

Overview of Event Nugget Track TAC KBP 2016

Teruko Mitamura
Zhengzhong Liu
Eduard Hovy
Carnegie Mellon University



TAC KBP Event Detection Tasks for English, Chinese and Spanish

- Goal: The task aims to identify the explicit mentioning of Events in text.
 1. Event Nugget Detection Task
Evaluation Window: September 20 – October 3
 2. Event Nugget Detection and Coreference Task
Evaluation Window: September 20 – October 3

1. Event Nugget Detection Task for English, Chinese and Spanish

Participating systems will extract the following items:

1. Event Nugget Span Identification (character string)
2. Event Type and Subtypes (subset types of Rich ERE)
3. REALIS Value (one of: ACTUAL, GENERIC, OTHER)

2. Event Coreference Task for English, Chinese, and Spanish

- Input: Newswire and Discussion Forum documents (not annotated)
- Output: Event Nugget and Coreference Links
- Follow the notion of an **Event Hopper** (less strict coreference in ACE and light ERE)
- Corpus: Newswire and Discussion Forum

2015 TAC KBP EN tasks: 9 Event Types/ 38 Subtypes from Rich ERE Annotation Guidelines

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)



2016 TAC KBP EN Tasks: 8 Event Types/18 Subtypes from Rich ERE Annotation Guidelines

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]



Evaluation

- Task 1: Event Nugget Detection (Span, Type, Realis, All)
 - English: 14 teams were submitted
 - Chinese: 5 teams were submitted
 - Spanish: 2 teams were submitted
- Task 2: Event Nugget and Coreference
 - English: 6 teams were submitted
 - Chinese: 4 teams were submitted
 - Spanish: 2 teams were submitted



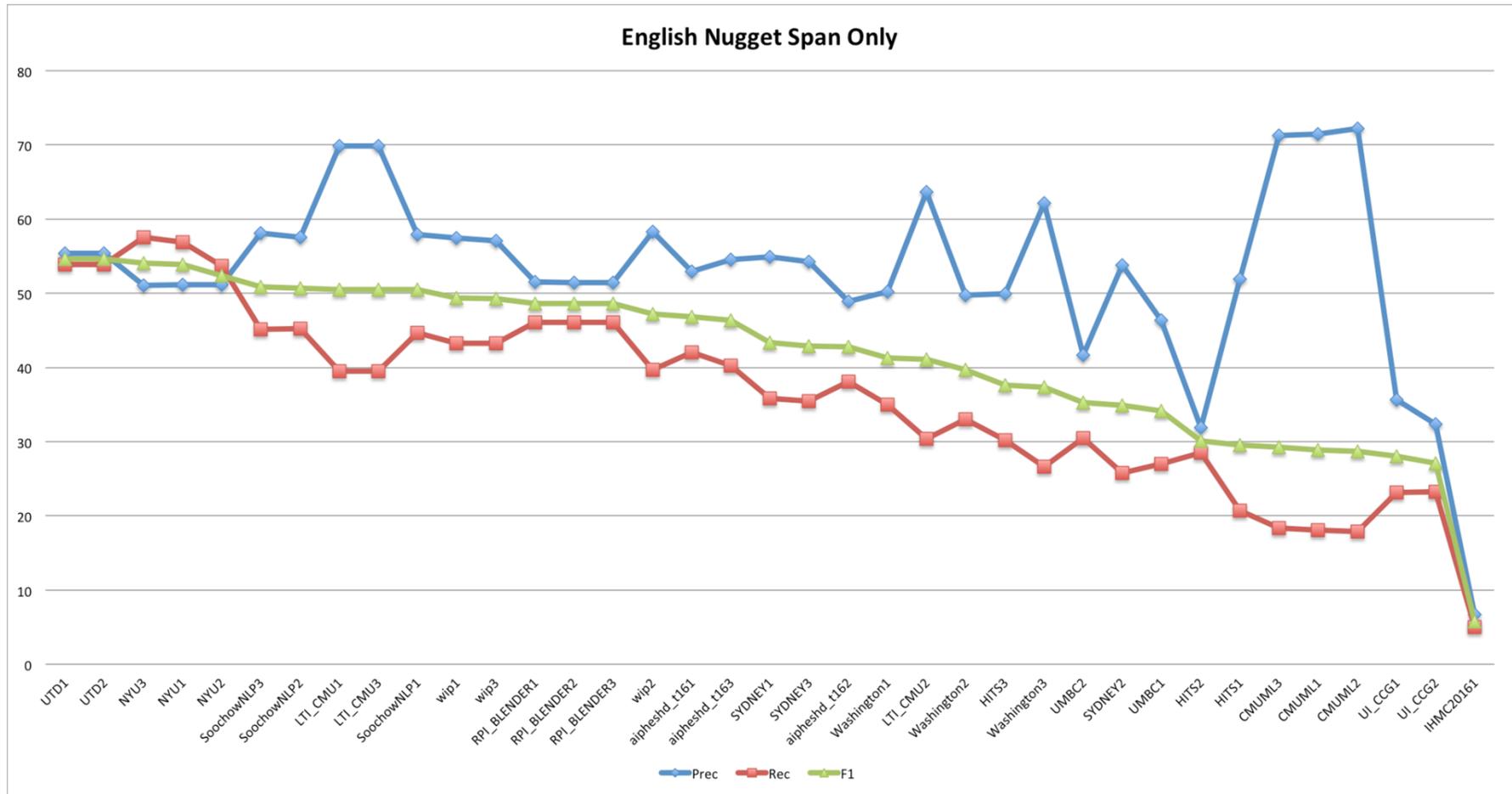
English Nugget Results (Span)

Highest score
from each
team

	Prec.	Recall	F1
UTD1	55.36	53.85	54.59
NYU3	51.02	57.52	54.07
SoochowNLP3	58.11	45.17	50.83
LTI-CMU1	69.82	39.54	50.49
wip1	57.49	43.29	49.39
RPI-BLENDER1	51.48	46.11	48.65
aipheshd-t161	52.89	42.06	46.85
SYDNEY1	54.87	35.80	43.33
Washington1	50.19	35.02	41.25
HITS3	49.91	30.22	37.64
UMBC2	41.70	30.51	35.24
CMUML3	71.27	18.37	29.21
UI-CCG1	35.68	23.14	28.07
IHMC20161	6.66	5.02	5.72



English Nugget (Span)



