



Overview of Linguistic Resources for the TAC KBP 2017 Evaluations

Jeremy Getman, Jennifer Tracey, Zhiyi Song,
Joe Ellis, Stephanie Strassel

*Linguistic Data Consortium
University of Pennsylvania*

- ◆ Linguistic resources for TAC KBP 2017
 - Ninth year LDC produced KBP resources
 - Twenty-one new data sets
- ◆ Five evaluation tracks
 - Entity Discovery & Linking
 - Event Arguments
 - Event Nuggets
 - Belief and Sentiment
 - Cold Start
- ◆ Resource creation methods similar to 2016



- ◆ Single source document collection for all evaluations
 - Approximately 90,000 documents for full evaluation corpus
 - 500-doc, manually-selected subset for gold standard data
 - 800-token max per doc (not including quote regions for DF)
 - Large, diverse set of features required, including:

Multiple mentions of all event types and subtypes in all languages and genres	Entities with nominal-only mentions in some documents, named resolutions in others
Cross-lingual, cross-document event mention clusters for at least half of the event subtypes	Other varieties of “ambiguous” entities
Multiple, cross-document instances of relatively simple, non-confusable events	Relatively short time span

Source Collections and Topic Selection

- ◆ Unexposed source collections
 - Newswire (NW)
 - 2013 English New York Times articles
 - 2013 Chinese, English and Spanish Xinhua articles
 - Discussion Forum threads (DF)
 - Harvested in 2016
- ◆ Topic selection
 - Entities, event types and ref doc for each topic.

2013 Savar building collapse



Aerial view of the building following the disaster

Time	08:45 am BST (UTC+06:00) ^[1]
Date	24 April 2013
Location	Savar Upazila, Dhaka District, Bangladesh
Coordinates	 23°50'46"N 90°15'27"E
Also known as	Rana plaza building collapse
Deaths	1,129 ^[2]
Non-fatal injuries	~2,500 ^[3]

- ◆ Manually-selected ‘core’ corpus
 - Data scouts search for and review on-topic documents, tally event mentions, and list relevant and/or confusable entities
 - Over 1,000 documents reviewed; 500 selected
- ◆ Automated selection for full corpus
 - Remaining 85.5K documents selected using fuzzy string matches with annotations from the core corpus to increase overlap with topic-relevant entities and events

Language	Genre	Docs	Words
Chinese	NW	83	33,683
Chinese	DF	84	49,932
English	NW	83	32,572
English	DF	84	42,891
Spanish	NW	83	29,615
Spanish	DF	83	42,850
Total		500	231,543

- ◆ Entities, Relations, and Events (Rich ERE)
 - Developed under DARPA Deep Exploration and Filtering of Text program (DEFT) Program
 - Ontology largely aligned with KBP
 - Exhaustive labeling of entities, relations and events and their attributes plus within-doc coreference

- ◆ Manual ERE annotation serves as input to downstream ED&L, EA, EN, and BeSt annotation/manual runs

ERE Annotation

The **Bo Xilai** Event was ignited by **Lijun** running into the **US consulate** in 2012 to bring **Bogu Kailai's** killing to light. Will **Bo Xilai** end up in **jail** due to bribery and corruption; what will **his wife** end up with?

Entity				
	NAM	Bo Xilai, Bo Xilai		
PER	PRO	his	SPC	IND
PER	NAM	Lijun	SPC	IND
	NAM	Bogu Kailai		
PER	NOM	his wife	SPC	IND
GPE	NAM	US	SPC	IND
LOC	NOM	US consulate	SPC	IND
			Non	
LOC	NOM	jail	SPC	
Relation				
R1	Physical.Located	<i>Lijun</i>		
		<i>into</i>		<i>US consulate</i>
R2	Personal.Family	<i>his</i>		
		<i>wife</i>		<i>his wife</i>

