

Overview of 2015 TAC KBP Event Nugget Tasks

Teruko Mitamura

Zhengzhong Liu

Eduard Hovy

Carnegie Mellon University



Three Tasks for Event Nugget

- Task 1: Event Nugget Detection
 - Evaluation Window: September 8–21, 2015
- Task 2: Event Nugget Detection and Coreference
 - Evaluation Window: September 8–21, 2015
- Task 3: Event Nugget Coreference
 - Evaluation Window: September 21–29, 2015



Task 1: Event Nugget Detection

- Detect explicit **mentions** of Events in text for English.
 - Must identify **all** event nuggets in the documents.
- For each Event Nugget, Event **Types/Subtypes** (9 types; 38 subtypes) must be identified.
- For each Event Nugget, **REALIS values** (*Actual*, *Generic*, *Other*) must be identified.
- Event Types/Subtypes and REALIS are defined in the Rich ERE guidelines created by LDC.



What are Events?

- Main verb
 - *The explosion **killed** 7 and **injured** 20.*
- Adjective or past-participle
 - *17 sailors were **killed**.*
 - *A **retired** congressman*
- Noun or pronoun
 - *The **attack** killed 7.*
- Resultatives and ongoing events
 - *All her grandparents are **dead**.*
 - *The newly **married** couple.* (state of being married)
 - *The **dying** man* (still in progress)

(from ERE Guidelines)



What are Events? (2)

- Single event
 - *Hamas launched an **attack**.*
 - *He carried out the **assassination**.*
 - *The hurricane left 20 **dead**.*
- Multiple words single event
 - *Foo Corp. had previously **filed Chapter 11** in 2001.*
- Two separate events
 - *Protestors **interrupted** their **meeting**.*
 - *An officer **witnessed** the **attack**.*
 - *Kennedy was **shot dead** by Oswald.*



What are Events? (3)

- Multiple verbs (aspectual verb + main verb)
 - .. continued to **bomb**
 - .. began **firing**

- Verb+Particle
 - Jane was **laid off** by XYZ Corp.
 - XYZ Corp **laid** Jane off.

(ERE Guideline: If the words occur non-contiguously, then we will only annotate the verb.)

- ERE guidelines do not allow any discontinuous event mentions.



9 Event Types/ 38 Subtypes from Rich ERE Annotation Guidelines: Events v2.6

1. **Life Events** (be-born, marry, divorce, injure, die)
2. **Movement Events** (transport-person, transport-artifact)
3. **Business Events** (start-org, merge-org, declare-bankruptcy, end-org)
4. **Conflict Events** (attack, demonstrate)
5. **Contact Events** (meet, correspondence, broadcast, contact)
6. **Personnel Events** (start-position, end-position, nominate, elect)
7. **Transaction Events** (transfer-ownership, transfer-money, transaction)
8. **Justice Events** (arrest-jail, release-parole, trial-hearing, charge-indict, sue, convict, sentence, fine, execute, extradite, acquit, appeal, pardon)
9. **Manufacture** (artifact)



REALIS Identification

- **ACTUAL:** the event actually happened
 - *The troops are **attacking** the city.* [Conflict.Attack, ACTUAL]
- **GENERIC:** the event is in general and not specific instance
 - *Weapon **sales** to terrorists are a problem.*
[Transaction.Transfer-Ownership, GENERIC]
- **OTHER:** the event didn't occur, future events, desired events, conditional events, uncertain events, etc.
 - *He plans to **meet** with lawmakers from both parties.*
[Contact.Meet, Other]



New challenge: Double Tagging

- (Type 1: 2 instances): *the murder of John on Tuesday and Bill on Wednesday.*
 - **murder**, argument=John, time=Tuesday
 - **murder**, argument=Bill, time=Wednesday
- (Type 2: 2 types): *the murder of John and Bill*
 - **Conflit.Attack**, murder
 - **Life.Die**, murder



Task 2: Event Nugget Detection and Coreference

- Task: Detect both Event Nugget and Coreference from the text
- Input: Unannotated document
- Output: Event Nugget Identification, Event Types/Subtypes, REALIS information, plus Event Coreference relations.