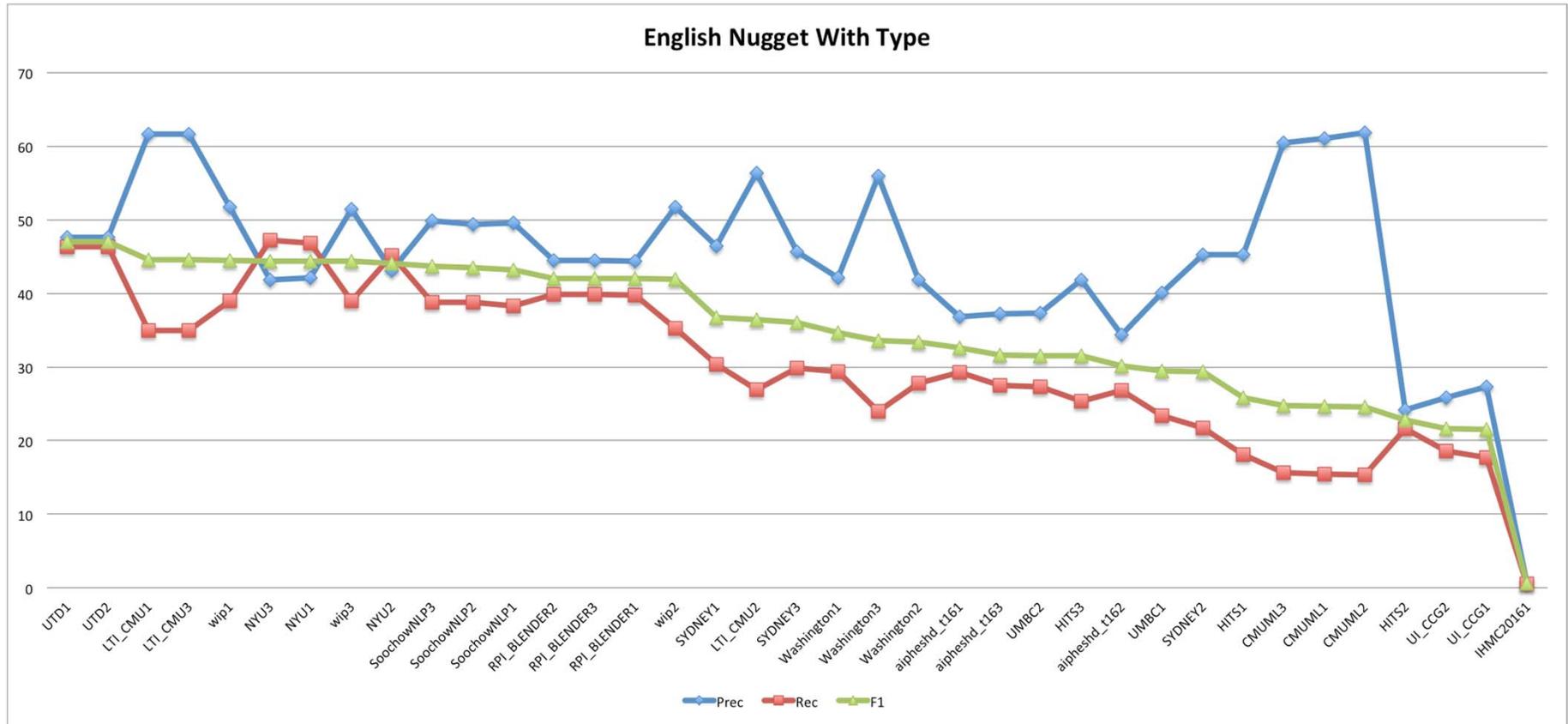
English Nugget Results (Type)

Highest
score from
each team

	Prec.	Recall	F1
UTD1	47.66	46.35	46.99
LTI-CMU1	61.69	34.94	44.61
wip1	51.76	38.98	44.47
NYU3	41.88	47.21	44.38
SoochowNLP3	49.92	38.81	43.67
RPI-BLENDER2	44.51	39.87	42.07
SYDNEY1	46.48	30.33	36.70
Washington1	42.15	29.41	34.65
aipheshd-t161	36.83	29.28	32.62
UMBC2	37.36	27.33	31.57
HITS3	41.79	25.30	31.52
CMUML3	60.44	15.58	24.77
UI-CCG2	25.81	18.53	21.57
IHMC20161	0.69	0.52	0.59



English Nugget (Type)



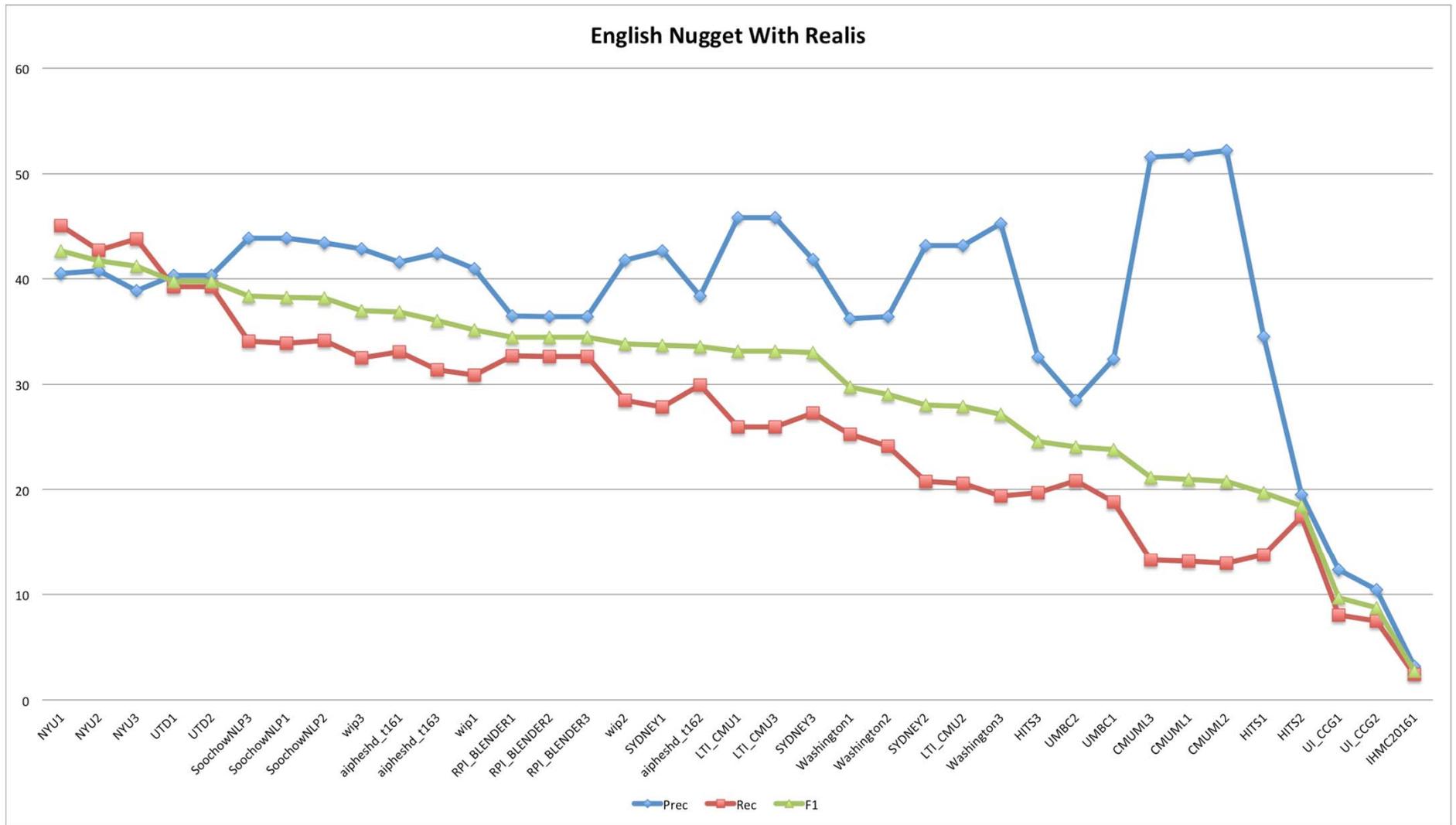
English Nugget Results (Realis)

Highest score
from each
team

	Prec.	Recall	F1
NYU1	40.53	45.07	42.68
UTD1	40.34	39.23	39.78
SoochowNLP3	43.84	34.08	38.35
wip3	42.86	32.49	36.96
aipheshd-t161	41.57	33.06	36.83
RPI-BLENDER1	36.47	32.67	34.46
SYDNEY1	42.67	27.84	33.69
LTI-CMU1	45.78	25.93	33.11
Washington1	36.20	25.25	29.75
HITS3	32.56	19.72	24.56
UMBC2	28.45	20.81	24.04
CMUML3	51.54	13.29	21.13
HITS1	34.52	13.77	19.68
UI-CCG1	12.33	8.00	9.70
IHMC20161	3.20	2.40	2.75



English Nugget (Realis)