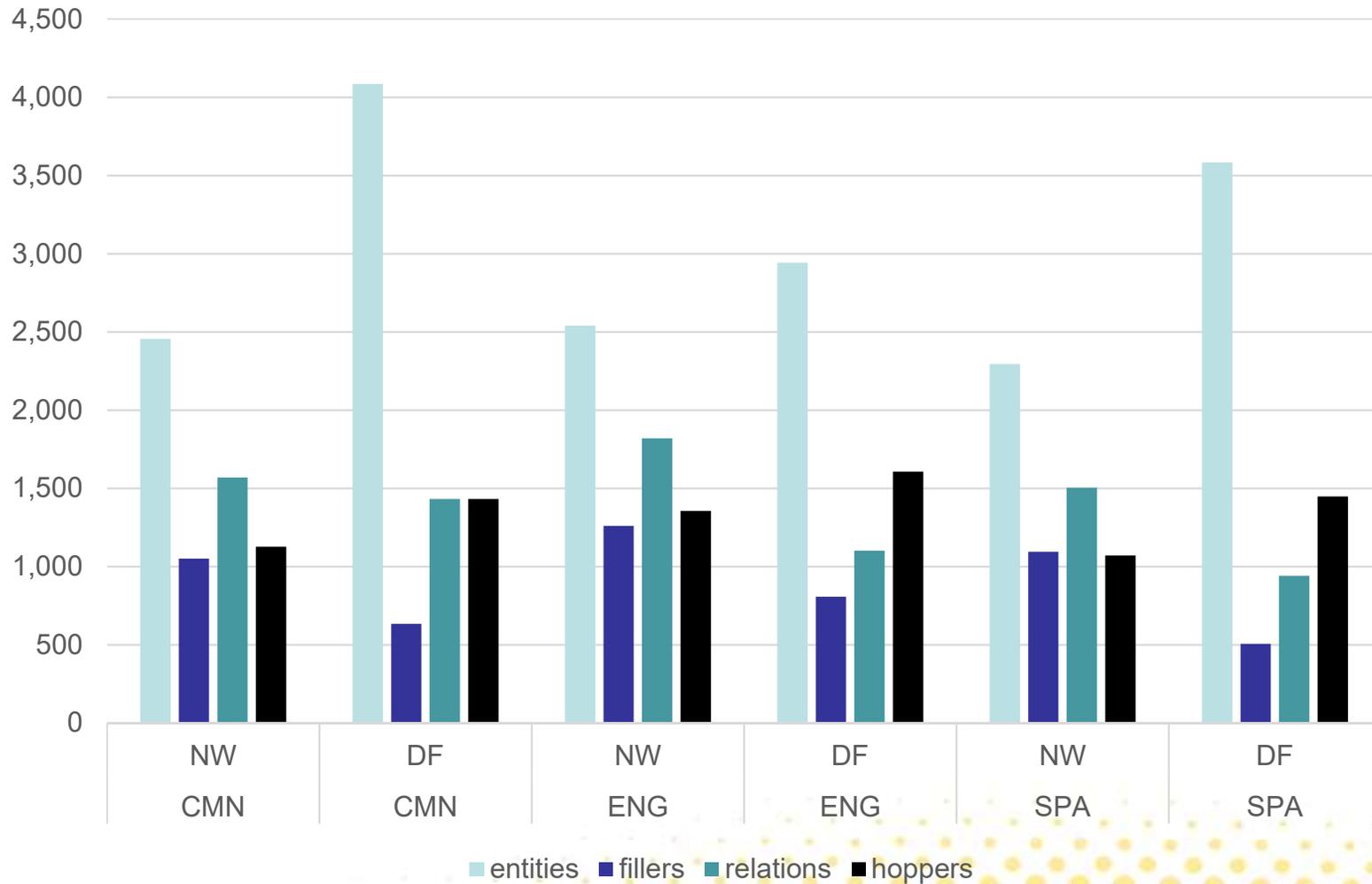
The **Bo Xilai** Event was ignited by **Lijun running** into the **US consulate** in 2012 to bring **Bogu Kailai's killing** to light. Will **Bo Xilai** end up in **jail** due to bribery and corruption; what will **his wife** end up with?

