TAC KBP 2016 Linguistic Resources: Data Selection, Entity Discovery & Linking (ED&L), and Cold Start

Joe Ellis (presenter), Neil Kuster, Jeremy Getman, Zhiyi Song, Ann Bies, Stephanie Strassel

Linguistic Data Consortium
University of Pennsylvania

Introduction and Overview

- Linguistic resources for TAC KBP 2016
  - Eighth year LDC produced KBP resources
  - Twenty-nine new data sets
- Two primary goals
  - Increase coordination across tracks
  - Increase multi-lingual evaluation tracks

- Today
  - Doc selection, ED&L, and Cold Start
- Tomorrow
  - Event Arguments, Event Nuggets, and Belief/Sentiment (BeSt)

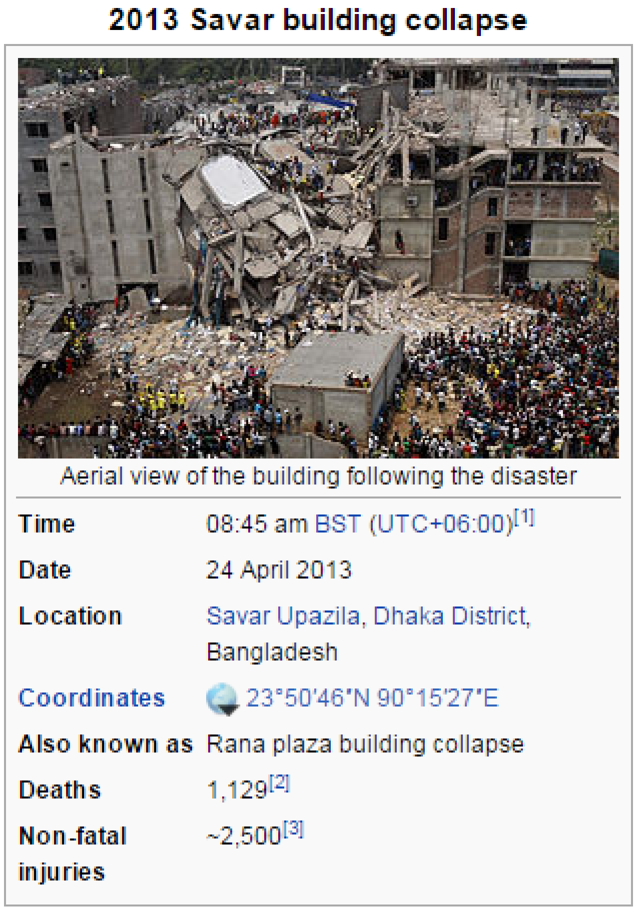
Source Document Selection

- 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)
    - Scouted online
- Topic selection
  - Entities, event types and ref doc for each topic.



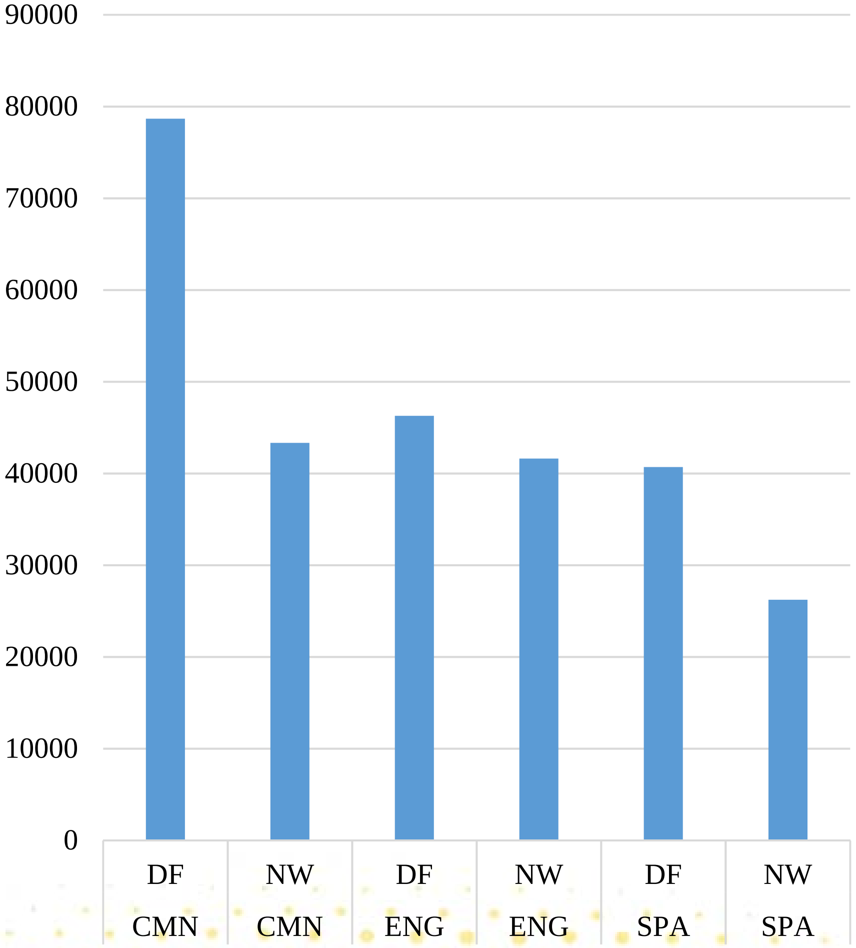
2013 Savar building collapse

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]

