

# The Universal Anaphora Scorer

Juntao Yu<sup>1</sup> , Sopan Khosla<sup>2,3</sup> , Nafise Sadat Moosavi<sup>4</sup> , Silviu Paun<sup>5</sup>

Sameer Pradhan<sup>6,7</sup> and Massimo Poesio<sup>5</sup>

<sup>1</sup>Univ. of Essex, UK; <sup>2</sup>Carnegie Mellon Univ., USA; <sup>3</sup>AWS AI, Amazon, USA;

<sup>4</sup>Univ. of Sheffield, UK; <sup>5</sup>Queen Mary Univ., UK; <sup>6</sup>LDC, Univ. of Pennsylvania, USA; <sup>7</sup> [cemantix.org](http://cemantix.org)

LREC 2022

# The Universal Anaphora Initiative

- Beyond Identity Anaphora:

- **Split-antecedent anaphora**

- E.g. [John]<sub>1</sub> met [Mary]<sub>2</sub>. [He]<sub>1</sub> greeted [her]<sub>2</sub>. [They]<sub>1,2</sub> went to the movies.

- **Discourse deixis**

- E.g. ... when suddenly a White Rabbit with pink eyes ran close by her. There was nothing so VERY remarkable in **[that]**; nor did Alice think it so VERY much out of the way to hear the Rabbit say to itself, 'Oh dear! Oh dear! I shall be late!' (when she thought it over afterwards, it occurred to her that she ought to have wondered at **[this]**, but at the time it all seemed quite natural); ....

- **Bridging references**

- E.g. she found herself in [a long, low hall, which was lit up by a row of lamps hanging from [the roof]].

# The Universal Anaphora Scorer

An extension to the Reference Coreference Scorer (Pradhan et al, 2014) scoring the interpretation of

- Identity Reference
- Singletons
- Split-antecedent anaphora
- Non-referring expression
- Discourse Deixis
- Bridging References

# The input: CONLL-UA Exploded Format

- an extension of the **CONLL-U** tabular format defined for Universal Dependencies broadly compatible with the **standard CONLL format** used by the Coreference Reference Scorer
- **Identity column:**
  - specifying the entity a markable refers to in the case of a referring markable and, optionally, whether the markable is referring or not, what its head is, and, for split antecedents, the set they belong to;
- **Bridging column:**
  - specifying the anchor, its most recent mention, and, optionally, the associative relation;
- **Discourse Deixis column:**
  - whose markables specify the non-nominal antecedents of discourse deixis, represented exactly as in the Identity layer. This makes it possible to adopt for discourse deixis the same metrics used for identity anaphora.

# The input: CONLL-UA Exploded Format

```
# global.columns = ID FORM LEMMA UPOS XPOS FEATS HEAD DEPREL DEPS MISC IDENTITY BRIDGING DISCOURSE_DEIXIS REFERENCE NOM_SEM
# newdoc id = Trains_93/D93_9_1
# turn_id = D93_9_1-t1
# speaker = s
# sent_id = D93_9_1-1
# text = s : hello can I help you
1 s - - - - -
2 : - - - - -
3 hello - - - - -
4 can - - - - -
5 I - - - - -
6 help - - - - -
7 you - - - - -

# turn_id = D93_9_1-t2
# speaker = u
# sent_id = D93_9_1-2
# text = u : okay um
8 u - - - - - (EntityID=1-DD|MarkableID=dd_markable_535|Min=8,23|SemType=dn -
9 : - - - - - - - - -
10 okay - - - - - - - - -
11 um - - - - - - - - -

# sent_id = D93_9_1-3
# text = I want to know how long alright how long does it take
12 I - - - - - - - - -
13 want - - - - - - - - -
14 to - - - - - - - - -
15 know - - - - - - - - -
16 how - - - - - (EntityID=1-Pseudo|MarkableID=markable_469|Min=16,17|SemType=quantifier - - - (MarkableID=markable_469|Entity_Type=unknown|Genericity=no-generic
17 long - - - - - ) - - - )
18 alright - - - - - - - - -
19 how - - - - - (EntityID=2-Pseudo|MarkableID=markable_470|Min=19,20|SemType=quantifier - - - (MarkableID=markable_470|Entity_Type=unknown|Genericity=no-generic
20 long - - - - - ) - - - )
21 does - - - - - - - - -
22 it - - - - - (EntityID=3-Pseudo|MarkableID=markable_6|Min=22|SemType=expletive) - - - (MarkableID=markable_6|Entity_Type=unknown|Genericity=no-generic)
23 take - - - - - - - - - ) - -
```