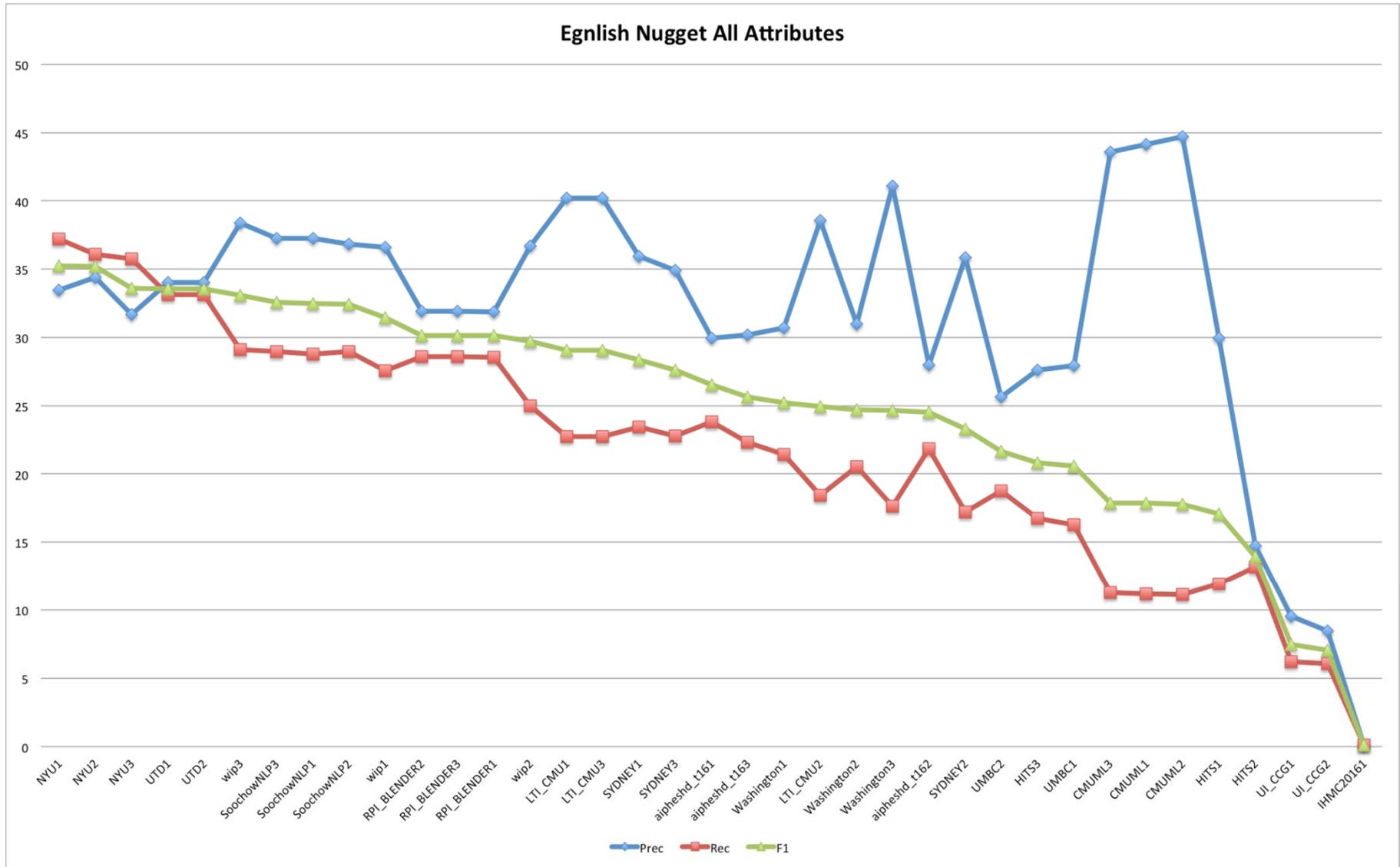
Task 1: English Nugget Results (All)

Highest score
from each
team

	Prec.	Recall	F1
NYU1	33.47	37.21	35.24
UTD1	34.05	33.12	33.58
wip3	38.38	29.1	33.1
SoochowNLP3	37.26	28.97	32.59
RPI-BLENDER2	31.92	28.59	30.16
LTI-CMU1	40.19	22.76	29.06
SYDNEY1	35.93	23.45	28.38
aipheshd-t161	29.94	23.81	26.53
Washington1	30.71	21.42	25.24
UMBC2	25.64	18.76	21.67
HITS3	27.63	16.73	20.84
CMUML3	43.6	11.24	17.87
UI-CCG1	9.52	6.18	7.49
IHMC20161	0.13	0.1	0.11



Task 1: English Nugget (All)

