP 2016 English Event Argument Linking Pilot Source Corpus	KBP
LDC2016E51	TAC KBP 2016 English Event Argument Linking Pilot Queries and Manual Run V1.1	KBP
LDC2016E52	TAC KBP 2016 Bilingual Spanish-English Cold Start Pilot Assessment Results V1.1	KBP
LDC2016E59	TAC KBP 2016 English Event Argument Linking Pilot Assessment Results V1.1	KBP
LDC2016E60	TAC KBP 2016 English Event Argument Linking Pilot Gold Standard	KBP
LDC2016E63	TAC KBP 2016 Evaluation Source Corpus V1.1	KBP
LDC2016E64	TAC KBP 2016 Evaluation Core Source Corpus	KBP
LDC2016E65	TAC KBP 2016 Cold Start Evaluation Queries	KBP
LDC2016E67	TAC KBP English Event Nugget Training Data - Character Based Format Conversion	KBP
LDC2016E68	TAC KBP 2016 Entity Discovery & Linking Evaluation Gold Standard Entity Mentions and Knowledge Base Links	KBP
LDC2016E69	TAC KBP 2016 Cold Start Evaluation Queries and Manual Run V1.1	KBP
LDC2016E71	TAC KBP 2016 Eval Core Set Rich ERE Annotation	KBP
LDC2016E72	TAC KBP 2016 Eval Core Set Event Nugget Annotation	KBP
LDC2016E73	TAC KBP 2016 Eval Core Set Rich ERE Annotation with Augmented Event Argument V2	KBP
LDC2016E74	TAC KBP 2016 English Event Argument Linking Evaluation Queries and Manual Run	KBP
LDC2017E02	TAC KBP English Event Nugget Detection and Coreference - Comprehensive Training and Evaluation Data 2014-2016	KBP

LDC2017E03	TAC KBP Entity Discovery and Linking Comprehensive Training and Evaluation Data 2014-2016 V1.1	KBP
LDC2017E04	TAC KBP English Cold Start Comprehensive Evaluation Data 2012-2016	KBP
LDC2017E05	TAC KBP English Event Argument Comprehensive Training and Evaluation Data 2014-2016	KBP
LDC2017E25	TAC KBP 2017 Evaluation Source Corpus V1.1	KBP
LDC2017E26	TAC KBP 2017 Cold Start Evaluation Queries V1.1	KBP
LDC2017E34	TAC KBP 2017 Cold Start Evaluation Queries and Manual Run V1.2	KBP
LDC2017E51	TAC KBP 2017 Evaluation Core Source Corpus	KBP
LDC2017E52	TAC KBP 2017 Entity Discovery & Linking Evaluation Gold Standard Entity Mentions and Knowledge Base Links	KBP
LDC2017E53	TAC KBP 2017 Eval Core Set Rich ERE Annotation	KBP
LDC2017E54	TAC KBP 2017 Eval Core Set Event Nugget Annotation	KBP
LDC2017E55	TAC KBP 2017 Eval Core Set Rich ERE Annotation with Augmented Event Arguments	KBP
LDC2017E56	TAC KBP 2017 Cold Start Evaluation Assessment Results V2.0	KBP
LDC2017E79	TAC KBP 2017 Event Sequencing Eval Source Data	KBP
LDC2017E80	TAC KBP 2017 Belief and Sentiment Evaluation Gold Standard Annotation	KBP
LDC2017E83	TAC KBP 2017 Event Sequencing Eval After Link Parent Child Annotation	KBP
LDC2017E81	REFLEX NMSU Chechen Language Pack for KBP 10-language EDL Pilot	KBP
LDC2017E82	REFLEX NMSU Chechen Language Pack for KBP 10-language EDL Pilot - Original Release	KBP
LDC2013E11	DEFT Phase 1 Sample Narrative Text Creation	Narrative text
LDC2013E19	DEFT Phase 1 Narrative Text Source Data R1	Narrative text
LDC2013E29	DEFT Phase 1 Narrative Text Source Data R2	Narrative text
LDC2013E95	DEFT Narrative Text Source Data R3	Narrative text
LDC2014R40	DEFT - Chinese Lexical Resources V1	Other
LDC2014E122	Chinese Lexical Resources V2	Other
LDC2013G01	Albany MPC Dialogue Chat Corpus	Site donation
LDC2013E06	Deep NLU Exploration - DEFT Pilot Source Audio	Source data
LDC2013E07	Deep NLU Exploration - DEFT Pilot Source Text and Annotations	Source data

