

# OpenEL: An Annotated Corpus for Entity Linking and Discourse in Open Domain Dialogue

LREC 2022  
Marseille

Wen Cui<sup>1</sup>

Leanne Rolston<sup>2</sup>

Marilyn A. Walker<sup>1</sup>

Beth Ann Hockey<sup>2</sup>

<sup>1</sup>University of California, Santa Cruz

<sup>2</sup>LivePerson Inc.

# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work

# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work

# Motivations

- Growth of interest in open-domain dialogue systems in the past few years
- Building open-domain dialogue systems converse like humans remains extremely challenging.



# Motivations

- Named Entity Linking (NEL) is the task of mapping named entity mentions to a target knowledge base.
- Coreference resolution is the KEY to NEL over dialogue

---

A1: The Other Woman Q14272676 is such a funny movie.  
B1: That Q14272676's the one with Cameron Diaz Q44380 and Leslie Mann Q229011, right?  
A2: That's right. Do you enjoy their Q44380, Q229011 movies?  
B2: Oh yes. Leslie Mann Q229011 plays neurotic characters very well.  
A3: You're right. She Q229011 was great in Knocked Up Q222800 as Debbie.  
B3: I liked her Q229011 in Big Daddy Q509025, too. That was the first movie I ever saw her Q229011 in.  
A4: Really? It's been awhile since I watched an Adam Sandler Q132952 movie.  
B4: I love his Q132952 movies. They always make me laugh.  
A5: They are pretty funny. Happy Gilmore Q1313063 is my favorite Adam Sandler Q132952 movie.  
B5: That Q1313063's a good one. The cameo by Bob Barker Q381178 always cracks me up.

---

Figure 1: An annotated conversation in OpenEL

# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work

# Challenges

- Name variations
- Open-ended nature
- Ambiguous reference

---

A1: You psyched about this coming NFL season?  
B1: Oh yeah. Can't wait to see my **Giants** in action.

---

A2: What do you think about Nicki Minaj?  
B2: My favorite is **Anaconda**.

---

Figure 2: potentially ambiguous reference resolved by dialogue context

# Contributions

- Release OpenEL with both NEL and anaphora annotations
- Analyze annotation quality
- Compare existing NEL tools and establish baselines



# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work

# OpenEL Corpus: Annotation

- Sampled existing open-domain corpus Edina (Krause et al., 2017)
- Preprocessed using DBpedia Spotlight (Mendes et al., 2011)
- 3 experienced annotators verified/corrected spans and Wikidata IDs, marked coreference.

```
"text": "Mandy Moore used to sing. Did they ever record together?",
```

```
"speaker-id": "A",
```

```
"entities": [ "annotator-1": [ "span": [0, 10],  
"surface-form": "Mandy Moore",  
"wikidata-id": ["Q187832"], "is-anaphora": false, "span": [31, 34],  
"surface-form": "they", "wikidata-id": ["Q160009", "Q187832"], "is-anaphora": true],
```

```
"annotator-2": [ "span": [0, 10],  
"surface-form": "Mandy Moore",  
"wikidata-id": ["Q187832"], "is-anaphora": false, "span": [31, 34],  
"surface-form": "they", "wikidata-id": ["Q160009", "Q187832"], "is-anaphora": true],
```

```
"annotator-3": [ "span": [0, 10],  
"surface-form": "Mandy Moore",  
"wikidata-id": ["Q187832"], "is-anaphora": false, "span": [31, 34],  
"surface-form": "they", "wikidata-id": ["Q160009"], "is-anaphora": true],
```

```
"ground-truth": [ "span": [0, 10 ],  
"surface-form": "Mandy Moore",  
"wikidata-id": ["Q187832"], "is-anaphora": false, {"span": [31, 34],  
"surface-form": "they", "wikidata-id": ["Q160009", "Q187832"], "is-anaphora": true}]]
```

Figure 3: JSON formatted examples in OpenEL

# OpenEL Corpus: Statistics

- First substantial entity linking corpus publicly available for open-domain dialogue

<b>Corpus Properties: Overall</b>	
conversations	179
conversational turns	2,570
topics	12
entity mentions (incl. anaphors)	2,263
entity mentions (excl. anaphors)	1,205
anaphors	1,058
unique entities	576
average conversation length (turns)	14.4
average mentions per entity	3.9
average mentions per conversational turn	0.9

Table 1: Counts and averages representing the composition of the corpus

# OpenEL Corpus: Statistics

Corpus Properties: By Topic			
	# Convs	# Turns	% Entity
music	33	20	0.60 / 0.40
pop	32	10	0.82 / 0.60
star wars	30	10	0.72 / 0.62
baseball	34	20	0.53 / 0.42
comedy	22	10	0.78 / 0.61
rap hiphop	5	10	0.76 / 0.68
action	4	10	0.72 / 0.55
basketball	4	20	0.68 / 0.60
horror	4	10	0.53 / 0.30
movies	3	20	0.80 / 0.55
nfl football	4	20	0.51 / 0.37
rock	4	10	0.78 / 0.63

Table 2: Number of conversations per topic, number of turns per conversation, proportion of turns with an entity (incl anaphors / excl anaphors).