Task 3: Event Nugget Coreference

- Task: Identify Full Event Coreference links, given the annotated Event Nuggets, Event types and subtypes, and Realis in the text.
- Input: Document with fully annotated events
- Output: Event coreference relations



Analysis of training corpus

Stat.	Newsire	Discussion Forum
# Docs	81	77
# Mentions	2219	4319
# Clusters	350	804
# Tokens	30,257	109,187
# Singleton	1112	1073
Average Mention per Doc	27.48	56.09
Average Token per Doc	373.54	1418.01
# Token / # Mention	13.64	25.28
Average Cluster Size	3.16	4.03

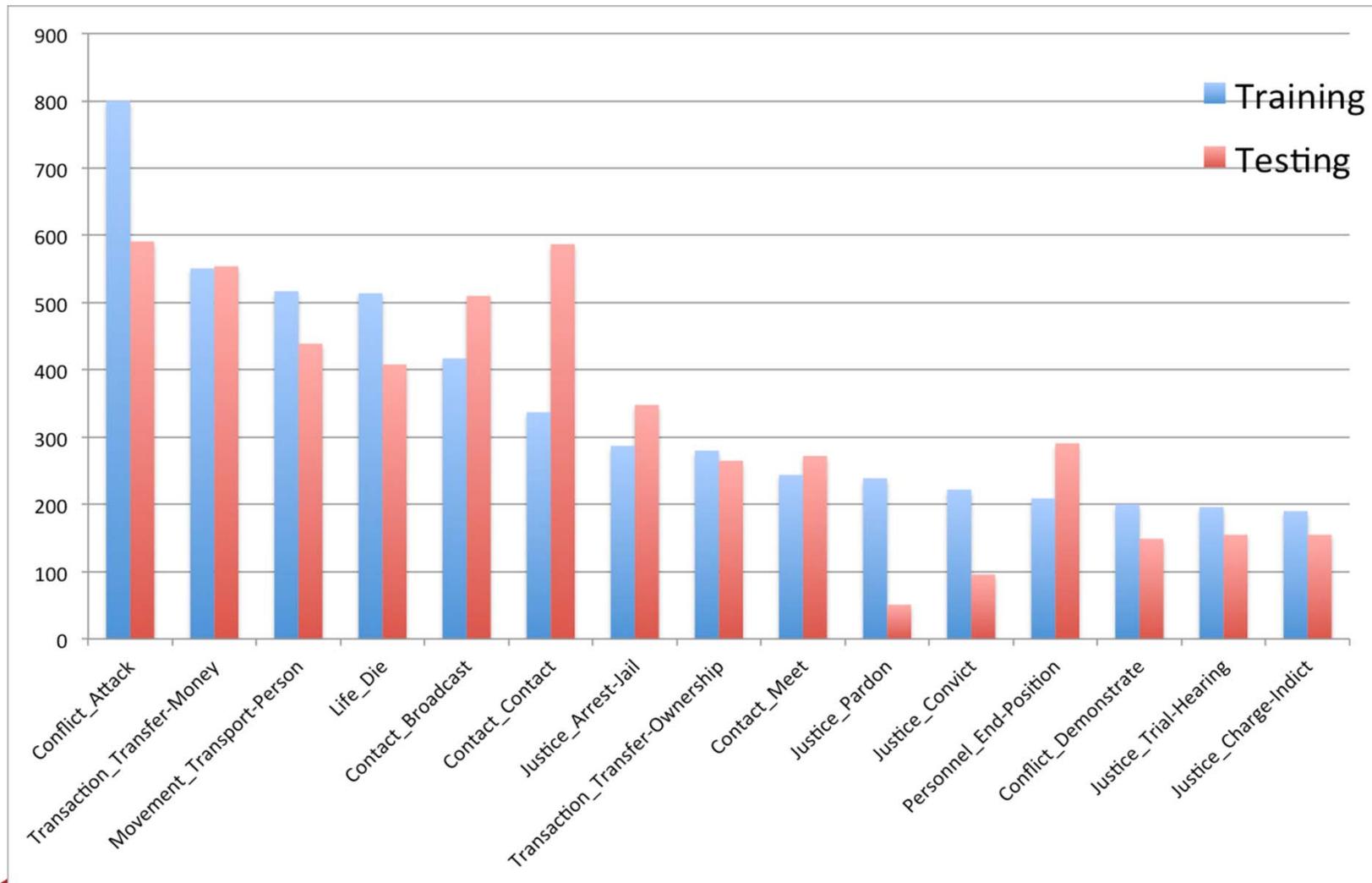


Comparison of training and testing dataset

Stat.	Training	Test
# Docs	158	202
# Mentions	6538	6438
# Clusters	1154	1050
# Tokens	139,444	98,414
# Singleton	2185	3075
Average Mention per Doc	41.38	31.88
Average Token per Doc	882.56	487.20
# Token / # Mention	21.33	15.29
Double Tagged Mentions	323	575
Average Cluster Size	3.77	3.20