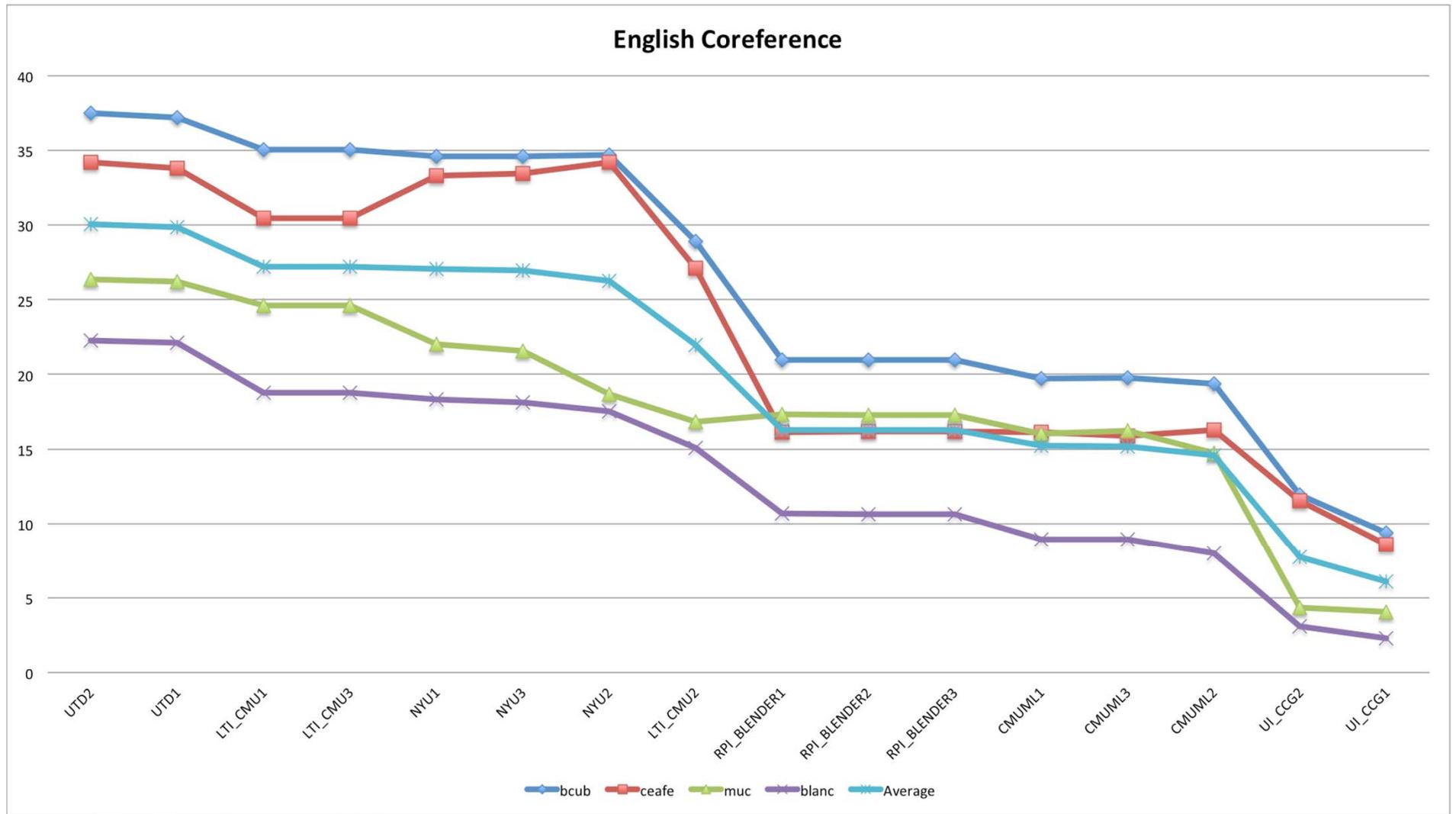


Task 2: English Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
UTD2	37.49	34.21	26.37	22.25	30.08
LTI-CMU1	35.06	30.45	24.60	18.79	27.23
NYU1	34.62	33.33	22.01	18.31	27.07
RPI-BLENDER1	20.96	16.14	17.32	10.67	16.27
CMUML1	19.74	16.13	16.05	8.92	15.21
UI-CCG2	11.92	11.54	4.34	3.10	7.73



Task 2: English Coreference



Observations on English Nugget and Coreference

- Most systems tend to have higher precision than recall.
- The best Event Type detection F1 score was 46.99, whereas the best F1 score from 2015 was 58.41.
- The average of Event Type F1 score is higher: 0.27, compared to 0.24 in 2015.
- The best Event Coreference F1 score: 30.08, compared to 39.12 in 2015.
- Part of the reasons may be caused by the reduction of Event Types/Subtypes to 18 from 38 and many difficult and ambiguous event types remained: Transaction, Contact, etc.

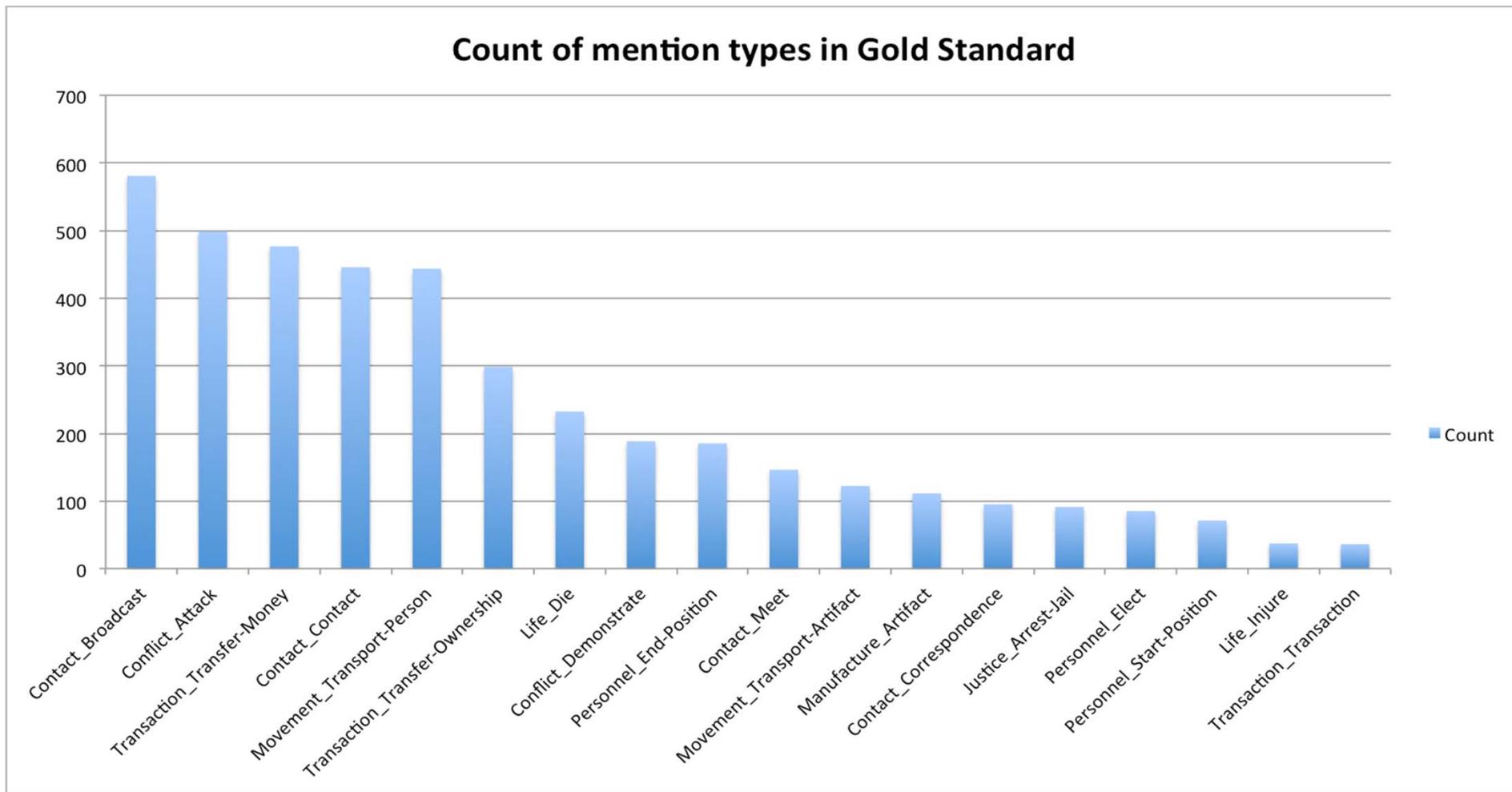


Observations on English Nugget and Coreference (2)