LDC2014E14	Spanish Discussion Forum Source Data R1	Source data
LDC2014E33	DEFT English Discussion Forum Source For Annotation R1	Source data
LDC2014R62	DEFT NW Data for Spanish Treebank Annotation V2.0	TB
LDC2014R71	DEFT DF Data for Spanish Treebank Annotation	TB
LDC2014E130	DEFT International Spanish NW Treebank V2.0	TB
LDC2015E66	DEFT Spanish Treebank V2	TB
LDC2013E79	DEFT - Phase 1 Textual Entailment Annotation R1	TE
LDC2014E07	DEFT - Phase 1 Textual Entailment Annotation R2	TE
LDC2015E106	DEFT - Chinese Textual Entailment Annotation	TE

5.0 CONCLUSIONS

The ReORIENT (Resources for Operationally Relevant Information Extraction from Non-explicit Text) research effort has developed a set of linguistic resources to support deep natural language understanding in the context of a diffuse, diverse and large community of researchers and stakeholders. To support Relational Analysis research we have developed data sets labeled for entities, relations, events, and AMR sembanking. To support Anomaly Analysis research we have developed resources labeled for sentiment and belief-based and event-based phenomena. To support Smart Filtering research we have created data sets labeled for textual entailment and inference.

The Event, Relation, Entity (ERE) Annotation Data sets represent the primary core resource developed for DEFT, producing over 2Mw of English, Spanish and Chinese data richly labeled for complex events including coreference. The Richer Event Descriptions (RED) and Rich Event Ontology annotation research efforts extended event annotation to include event relations as well as broader ontology coverage. The AMR Sembank developed under DEFT represents a signficant new semantic resource. Thousands of sentences labeled for text entailment and inference supported DEFT smart filtering research. The BeSt task enabled DEFT performers working on belief and sentiment to connect their task to the work being done by those working on entities, relations, and events. Dozens of DEFT-relevant data sets were produced to support training and evaluation of system capablities under the annual TAC KBP evaluation series. We produced a new DEFT Spanish Treebank, Chinese Lexical Resources and additional DEFT data in low resource languages.

Technology transfer is achieved by distributing labeled data sets to DEFT performers to enable DEFT algorithm research and the development and testing of DEFT technology. A total of 240 distinct data sets were developed under this effort and distributed to DEFT performers during the program. These resources have been consolidated into 34 corpora that have been or will soon be published in LDC's public catalog, making DEFT data available to the wider research community, thus amplifying the government's investment in linguistic data and stimulating relevant research outside of the program.

APPENDIX A – PUBLICATIONS

Teruko Mitamura, Yukari Yamakawa, Susan Holm, Zhiyi Song, Ann Bies, Seth Kulick, Stephanie Strassel. *Event Nugget Annotation: Processes and Issues*. NAACL HLT 2015: Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies. 3rd Workshop on EVENTS: Definition, Detection, Coreference, and Representation.

Yu Hong, Tongtao Zhang,Tim O'Gorman, Sharone Horowit-Hendler, Heng Ji, Martha Palmer. *Building a Cross-document Event-Event Relation Corpus*. 10th Linguistic Annotation Workshop (LAW-X 2016), held in conjunction with ACL 2016.

C. Bonial, S. W. Brown, M. Palmer. *A lexically informed upper-level event ontology*. ISA-12, In the Interoperability of Semantic Annotation (ISA-12), with LREC 2016.

R. Ikuta, W. Styler IV, M. Hamang, T. O'Gorman, M. Palmer. *Challenges of adding causation to Richer Event Descriptions*. Proceedings of the 2nd Events Workshop, held in conjunction with ACL 2014.

C. Bonial, D. Tahmoush, S. W. Brown, M. Palmer. *Multimodal use of an upper-level event ontology*. Proceedings of the 4th Workshop on Events: Definition, Detection, Coreference, and Representation, held in conjunction with NAACL 2016.

T. O'Gorman, K Wright-Bettner, M. Palmer. *Richer Event Description: Integrating event coreference with temporal, causal and bridging annotation*. Computing News Storylines Workshop at EMNLP 2016, pp. 47-56.

S. W. Brown, C. Bonial, L. Obrst, M. Palmer. *The Rich Event Ontology*. Proceedings of the Events and Stories in the News Workshop, ACL 2017, pp. 87-97.

M. Pust, U. Hermjakob, K. Knight, D. Marcu, and J. May. *Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation.* EMNLP 2015: Conference on Empirical Methods on Natural Language Processing.

Jonathan May. *SemEval-2016 Task 8: Meaning Representation Parsing*. SemEval 2016.

Nima Pourdamghani, Yang Gao, Ulf Hermjakob, Kevin Knight. *Aligning English Strings with Abstract Meaning Representation Graphs.* EMNLP, 2014: Conference on Empirical Methods on Natural Language Processing.