Comparison: number of event nuggets by top 15 Event Types (Training vs Testing)



Submission format for all 3 tasks

- system-ID: unique ID assigned to each system run
- doc-ID: unique ID assigned to each source document
- mention ID: ID of the event nugget
- token ID list: list of IDs for the token(s) of the current mention
- mention-string: actual character string of event mention
- event-type: type.subtype
- Realis-value: one of ACTUAL, GENERIC, OTHER
- Confidence scores of event span: score between 0 and 1 inclusive (optional)
- Confidence scores of event type: score between 0 and 1 inclusive (optional)
- Confidence scores of Realis-value: score between 0 and 1 inclusive (optional)



Coreference format

- Relation name: this should always be @Coreference
- Relation Id: This is for bookkeeping purposes, which will not be read by the scorer. The relation id used in the gold standard files will be in form of “R[id]” (e.g., R3)
- Mentions Id list: list of event mentions in this coreference cluster, separated by comma (.). In terms of coreference, the ordering of event mentions does not matter.



Scoring

- For Event Nugget, systems were scored on the F-1 score of Precision and Recall over the gold standard.
- For Event Nugget coreference, systems were scored using the evaluation metrics used in CoNLL shared tasks.
- We ran four metrics (B^3 , CEAF-E, MUC, BLANC) and averaged the scores.



Evaluation

- Task 1: 38 runs were submitted by 14 teams:
 - RPI_BLENDER, LTI, UKP, wip, SYDNEY, LCC, UI_CCG, HITS, TEA_ICT, CMU_CS_event, BUPT_PRIS, ZIU-Insight, UMBC, IHMC
- Task 2: 19 runs were submitted by 8 teams:
 - RPI_BLENDER, LCC, UI_CCG, OSU, ZIU_Insight, UTD, BUPT_PRIS, UMBC
- Task 3: 16 runs were submitted by 6 teams:
 - LCC, UI_CCG, LTI, UKP, RPI_BLENDER, ntnu



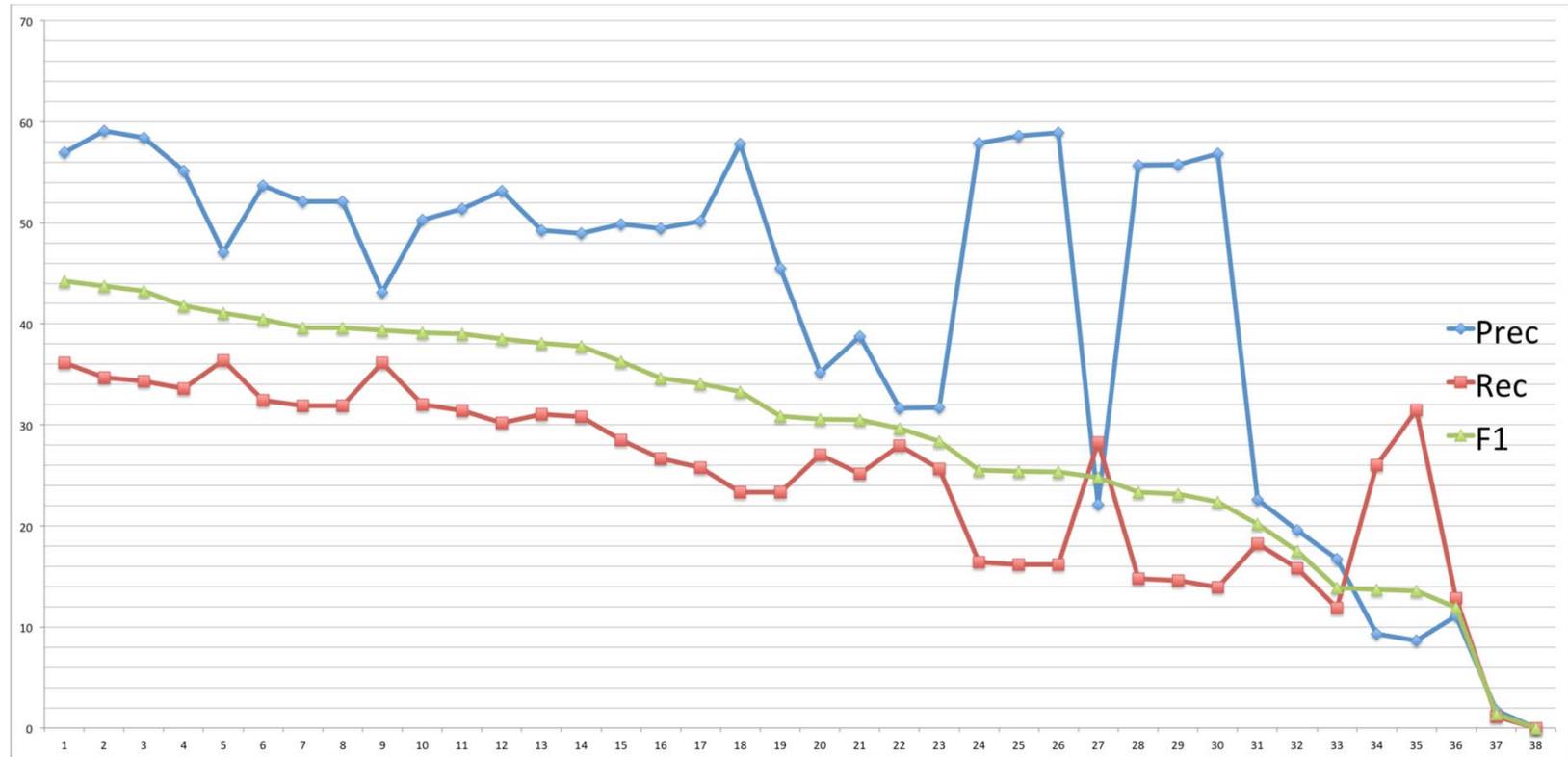
Task 1.
Event
Nugget
Detection
Results:

Highest
score from
each team
(38 runs, 14
teams, micro-
average F1)

	Plain	Type	Realis	All
1	65.31	58.41	49.16	44.24
2	63.66	57.18	48.70	41.77
3	62.49	55.83	47.05	41.04
4	60.77	55.56	45.54	39.58
5	60.30	53.97	43.89	39.33
6	59.80	51.97	42.87	38.06
7	59.68	49.42	40.35	36.28
8	57.36	48.16	38.30	33.27
9	55.38	42.73	37.44	29.67
10	51.38	41.57	37.04	28.35
11	46.03	35.17	31.21	25.54
12	38.53	34.67	28.16	24.81
13	34.50	32.60	24.27	23.32
14	33.81	26.93	18.09	13.89



Task 1: Event Nugget Detection Results (All systems' runs)



Task 2. Event Nugget and Coreference

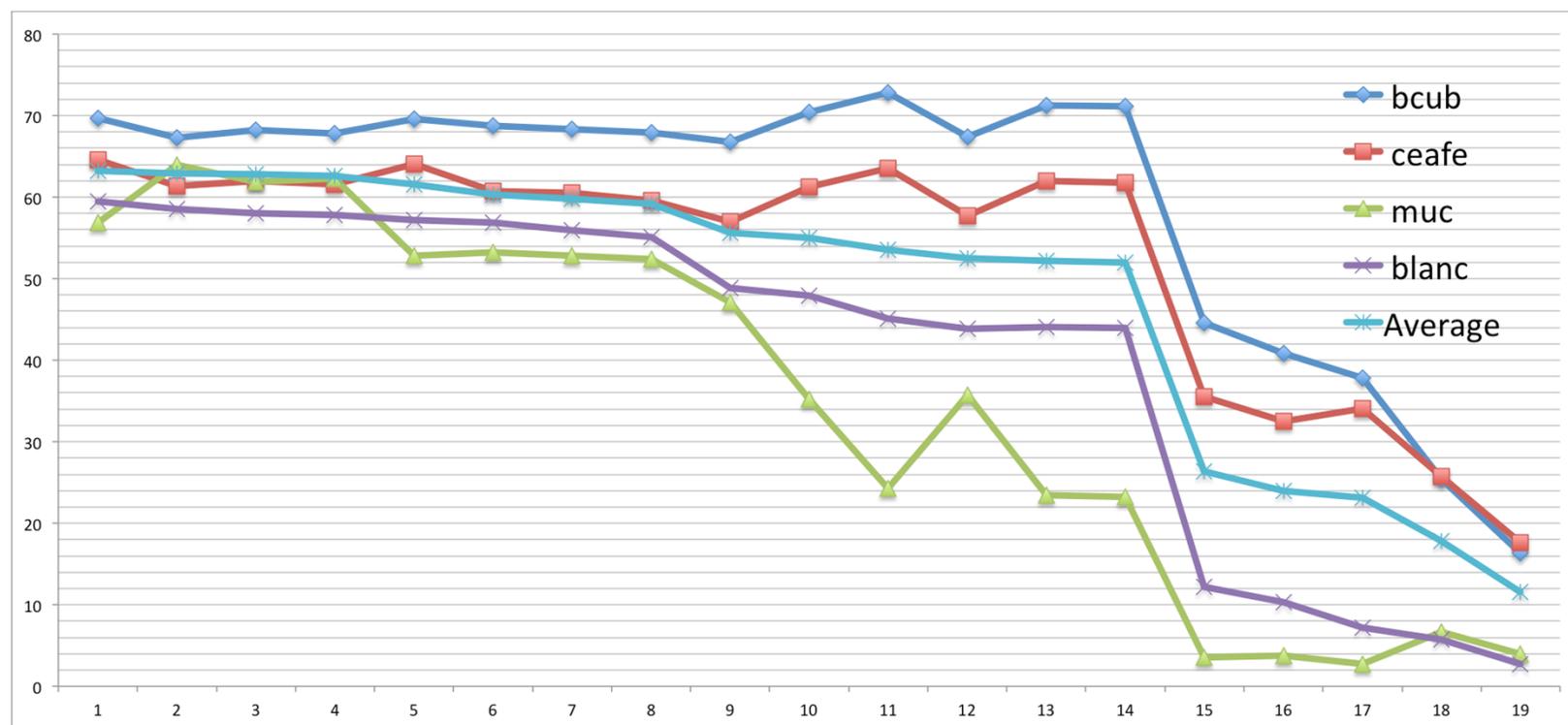
Highest score from each team (19 runs, 8 teams)

