

Universal Anaphora: The First Three Years

**Massimo Poesio, Maciej Ogrodniczuk, Vincent Ng,
Sameer Pradhan, Juntao Yu, Nafise Sadat Moosavi,
Silviu Paun, Amir Zeldes, Anna Nedoluzhko,
Michal Novák, Martin Popel, Zdeněk Žabokrtský,
and Dan Zeman**

`http://universalanaphora.org
https://arciduca.org/`

COLING, May 2024

Motivations

- ▶ The **Universal Anaphora** initiative (UA) was launched in 2020 to coordinate efforts to push forward the state of the art in anaphora research.
- ▶ The initiative, modelled on Universal Dependencies, aims to
 - ▶ expand the aspects of anaphoric interpretation which are or can be reliably annotated in anaphoric corpora,
 - ▶ produce unified standards to annotate and encode these annotations,
 - ▶ deliver datasets encoded according to these standards,
 - ▶ and develop methods for evaluating this type of interpretation.
- ▶ In parallel, the COREFUD project was also launched, with the related aim of developing standards for adding anaphoric information to corpora annotated according to Universal Dependencies.

This paper

- ▶ First published paper on the objectives of Universal Anaphora and the proposed coverage
- ▶ Summary of the progress since 2020, including
 - ▶ Markup formats: CONLL-UA, COREFUD
 - ▶ Datasets: CODI/CRAC 2021 and 2022, COREFUD repository
 - ▶ Universal Anaphora Scorer 1.0 and 2.0
 - ▶ Shared tasks: CODI/CRAC 2021 and 2022, CRAC 2023

A Fuller Picture of Anaphoric Reference

- ▶ One of the objectives of Universal Anaphora is to encourage the move towards looking at the full range of anaphoric phenomena
- ▶ To this end, our markup standards / scoring methods were designed to support the fuller picture of anaphoric reference annotated in modern corpora:
 - ▶ A richer notion of MARKABLE
 - ▶ A wider RANGE OF ANAPHORIC RELATIONS

Beyond continuous nominal markables

▶ DISCONTINUOUS MARKABLES

- (1) u1 M [a tanker
 u2 S yeah
 u3 M of orange juice]

▶ ZERO ANAPHORA

- (2) [IT] [Giovanni]_i è in ritardo, così [∅]_i mi ha chiesto se
 posso incontrarlo al cinema.
 [EN] [John]_i is late so [he]_i asked me if I can meet him
 at the movies.

Beyond basic coreference

▶ SPLIT-ANTECEDENT ANAPHORA

(3) [John]₁ met [Mary]₂. [He]₁ greeted [her]₂. [They]_{1,2} went to the movies.

▶ DISCOURSE DEIXIS

(4) ... 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);

Markup

- ▶ Two mutually interchangeable markup formats:
 - ▶ CONLL-UA: based on the CONLL-U-Plus tabular format proposed in Universal Dependencies for corpora containing additional linguistic annotations. It introduces new layers specifically for reference and nominal semantics
 - ▶ COREFUD: a more 'compact' format that can be used to pack the anaphoric information representable in CONLL-UA in the 'MISC' column of the CONLL-U format, and is fully compatible with the Universal Dependencies (= the current UD parser can process it)

CONLL-UA

The full UA format specifies four additional CONLL-U layers:

- ▶ **Identity** (required), for specifying the markables for anaphoric reference, and the entity a markable refers to in the case of a referring markable (as in the CONLL coreference scheme). In addition to coreference information, this layer may contain additional optional attributes specifying whether the markable is referring or not (attribute `SemType`), and what its head is (attribute `Min`).
- ▶ **Bridging** (optional), specifying the anchor, its most recent mention, and, optionally, the associative relation.
- ▶ **Discourse_Deixis** (optional), 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.
- ▶ **Nom_Sem** (optional), for information about nominal semantics not already included in the CONLL-U layers - ontological category, genericity, etc.