# OpenEL Corpus: Agreement

Pairwise Annotator Agreement		
	Cohen Kappa	F-Score
Span & Wikidata ID	0.79	0.61
	0.82	0.67
	0.82	0.68
(average)	0.81	0.67
Span	0.83	0.73
	0.86	0.74
	0.84	0.75
(average)	0.84	0.74
Wikidata ID	0.85	0.76
	0.86	0.77
	0.89	0.83
(average)	0.87	0.79

Table 3: Pairwise annotator agreement: Cohen Kappa and F-scores

# OpenEL Corpus: Disagreement

---

## Annotation

---

(Star Wars topic)

A1:Do you like the

star wars movies *Q22092344*?

A2:Do you like the star wars movies *Q462*?

A3:Do you like the star wars *Q462* movies?

---

(Baseball topic)

A1:Wouldn't think of missing it. I never miss

that *Q213417* matchup.

A2:Wouldn't think of missing it. I never miss

that matchup *Q213417*.

A3:Wouldn't think of missing it. I never miss  
that matchup.

---

# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work



# Experiments

- Setup: **UTTERANCE**, **DIALOGUE**, **DISCOURSE**

A: Now if Belt will just start hitting ...

B: Dude, if he can keep getting on base by taking walks, I don't care if he never gets another home run.

**DIALOGUE**

A: A walk's as good as a single. I'll take it.

B: You hate to see all that power go to waste, though.

A: Yeah. Maybe Bonds could work with him. Give him some pointers.

**UTTERANCE**

B: Looks like Posey's all recovered from his concussion.

# Experiments

- Setup: **UTTERANCE**, **DIALOGUE**, **DISCOURSE**

**DISCOURSE**

A: Now if Belt will just start hitting ...

B: Dude, if **he** can keep getting on base by taking walks, I don't care if **he** never gets another home run.

A: A walk's as good as a single. I'll take it.

B: You hate to see all that power go to waste, though.

A: Yeah. Maybe Bonds could work with **him**. Give **him** some pointers.

B: Looks like Posey's all recovered from his concussion.

# Experiments

- Setup: **UTTERANCE**, **DIALOGUE**, **DISCOURSE**

**DISCOURSE**

A: Now if Belt will just start hitting ...

B: Dude, if Belt can keep getting on base by taking walks, I don't care if Belt never gets another home run.

A: A walk's as good as a single. I'll take it.

B: You hate to see all that power go to waste, though.

A: Yeah. Maybe Bonds could work with Belt. Give Belt some pointers.

B: Looks like Posey's all recovered from his concussion.



# Experiments

- Baseline systems
  - DBpedia Spotlight (Mendes et al.,2011)
  - WAT (Piccinno et al.,2014): MD (SVM) + ED (PageRank) + Pruning (SVM) trained on Wikipedia articles
  - REL (Hulst et al.,2020): NER (Flair) + ED (MLP) trained on AIDA
  - Flair (Akbik et al.,2019) + BLINK (Li et al., 2020): BLINK is a BERT based ED system