Document Selection

- Manually-selected ‘core’ corpus
  - Data scouts review documents related to topics and tally features
  - Over 1,100 documents reviewed; 500 selected
- Automated selection for full corpus
  - Remaining 85.5K documents selected using fuzzy string matches with annotations from the core corpus



Linguistic Data Consortium

Entity Discovery & Linking

- Overall goal consistent with 2015
  - Exhaustive cross-document, cross-lingual, entity extraction, clustering, and linking.
- Changes to approach in 2016
  - Import entity mentions from Rich ERE (Entities, Relations, Events) data
  - New knowledge base (KB) search
  - Other annotation changes
    - Add'l language and entity types for NOM mentions, no TTLs, no intra-token mentions

Rich ERE

- Entities, Relations, and Events (ERE),
  - Ongoing annotation task developed by LDC for DARPA's Deep Exploration and Filtering of Text program (DEFT)
  - Exhaustive labeling of entities, relations and events and their attributes.
  - Used as input for many downstream annotation tasks supporting multiple tracks in KBP 2016
  - To address ED&L requirements for KB linking on individual entities only, added Individual, Group, or Indeterminate label for Specific entities

ERE Annotation

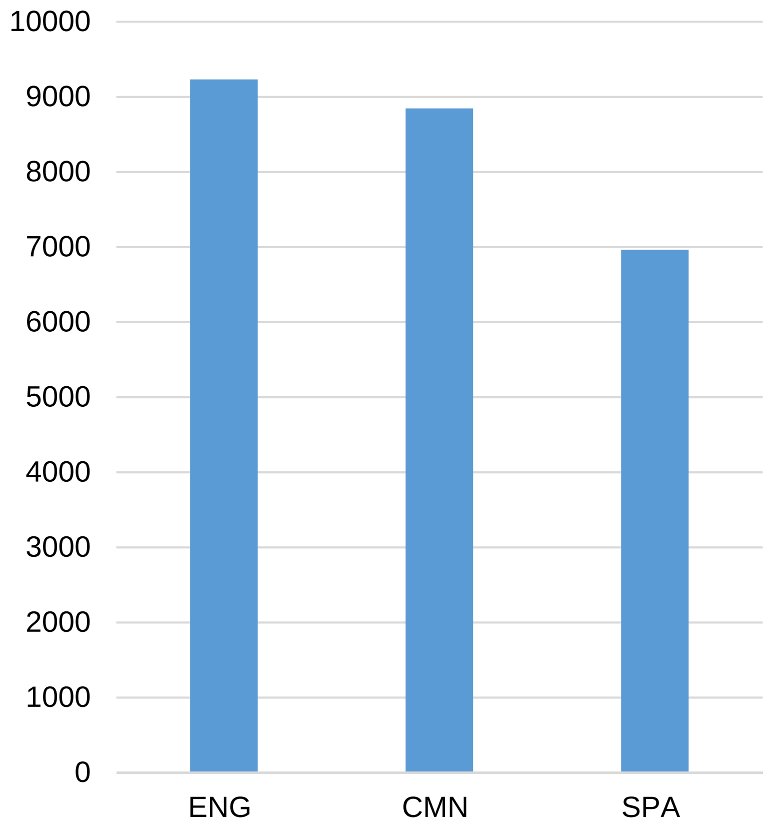
The Bo Xilai Event is a big event that shocked the whole world. It was ignited by Lijun running into US consulate in 2012 to bring Bogu Kailai’s killing to light. Will Bo Xilai end up in jail due to bribery and corruption; what will his wife end up with?