Event				
	Type.Subtype	Realis	Trigger	Arguments
H1	Movement. Transport Person	Actual	running	Lijun
				US consulate
				2012
H2	Conflict. Attack	Actual	killing	Bogu Kailai
H3	Justice. Jail	Other	jail	Bo Xilai
				bribery and corruption
H4	Life.Die	Actual	killing	Bogu Kailai
H5	Movement. Transport Person	Other	end up	Bo Xilai
				jail

- ◆ ERE event type inventory for 2016 & 2017
 - 8 types and 18 subtypes
 - Types evaluated in Event Argument and Event Nugget

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

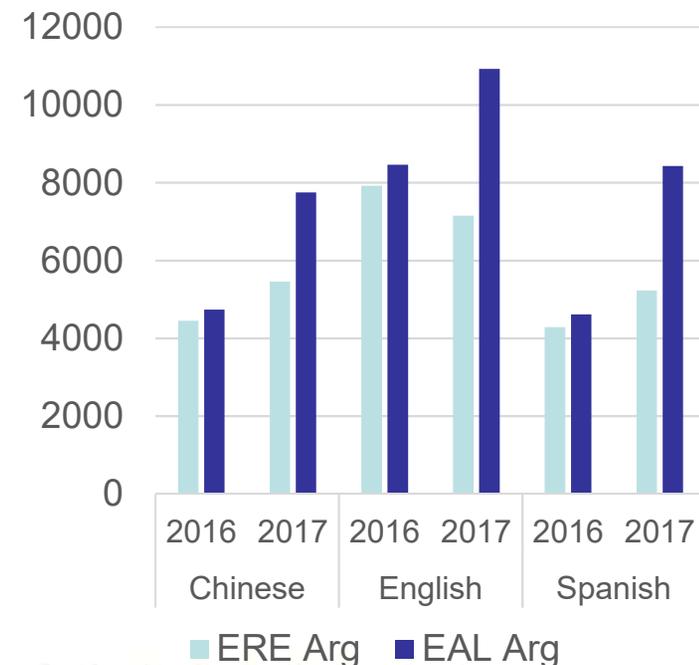
ERE annotation counts, 500 core documents



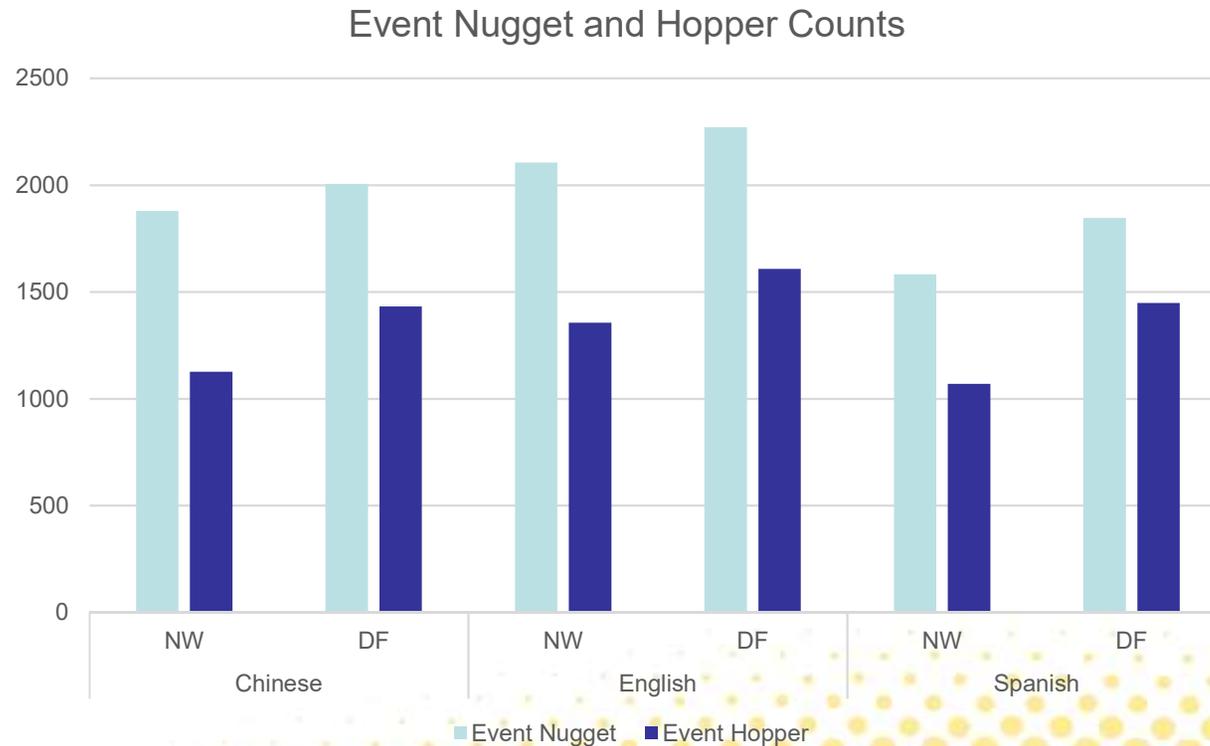
- ◆ Overall goals consistent with 2016
 - Exhaustive cross-document, cross-lingual, entity extraction, clustering, and linking
 - Import specific, individual entity mentions from Rich ERE data
- ◆ Also provided low resource language packs to performers participating in 10-language EDL