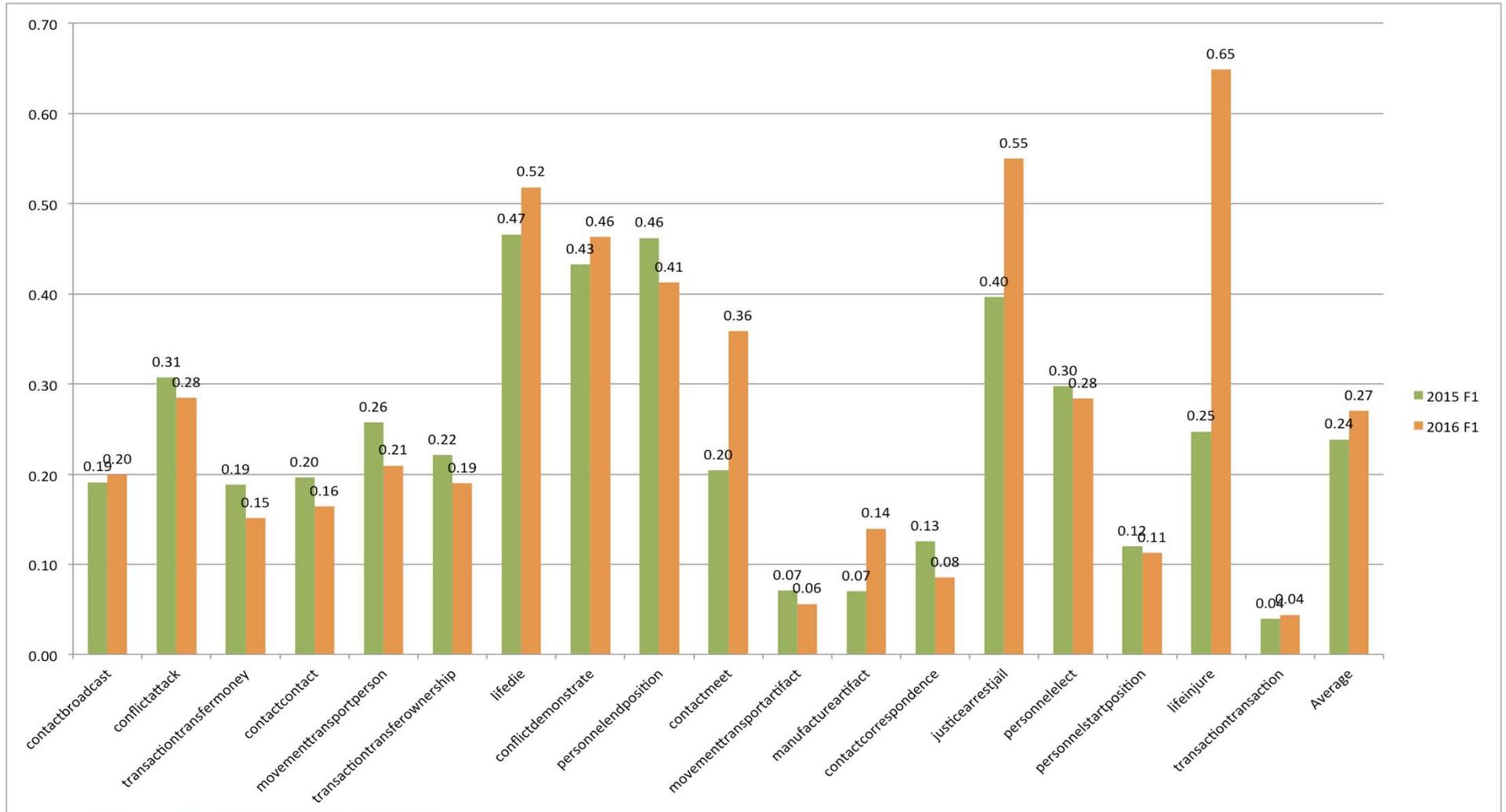
- *Contact-Broadcast, Contact-Contact, Transaction-TransferMoney, Transaction-TransferOwnership* event types contribute around 50% of the total misses, while they appear 43% in the test data.
- *Transaction-TransferOwnership* and *Transaction-Transaction* are easily misclassified.
- *Movement-TransportArtifact* was easily misclassified with *Movement-TransportPerson*.
- *Contact-Broadcast* was easily misclassified with *Contact-Contact*.