Micro Average of 4 metrics

	Plain	Type	Realis	Type+Realis	Coref
1	64.56	58.41	48.70	44.24	63.23
2	63.66	57.45	45.21	39.67	62.95
3	60.77	57.18	42.87	38.06	60.33
4	59.80	49.42	40.35	36.28	55.67
5	51.38	39.47	37.44	27.44	53.57
6	46.67	35.17	32.13	24.81	52.48
7	34.50	32.60	24.27	23.32	26.33
8	33.81	26.93	18.09	13.89	17.80



Task 2. Event Nugget Coreference Results



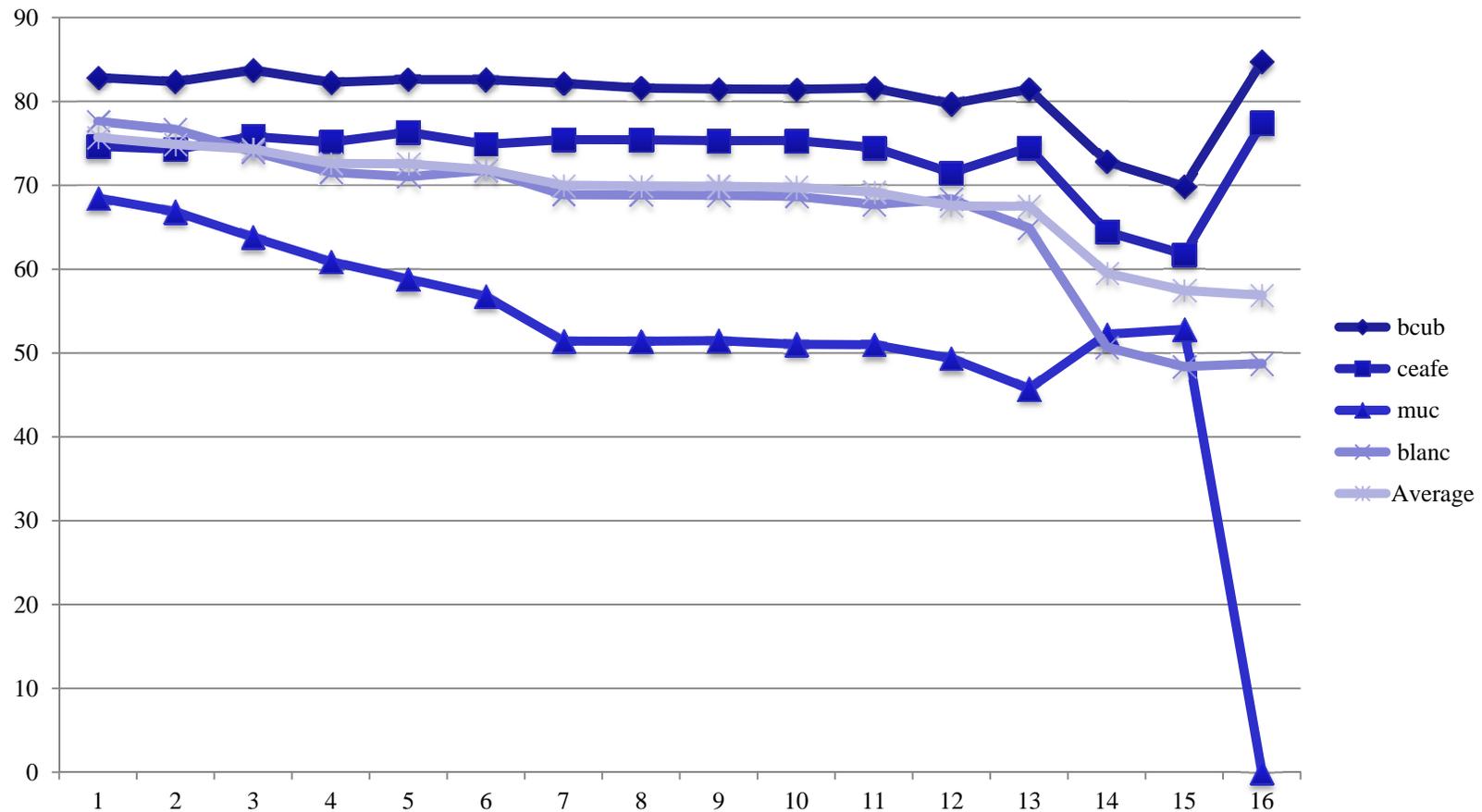
Task 3. Event Coreference Results:

Highest Score from each team
16 runs, 6 teams

	Average CoNLL Score
1	75.69
2	74.28
3	72.60
4	70.02
5	69.94
6	56.88



Task 3. Event Coreference Results



Example of Event Coreference

- *Lebanese Shiites rejoice at 'Night of Destiny' helicopter_{agent} [**crash**]_{e1}*
- *Hezbollah guerrillas fired shots into the air to rejoice at Tuesday's_{time} mid-air [**crash**]_{e2} between two Israeli helicopters_{agent} which killed more than 70 soldiers.*
- *Two Sikorsky troop carriers_{agent} [**collided**]_{e3} over northern Israel as they were flying to the occupied zone in south Lebanon.*
- *News of the [**crash**]_{e4} was greeted by automatic weapon fire which lasted around half an hour.*



Example of Coreference Resolution

- We analyze sentence roles to fill in the blanks
 - Compare information by role
 - Cycle over entities and events, propagating role fillers

	Event	Agent (-like)	Patient (-like)	Location	Time
	crash_e1	Helicopter			Tuesday
	crash_e2	<i>two Israeli helicopters</i>			Tuesday
Reason: time match, headword match, determiner	collided_e3	<i>Two Sikorsky troop carriers</i>		<i>northern Israel</i>	
	crash_e4	<i>two Israeli helicopters</i>			Tuesday

Reason: title_and_first_sentence , agent match, trigger match

Reason: Sentence proximity, agent match, trigger similar

Reason: time match, headword match, determiner

Language technologies institute

Baseline system for Task 3

- Singleton baseline: generated by putting each individual mention into cluster.
- Matching baseline: all mentions that have the same mention type and Realis are coreferent.



Baseline system results for Task 3

	B³	CEAF-E	MUC	BLANC	Average
Participants Systems Ave.	80.83	73.55	52.01	66.67	68.72
Singleton Baseline	78.10	68.98	0	48.88	52.01
Simple Type + Realis Match Baseline	78.40	65.82	69.83	76.29	71.94



Conclusion

- Event Nugget tasks attracted a lot of participants

Number of Participants and Runs

	# Teams	# Runs
Task 1	14	38
Task 2	8	19
Task 3	6	16
Total	28	73
Unique Teams	17	



Conclusion

- Event Nugget tasks are not easy
- It is harder to identify Event Types/Subtypes and Realis values
- Event Nugget Information is given (Task 3), the event coreference resolutions show high score



What is next?

- Event Nugget detection on cross documents of the same language or cross lingual documents?
- Event Nugget and Argument detection together?
- Event sequence detection with temporal ordering? (a pilot evaluation in 2016?)