Status	Chinese	English	Spanish
Linked to KB	7,673	4,572	4,982
NIL	2,573	2,343	2,230
Total	10,246	6,915	7,212

- ◆ Gold Standard
 - ERE annotations on core source corpus
- ◆ Augmentation pass
 - Inferred arguments
 - Locational containment
 - 2016
 - BBN script run over ERE data
 - Annotators review results
 - 2017
 - Manual augmentation pass



- ◆ No separate EN annotation task
 - Data are entirely produced by running a script over ERE data to extract and reformat a subset for use by EN



- ◆ Only ERE entities are holders of belief and sentiment
 - NEW in 2017: “author” and “other” flags when belief/sentiment holder is not an ERE annotated entity
- ◆ Only ERE entities, relations and events are targets of belief and sentiment
- ◆ Belief values: committed, non-committed, reported, n/a
- ◆ Sentiment values (polarity): positive, negative
- ◆ Annotated but not evaluated:
 - Sarcasm flag indicating when polarity annotated is opposite of literal meaning (based on context)
 - Belief about role of arguments in events
 - Belief polarity

Belief Annotation Example

Ominous new action by **UKRAINE'S SECURITY FORCES** on Monday, including a **raid** on **AN OPPOSITION PARTY'S HEADQUARTERS**, appears to diminish prospects for **talks** between **PROTEST LEADERS**, a way to defuse