Count of mention types in Gold Standard



Event Types F1 Score Comparisons between 2015 and 2016



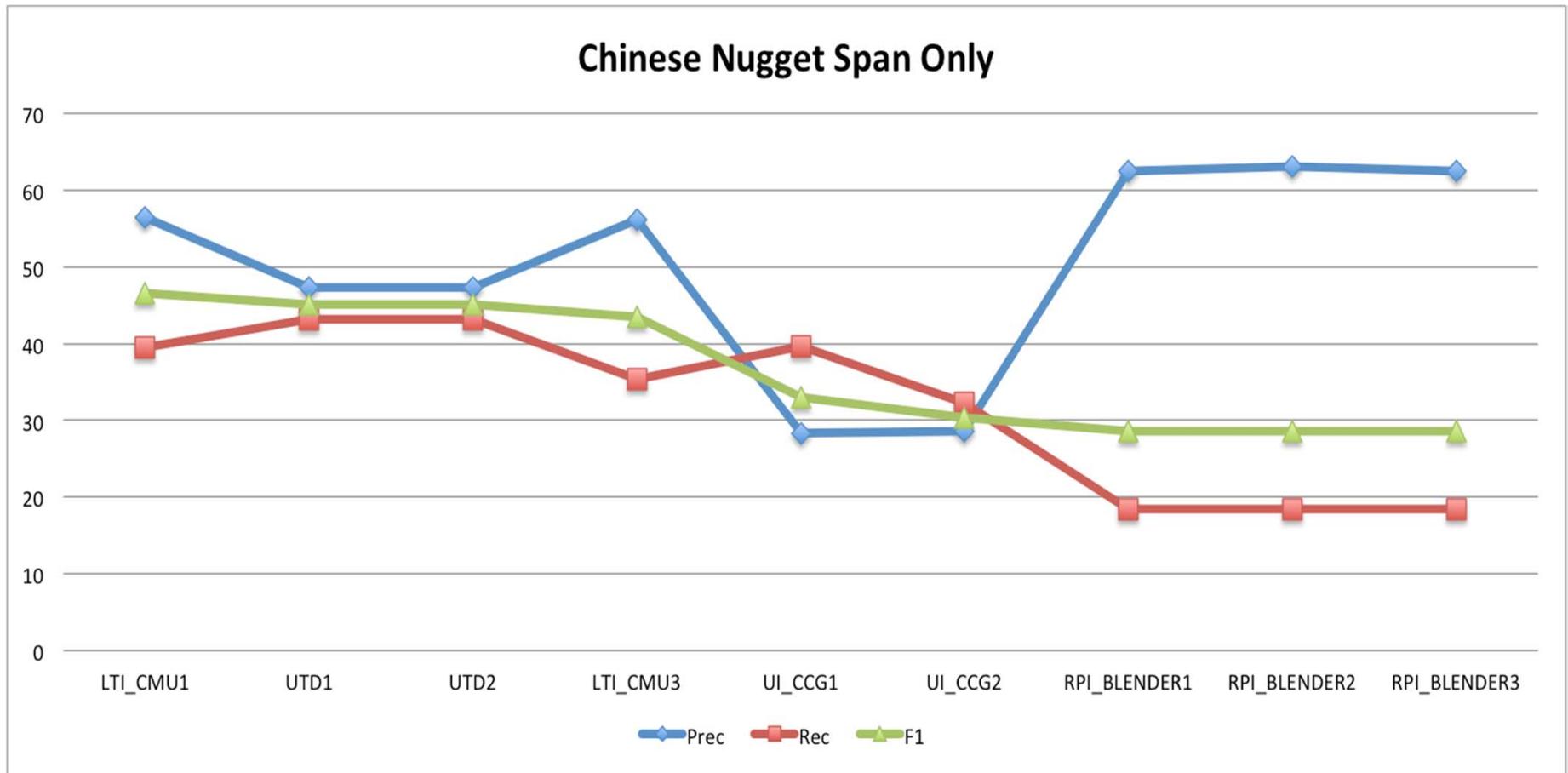
Chinese Nugget Results

Highest score
from each team

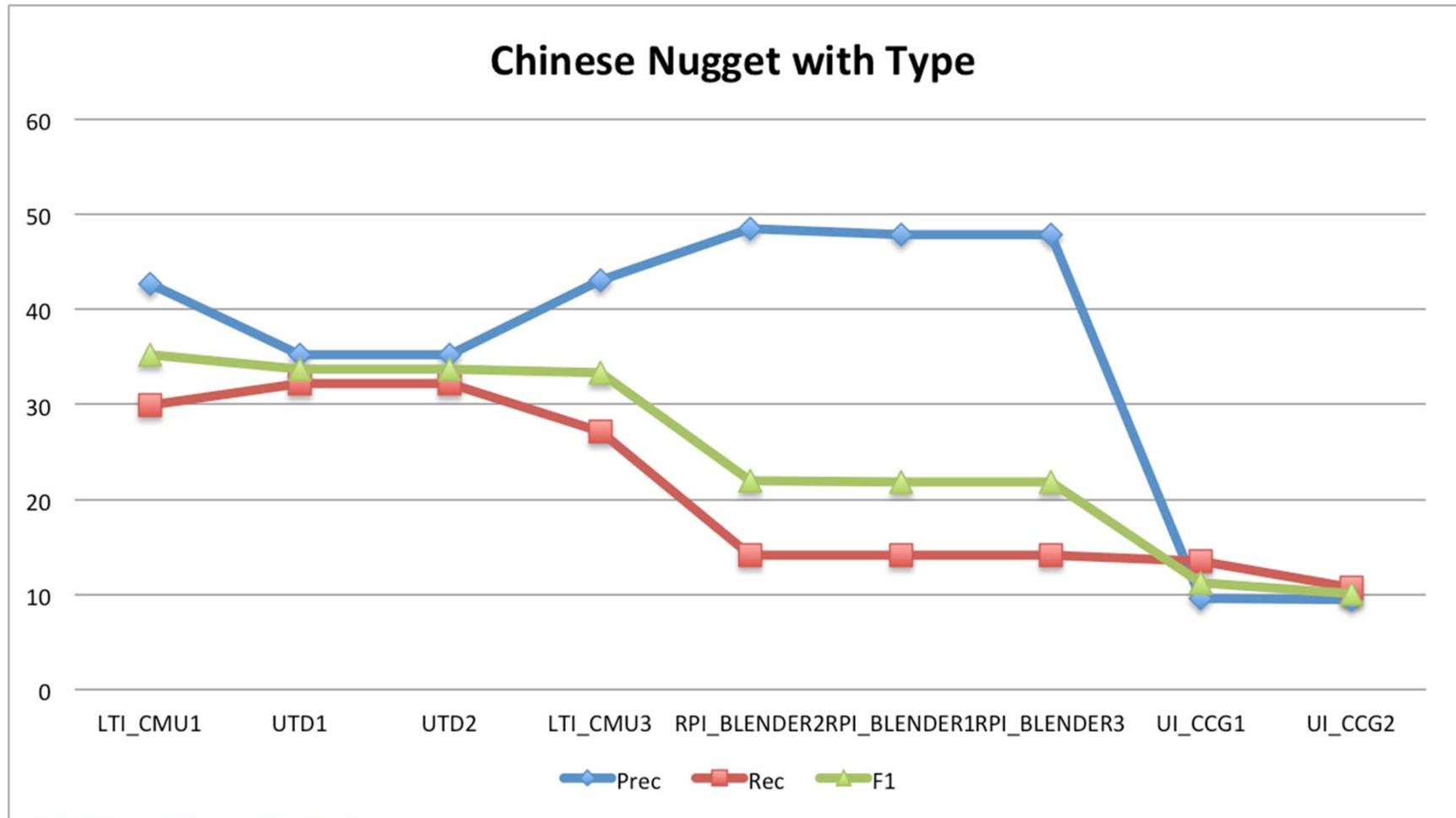
		Prec.	Recall	F1
Span	LTI-CMU1	56.46	39.55	46.52
	UTD1	47.23	43.16	45.1
	LTI-CMU3	56.19	35.35	43.4
	UI-CCG1	28.34	39.61	33.04
	RPI-BLENDER1	62.46	18.48	28.52
Type	LTI-CMU1	50.72	35.53	41.79
	UTD1	41.9	38.29	40.01
	LTI-CMU3	49.7	31.26	38.38
	UI-CCG1	24.01	33.55	27.99
	RPI-BLENDER2	59.87	17.5	27.08
Realis	LTI-CMU1	42.7	29.92	35.18
	UTD1	35.27	32.23	33.68
	LTI-CMU3	43.11	27.12	33.29
	RPI-BLENDER2	48.46	14.16	21.92
	UI-CCG1	9.65	13.49	11.25
All	LTI-CMU1	38.91	27.26	32.06
	UTD1	31.76	29.02	30.33
	LTI-CMU3	38.54	24.25	29.77
	RPI-BLENDER2	46.69	13.65	21.12
	UI-CCG1	8.31	11.62	9.69



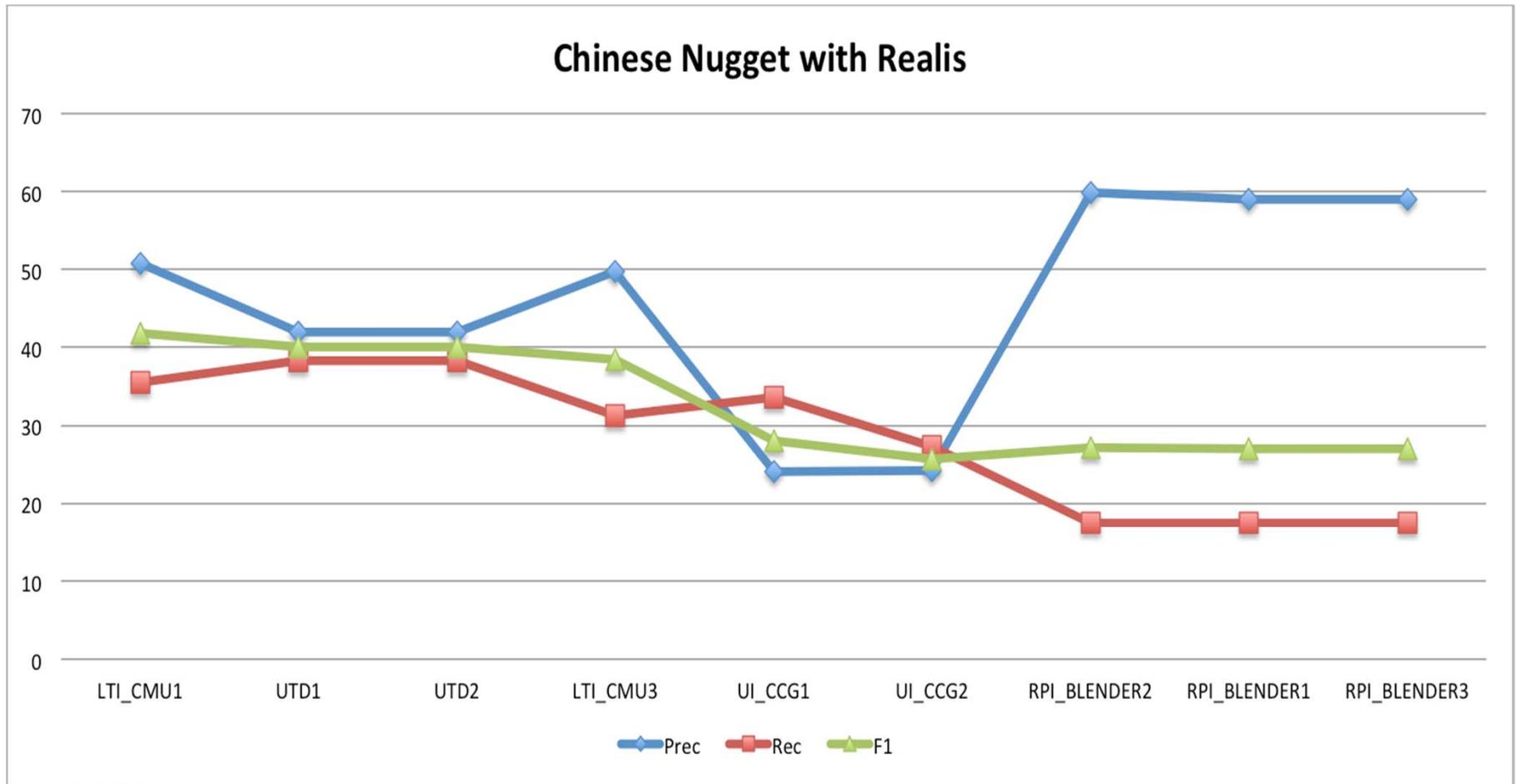
Results: Chinese Event Nuggets (Span)