# Identity Anaphora

## Identity Column

(EntityID=10|MarkableID=markable\_11|Min=5|SemType=do|ElementOf=23)

- Evaluating coreference relations only (e.g. in CoNLL 2012)
- Evaluating coreference relations and singletons (e.g. in CRAC 2018)
- Evaluating coreference relations (include split-antecedents) and singletons (e.g. in CODI-CRAC 2021)

Supported metrics: MUC, B<sup>3</sup>, CEAF, BLANC, LEA

# Split-antecedent anaphora

## Identity Column

- Implements a new method proposed by Paun et al. (2022).
- Treating the antecedents of split-antecedent anaphors as a new type of mention: **accommodated sets**.
- E.g. [John]<sub>1</sub> met [Mary]<sub>2</sub>. [He]<sub>1</sub> greeted [her]<sub>2</sub>. [They]<sub>1,2</sub> went to the movies.
  - [He]<sub>1</sub> ∈ Coref Chain 1 (John) = { [John], [He] }
  - [her]<sub>2</sub> ∈ Coref Chain 2 (Mary) = { [Mary], [her] }
  - [They]<sub>1,2</sub> ∈ Coref Chain 3 ({John, Mary}) = { {1,2}, [they] }

# Split-antecedent anaphora

## Identity Column

- Evaluation Steps
  - Identifies all accommodated sets in the key and response.
  - The relevant F1 scores are then calculated for pairs (key-response) of accommodated sets to create a similarity matrix between all accommodated sets in the key and response.
  - The KuhnMunkres algorithm (Kuhn, 1955; Munkres, 1957) is used to search for the best alignments.
  - The standard metrics are adjusted to allow partially matched mentions (i.e. the accommodated sets).
- The procedures for the standard mentions are unchanged, while for computation associated with accommodated sets, **partial credit** is awarded on how well the accommodated sets are resolved.
- Evaluating only split-antecedents
  - The micro-average F1 of all the split-antecedents in the key and response.



# Non-referring expressions

Identity Column

(EntityID=4-**Pseudo**|MarkableID=markable\_6|Min=17|SemType=predicate)

- E.g. [It] was late at night
- Follow CRAC 2018, non-referring expressions were scored separately.
- An F1 score is computed between the collection of non-referring expressions in the key and the response.

# Discourse Deixis

## Discourse Deixis Column

- Discourse deixis is similar to coreference
  - Both form clusters by linking the anaphors to their antecedents.
  - Both have split-antecedent anaphors that refer to multiple antecedents
- The main difference:
  - In coreference, antecedents are introduced using nominal phrases.
  - In discourse deixis they are introduced using non-nominal phrases (segments).

# Discourse Deixis

## Discourse Deixis Column

(EntityID=1-DD|MarkableID=dd\_markable\_2|Min=19,32|SemType=dn|ElementOf=6-DD)

- Discourse deixis is evaluated in the same way as entity anaphora;
  - Discourse deixis evaluation now works with predicted mentions/segments
  - By adopting the generalization of the standard identity reference metrics to split antecedents, we can use the scorer for the very common case of discourse deixis with more than one segment antecedent.

# Bridging References

Bridging Column

(MarkableID=markable\_9|Rel=subset-inv|MentionAnchor=markable\_1|EntityAnchor=3)

- Mention-based F1
- Entity-based F1
- Anaphora recognition F1

# The CODI/CRAC 2021 Shared Task

- The new UA scorer was used as the official scorer for the CODI-CRAC 2021 shared task.
- Identity (Task 1)
  - coreference relations (including split-antecedents) and singletons
- Bridging (Task 2)
- Discourse Deixis (Task 3)
  
- We run the CODI/CRAC Shared Task for the second year with new datasets and additional gold anaphora phrases.
- <https://codalab.lisn.upsaclay.fr/competitions/614>

# Future Work

- Extending the scorer to cover discontinuous mentions—mentions broken into separate segments of text
- To cover cases of disagreement on anaphoric interpretation.

# Code

- The code is available from our GitHub pages:
- <https://github.com/juntaoy/universal-anaphora-scorer>