Event: raid

Event: talks

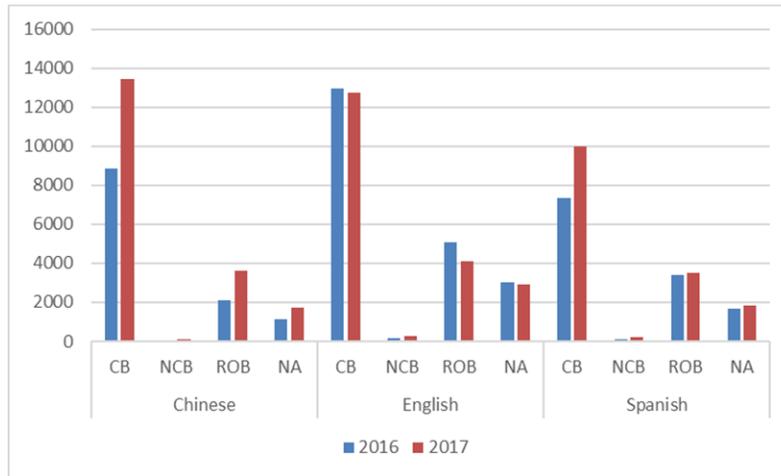
NA

Arguments also

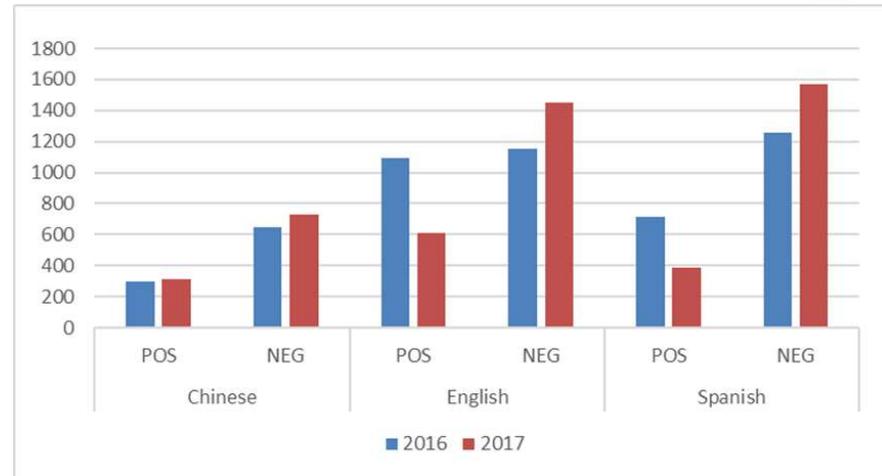
NA

Language	Belief annotations		Sentiment annotations	
	2016	2017	2016	2017
Chinese	12,163	18,854	18,982	23,761
English	21,188	20,030	25,358	22,370
Spanish	12,546	15,528	17,353	19,622

- ◆ In 2016, English had a significantly higher total number of annotations
- ◆ In 2017, the difference across languages was smaller



Belief annotations in 2016 and 2017



Sentiment annotations in 2016 and 2017

- ◆ Distributions similar between 2016 and 2017 data
 - Very large numbers of Committed Beliefs
 - Very small numbers of Non-Committed Beliefs
 - Presence of sentiment in the data remains low

Lance Barrett, 23, of London, KY, was charged with first-degree attempted burglary, theft of a firearm, and carrying a concealed weapon.

Lesa Bailey, 44, of London, KY, was charged with criminal conspiracy to make meth, unlawful possession of meth precursors and possession of a controlled substance.

London – *gpe:residents_of_city* – *per:charges*

• **Lance Barrett**

- first-degree attempted burglary
- theft of a firearm
- carrying a concealed weapon

• **Lesa Bailey**

- criminal conspiracy to make meth
- unlawful possession of meth precursors
- possession of a controlled substance

- ◆ Chains of entities connected by KBP slots
 - Cold Start queries comprised of
 - Entity – Slot 0 – Slot 1
- ◆ Cold Start annotation & query development (QD) separate
 - Though some produced during QD to count
 - Productive queries
 - Multi-lingual queries
 - Annotation much closer to exhaustive

◆ Overall goal consistent with 2016

- Create queries to sample KBs, a “manual run”, and assess responses

◆ Changes in 2017

- New slot categories: events & sentiment
- Pipeline for query & manual run development extended from 6 stages to 12
 - 2016: 3 for QD (1/language) + 3 for MR
 - 2017: 9 for QD (1/category/language) + 3 for MR

Slot0	Slot1	Targeted # of Queries
Event	-	300
SF (string)	-	100
SF (entity)	SF (any)	300
SF (entity)	Sentiment	100
SF (entity)	Event	100
Sentiment	SF (any)	100
Sentiment	Sentiment	100
Sentiment	Event	100