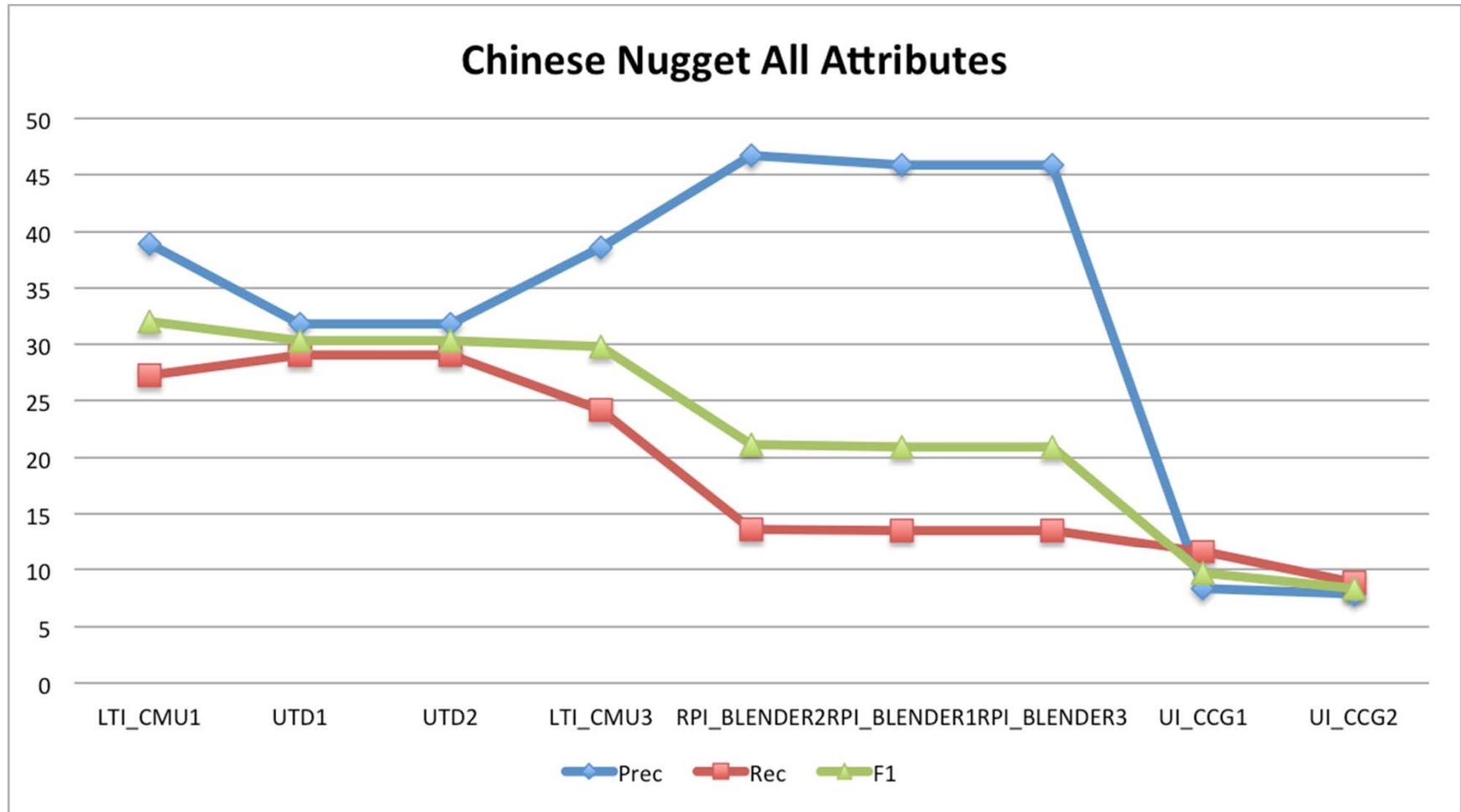
Results: Chinese Event Nuggets (Type)



Results: Chinese Event Nuggets (Realis)



Results: Chinese Event Nuggets (All)

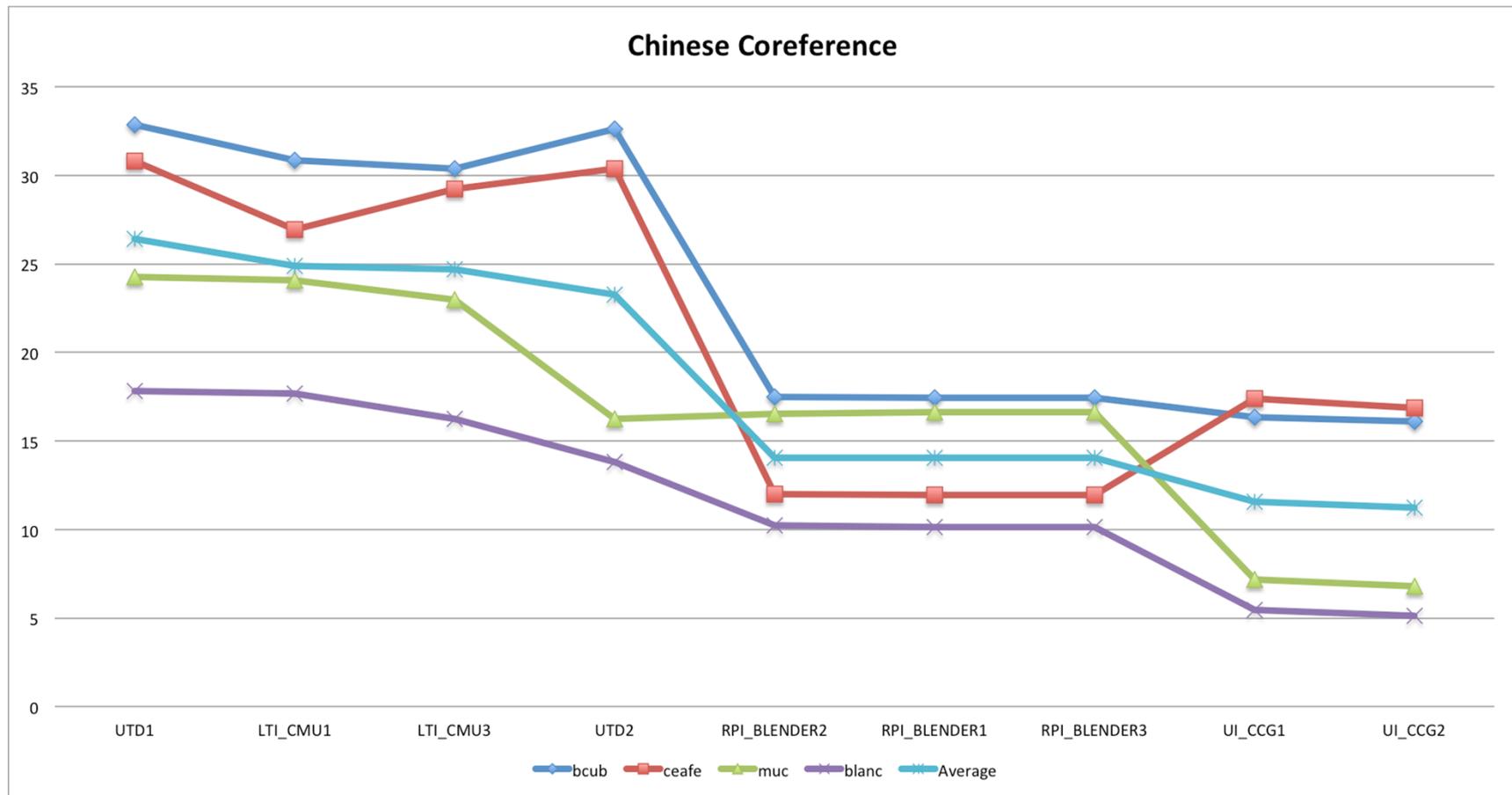


Chinese Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
UTD1	32.83	30.82	24.27	17.80	26.43
LTI-CMU1	30.83	26.95	24.07	17.67	24.88
RPI-BLENDER2	17.46	11.97	16.51	10.23	14.04
UI-CCG1	16.34	17.38	7.15	5.46	11.58



Results: Chinese Coreference



Observations on Chinese Nugget and Coreference

- Datasets are all from discussion forum (no newswire data annotated) for training
- 4 teams participated in Chinese
- The best performance of F1 All is 32.06, whereas 35.24 in English.
- Tokens in Chinese may be composed by several characters.
- One character tokens are more ambiguous and difficult to detect event types.
e.g. 打 in “attack” (打人) “call by phone” (打电话)
- There are 17 single-character nuggets in top 20 most frequent event nuggets.