Michael Pust, Ulf Hermjakob, Kevin Knight, Daniel Marcu, Jonathan May. *Parsing English into Abstract Meaning Representation Using Syntax-Based Machine Translation*. EMNLP 2015: Conference on Empirical Methods on Natural Language Processing

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, Joe Ellis. *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. 2nd Workshop on Events: Definition, Detection, Coreference, and Representation.

Ann Bies. *Balancing the Existing and the New in the Context of Annotating Non-Canonical Language*. NAACL HLT: Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies.

Zhiyi Song, Ann Bies, Stephanie Strassel, Tom Riese, Justin Mott, Joe Ellis, Jonathan Wright, Seth Kulick, Neville Ryant and Xiaoyi Ma. *From Light to Rich ERE: Annotation of Entities, Relations, and Events*. NAACL HLT 2015: 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies.

Seth Kulick, Ann Bies, Justin Mott. *Inter-annotator Agreement for ERE Annotation*. ACL 2014: 52nd Annual Meeting of the Association for Computational Linguistics.

Justin Mott, Ann Bies, Zhiyi Song, Stephanie Strassel. *Parallel Chinese-English Entities, Relations and Events Corpora*. LREC 2016: 10th Edition of the Language Resources and Evaluation Conference.

Mariona Taulé, M Antonia Martí, Ann Bies, Aina Garí, Montserrat Nofre, Zhiyi Song, Stephanie Strassel and Joe Ellis. *Spanish Treebank Annotation of Informal Non-Standard Web Text*. ICWE 2015.

Vinodkumar Prabhakaran, Tomas By, Julia Hirschberg, Owen Rambow, Samira Shaikh, Tomek Strzalkowski, Jennifer Tracey, Michael Arrigo, Rupayan Basu, Micah Clark, Adam Dalton, Mona Diab, Louise Guthrie, Anna Prokofieva, Stephanie Strassel, Gregory Werner, Janyce Wiebe and Yorick Wilks. *A New Dataset and Evaluation for Belief/Factuality*. *SEM 2015: Fourth Joint Conference on Lexical and Computational Semantics.

Zhiyi Song, Ann Bies, Stephanie Strassel, Joe Ellis, Teruko Mitamura, Hoa Dang, Yukari Yamakawa, Sue Holm. *Event Nugget and Event Coreference Annotation*. NAACL HLT 2016: 15th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies
The 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation San Diego, June 17.

Ann Bies, Zhiyi Song, Jeremy Getman, Joe Ellis, Justin Mott, Stephanie Strassel, Martha Palmer, Teruko Mitamura, Marjorie Freedman, Heng Ji, Tim O'Gorman. *A Comparison of Event Representations in DEFT*. NAACL HLT 2016: 15th Annual Conference of the North American Chapter of the Association for Computational

Linguistics: Human Language Technologies.

Xiaoman Pan, Taylor Cassidy, Ulf Hermjakob, Heng Ji, Kevin Knight. *Unsupervised Entity Linking with Abstract Meaning Representation*. NAACL HLT 2015: 14th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Jeremy Getman, Joe Ellis, Stephanie Strassel, Zhiyi Song, Jennifer Tracey. *Laying the Groundwork for Knowledge Base Population: Nine Years of Linguistic Resources for TAC KBP*. LREC 2018: 11th Edition of the Language Resources and Evaluation Conference.

Zhiyi Song, Ann Bies, Justin Mott, Xuansong Li, Stephanie Strassel, Christopher Caruso. *Cross-Document, Cross-Language Event Coreference Annotation Using Event Hoppers*. LREC 2018: 11th Edition of the Language Resources and Evaluation Conference.

GLOSSARY: ABBREVIATIONS, ACRONYMS

AMR	Abstract Meaning Representation
BBN	Raytheon BBN
BeSt	Belief and Sentiment
CMU	Carnegie Mellon University
DF	Discussion Forum
EA	Event Argument
EDL	Entity Discovery and Linking
EN	Event Nugget
ERE	Entities, Relations and Events
EWG	Event Working Group
HLT	Human Language Technology
IL(#)	Incident Language; IL# refers to a particular one of a set of incident languages
Kw	Thousand (of) words
Mw	Million (of) words
NAACL	North American Chapter of the Association for Computational Linguistics
NLU	Natural Language Understanding
NW	Newswire
OSC	Open Source Center
RED	Richer Event Description
RPI	Rensselaer Polytechnic Institute
SF	Slot Filling
SSF	Sentiment Slot Filling
TAC KBP	Text Analysis Conference - Knowledge Base Population
TSF	Temporal Slot Filling