- ◆ Increase in precision and recall in 2017
 - Second year of cross-lingual Cold Start
 - Learned from challenges faced in 2016
- ◆ Cross-lingual coref
 - Still challenging
 - KB and SF scores slightly different

Year	Lang.	Cond.	Prec.	Recall	F1
2017	ENG	KB	94.7%	41.3%	57.5%
2017	CMN	KB	94.5%	41.4%	57.6%
2017	SPA	KB	88.0%	78.3%	82.9%
2017	X-ling	KB	89.7%	38.2%	53.6%
2017	ENG	SF	94.5%	41.4%	57.6%
2017	CMN	SF	94.5%	41.4%	57.6%
2017	SPA	SF	88.2%	78.3%	83.0%
2017	X-ling	SF	89.6%	38.4%	53.8%
2016	ENG	SF	80%	34%	48%
2016	CMN	SF	76%	25%	38%
2016	SPA	SF	87%	64%	74%
2016	X-ling	SF	78%	35%	49%

- ◆ Supported 5 evaluation tracks
- ◆ Created 10 new resources (21 releases)

- ◆ Ongoing work to publish KBP resources in LDC catalog
 - Consolidated packages from prior years
 - Updates to add 2016 and 2017 corpora

Newly Created Datasets for TAC KBP 2017

Track	Catalog ID	Title
All	LDC2017E25	TAC KBP 2017 Evaluation Source Corpus V1.1
Cold Start	LDC2017E26	TAC KBP 2017 Cold Start Evaluation Queries V1.1
Cold Start	LDC2017E34	TAC KBP 2017 Cold Start Evaluation Queries and Manual Run V1.2
All	LDC2017E51	TAC KBP 2017 Evaluation Core Source Corpus
ED&L	LDC2017E52	TAC KBP 2017 Entity Discovery and Linking Evaluation Gold Standard Entity Mentions and Knowledge Base Links
BeSt	LDC2017E53	TAC KBP 2017 Eval Core Set Rich ERE Annotation
Event Nugget	LDC2017E54	TAC KBP 2017 Eval Core Set Event Nugget Annotation
Event Argument	LDC2017E55	TAC KBP 2017 Eval Core Set Rich ERE Annotation with Augmented Event Arguments
Cold Start	LDC2017E56	TAC KBP 2017 Cold Start Evaluation Assessment Results V2.0
BeSt	LDC2017E80	TAC KBP 2017 Belief and Sentiment Evaluation Gold Standard Annotation

Consolidated Datasets of Past KBP Resources

Catalog ID	Title
LDC2015E17	TAC KBP Chinese Entity Linking Comprehensive Training and Evaluation Data 2011 - 2014
LDC2015E19	TAC KBP English Entity Linking Comprehensive Training and Evaluation Data 2009 - 2013
LDC2015E45	TAC KBP Comprehensive English Source Corpora 2009-2014
LDC2015E46	TAC KBP English Regular Slot Filling Comprehensive Training and Evaluation Data 2009-2014
LDC2015E47	TAC KBP English Sentiment Slot Filling Comprehensive Training and Evaluation Data 2013-2014
LDC2015E49	TAC KBP English Surprise Slot Filling Comprehensive Training and Evaluation Data 2010
LDC2015E50	TAC KBP English Temporal Slot Filling Collected Training and Evaluation Data Sets 2011 and 2013
LDC2016T26	TAC KBP Spanish Entity Linking Comprehensive Training and Evaluation Data 2012 - 2014
LDC2016E35	TAC KBP Chinese Regular Slot Filling Comprehensive Training and Evaluation Data 2014
LDC2017E02	TAC KBP Event Nugget Detection and Coreference Comprehensive Training and Evaluation Data 2014-2016
LDC2017E03	TAC KBP Entity Discovery and Linking Comprehensive Training and Evaluation Data 2014-2016 V1.1
LDC2017E04	TAC KBP Cold Start Comprehensive Evaluation Data 2012-2016
LDC2017E05	TAC KBP Event Argument Comprehensive Training and Evaluation Data 2014 - 2016