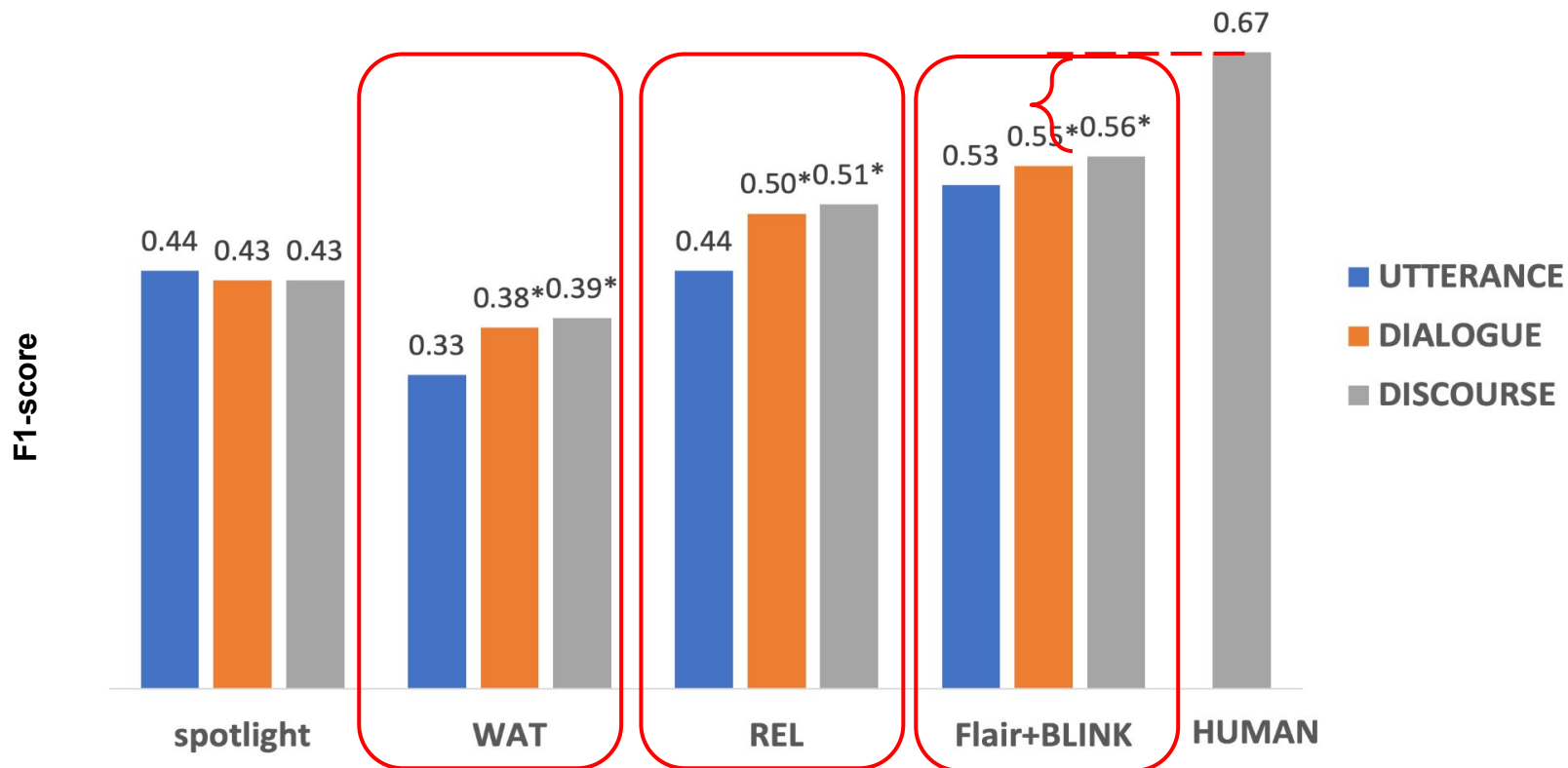


Figure 4: NEL evaluation result (strict match). \*indicates statistically significant compared to UTTERANCE setting with  $p < 0.05$

UTTERANCE	B: Looks like <i>Posey's</i> all recovered from his concussion.	(Posey, Posey County Q6307475)
DIALOGUE	<p>A: Now if Belt will just start hitting . . .</p> <p>B: Dude, if he can keep getting on base by taking walks, I don't care if he never gets another home run.</p> <p>A: A walk's as good as a single. I'll take it.</p> <p>B: You hate to see all that power go to waste, though.</p> <p>A: Yeah. Maybe Bonds could work with him. Give him some pointers.</p> <p>B: Looks like <i>Posey's</i> all recovered from his concussion.</p>	(Posey, James Posey, Q717793)
DISCOURSE	<p>A: Now if Belt will just start hitting . . .</p> <p>B: Dude, if <b>Belt</b> can keep getting on base by taking walks, I don't care if <b>Belt</b> never gets another home run.</p> <p>A: A walk's as good as a single. I'll take it.</p> <p>B: You hate to see all that power go to waste, though.</p> <p>A: Yeah. Maybe Bonds could work with <b>Belt</b> . Give <b>Belt</b> some pointers.</p> <p>B: Looks like <i>Posey's</i> all recovered from his concussion.</p>	(Posey, Buster Posey, Q971912) ✓

Figure 5: Model outputs

# Outline

- Motivations
- Challenges
- The OpenEL Corpus
- Experiments
- Conclusion & Future Work

# Conclusion & Future Work

- Presented OpenEL corpus with very high IAA
- Demonstrated the gap between baselines and human performance
- Future work
  - Full discourse representation
  - Off-the-shelf coreference



# Thank you!

For code and data:

[https://github.com/wenzi3241/OpenEL\\_corpus](https://github.com/wenzi3241/OpenEL_corpus)