Chinese Dataset Issue

- Chinese dataset doesn't seem to be fully annotated.
- Top 5 double character nuggets in RichERE
- Top 5 double character nuggets in ACE 2005

Token	Annotated	Total	%
战争	96	223	43.05%
死亡	24	33	72.73%
暗杀	22	40	55.00%
入侵	18	22	81.82%
自杀	17	33	51.52%

Token	Annotated	Total	%
冲突	100	119	84.03%
访问	64	90	71.11%
受伤	53	59	89.83%
死亡	46	50	92.00%
前往	44	52	84.62%



Spanish Nugget Results

		Prec.	Recall	F1
Span	RPI-BLENDER1	48.11	51.31	49.66
	UI-CCG1	35.12	22.85	27.69
Type	RPI-BLENDER2	36.06	38.47	37.23
	UI-CCG1	26.01	16.92	20.50
Realis	RPI-BLENDER1	35.90	38.29	37.06
	UI-CCG1	11.03	7.17	8.69
All	RPI-BLENDER2	26.67	28.45	27.53
	UI-CCG1	8.43	5.48	6.64



Spanish Event Coreference

	B^3	CeafE	MUC	BLANC	Aver.
RPI-BLENDER1	22.05	18.56	19.04	12.43	18.02
UI-CCG1	11.25	10.5	6.76	3.24	7.94

- Only 2 teams participated in Spanish
- The scores in Event Nugget and Coreference tasks are lower than English and Chinese.



What is next?

TAC KBP 2017 Event Nugget Tasks

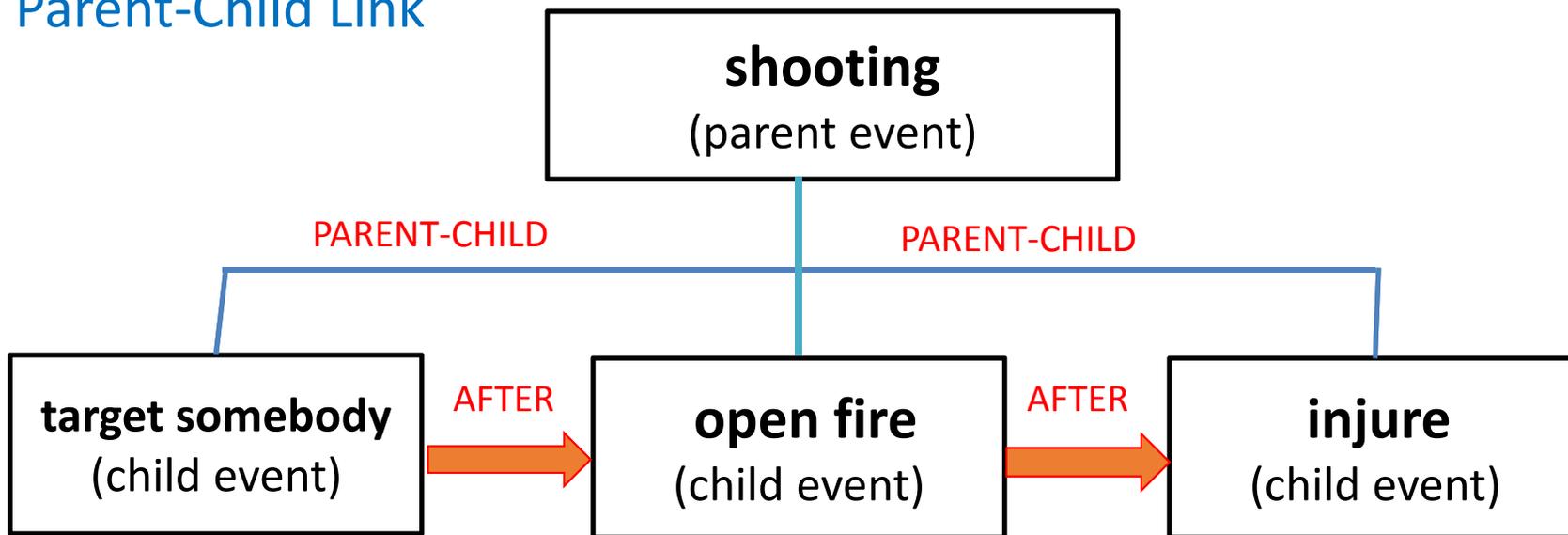
Tasks are under-discussion

1. Event Nugget Detection Task for English, Chinese, Spanish (Multilingual, Cross-Doc?)
2. Full Event Coreference Task for English, Chinese, Spanish (Multilingual, Cross-Doc?)
3. Subsequence Linking task (after DEFT pilot evaluation) for English, will be organized by CMU

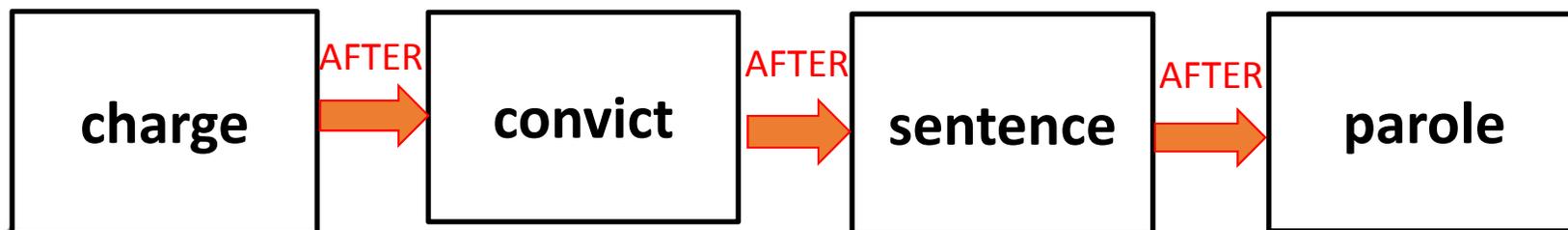


After Link
Parent-Child Link

SUB-SEQUENCE OF SHOOTING



SUB-SEQUENCE OF JUDICIAL PROCESS



Two Types of Event-Event Relation Linking: AFTER Link and Parent-Child Link

- **AFTER Link Relation:**
 - Represents a **temporal sequence** between child events in a subevent cluster
 - Can be linked between child events **with or without** a parent event
- **Parent-Child Link Relation:**
 - Sub-event cluster detection



Event Subsequence Linking Tasks for English in 2017

- Goal: Extract Subsequence of events within Doc
 - Input: Event nugget annotated files
 - Outputs: (1) After links; (2) Parent-Child links
- Corpus: Newswire and Discussion Forum in English
- Training data and Annotation Guidelines will be available for interested participants
 - Annotation tool: Modified Brat tool
- Scorer, submission validation scripts and submission format will be created by CMU

Joint Workshop in ACL 2017

Event Workshop Series (NAACL 2013, 2014, 2015, 2016) + Computing News Storylines Workshop Series (ACL 2015, EMNLP 2016) = Events and Stories in the News



Questions?