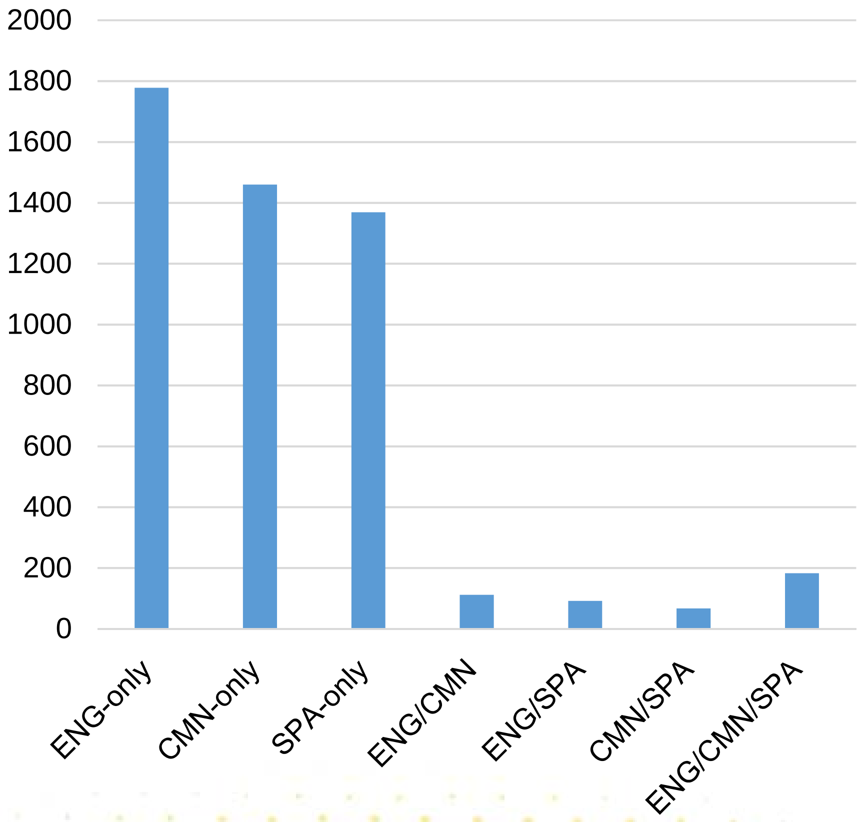
Entity				
PER	NAM	Bo Xilai, Bo Xilai	SPC	IND
	PRO	his		
PER	NAM	Lijun	SPC	IND
PER	NAM NOM	Bogu Kailai his wife	SPC	IND
GPE	NAM	US	SPC	IND
LOC	NOM	US consulate	SPC	IND
LOC	NOM	jail	Non SPC	

ED&L: New Knowledge Base Search

- Freebase went offline
- NIST took up development of new search engine
  - LDC assisted by testing, reporting problematic results
- Significant progress made but still work to be done
  - Some known problematic entities for search results rankings
  - Occasional slow response times

ED&L: Results





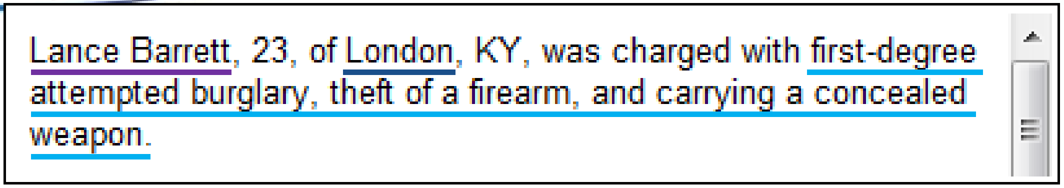
ED&L 2016 entity mentions

ED&L 2016 entity clusters

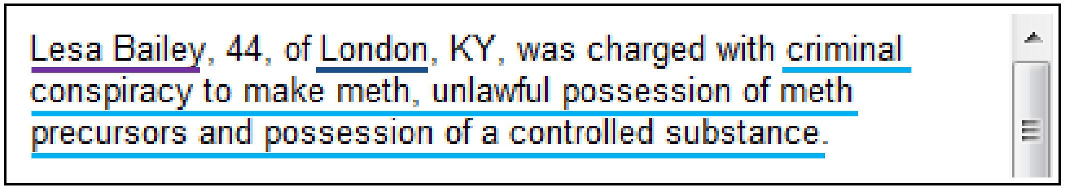
Cold Start

- Overall goal consistent with 2015
  - Create queries to navigate KBs, a “manual run”, and assess responses
- Changes to approach in 2016
  - Mono- to cross-lingual
  - Nominal entity mentions
  - Separate query and manual run development
    - In order to improve low recall seen in 2015

Cold Start QD & Manual Run



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

Cold Start: Other Changes

- Null queries produced manually
  - Auto-generated in 2015
- Multi-lingual support
  - Total pipeline extended from 4 passes to up to 18 across query and manual run development and assessment.
- Assessing nominals
  - Label correct and inexact responses as NAM or NOM
  - For NOM-only clusters, add'l corpus search for NAM mentions

Cold Start: Results

Year	Lang	Precision	Recall	F1
2015	ENG	81%	19%	30%
2016	ENG	79%	34%	48%
2016	CMN	75%	25%	38%
2016	SPA	86%	64%	73%
2016	Cross-lingual	67%	35%	36%

- Separation of QD and manual run had an impact!
  - Recall for English up 15%
  - Precision down 2%?

- Manual Run stats
  - 4,739 responses
  - 944 responses (20%) directly* assessed
    - 83% Correct
    - 14% Wrong
    - 3% Inexact

The End

- Questions?