Datasets

- ▶ UNIVERSAL ANAPHORA repository
<https://github.com/UniversalAnaphora>:
 - ▶ Existing datasets converted to CONLL-UA: e.g., *Phrase Detectives*
 - ▶ New datasets created as part of the Universal Anaphora effort: the CODI/CRAC 2022 corpus, used for the CODI/CRAC 2022 Shared Task on Anaphoric Reference in Dialogue
- ▶ COREFUD repository
<https://ufal.mff.cuni.cz/corefud>:
 - ▶ a collection of 21 datasets in 13 languages, many of which stored in full UD format (i.e., also including morpho syntactic information)

The CorefUD Repository

Dataset	Doc	Tok	Ent	Mark	Sgl	App	Pred	Split	DD	Bdg
Catalan-AnCora	1,298	429	18	62	0	✓	✓	✓	✓	X
Czech-PCEDT	2,312	1,156	49	168	3	(✓)	(✓)	✓	✓	✓
Czech-PDT	3,165	835	47	155	32	(✓)	(✓)	✓	✓	✓
English-GUM	195	187	7	32	20	✓	✓	✓	✓	✓
English-ParCorFull	19	11	0	1	0	✓	(✓)	✓	✓	X
French-Democrat	126	285	7	46	32	X	X	X	X	X
German-ParCorFull	19	11	0	1	0	✓	(✓)	✓	✓	X
German-PotsdamCC	176	33	1	3	3	✓	✓	X	✓	X
Hungarian-KorKor	94	25	1	4	0	?	?	X	?	X
Hungarian-SzegedKoref	400	124	5	15	0	✓	?	X	✓	✓
Lithuanian-LCC	100	37	1	4	0	X	X	✓	X	X
Norwegian-BokmaalNARC	346	246	6	27	48	?	?	✓	?	✓
Norwegian-NynorskNARC	394	207	5	22	40	?	?	✓	?	✓
Polish-PCC	1,828	539	22	83	106	✓	✓	X	✓	✓
Russian-RuCor	181	157	4	16	0	✓	✓	X	X	X
Spanish-AnCora	1,356	458	19	71	1	✓	✓	✓	✓	X
Turkish-ITCC	24	55	1	4	0	?	?	X	?	X
Dutch-COREA	844	140	3	9	25	✓	✓	X	✓	✓
English-ARRAU (ARRAU 1)	413	229	8	32	40	✓	✓	✓	✓	✓
English-OntoNotes	3,493	1,632	51	209	0	✓	X	X	✓	X
English-PCEDT	2,312	1,174	39	139	15	(✓)	(✓)	✓	✓	X

The Universal Anaphora Scorer(s)

- ▶ A Python scorer covering the varieties of markables and anaphoric reference in the scope of the Universal Anaphora catalogue
 - ▶ builds on the original Reference Coreference scorer (Pradhan et al, 2014)
 - ▶ and its reimplement in Python by Moosavi for the CRAC 2018 shared task (Poesio et al 2018), which also added evaluation for identification of non-referring expressions, bridging reference, and discourse deixis.
- ▶ UA Scorer 1.0: added evaluation of split antecedent plurals and new evaluation for discourse deixis (CODI/CRAC 2022)
- ▶ COREFUD scorer: added support for discontinuous markables (CRAC 2022)
- ▶ UA Scorer 2.0: merged UA 1.0 and COREFUD 1.0, added evaluation of zero anaphora resolution

Scoring Split Antecedent Anaphora

- ▶ The UA scorer implements the new method for scoring split-antecedent anaphora proposed by (Paun et al, 2023)
- ▶ Key idea: the antecedents of split-antecedent anaphors are a new type of mention, **accommodated sets**—set denoting entities which have the split antecedents as elements.
- ▶ E.g., in (3), split-antecedent anaphor [*They*]_{1,2} is encoded as belonging to a coreference chain whose first element is the accommodated set {1,2} with the coreference chains for *John* and *Mary* as elements.

(3) [John]₁ met [Mary]₂. [He]₁ greeted [her]₂. [They]_{1,2} went to the movies.

Scoring Discourse Deixis Resolution

- ▶ The UA scorer implements an entirely new approach to evaluation of discourse deixis.
- ▶ Key idea: discourse deixis is similar to coreference, in that both form clusters by linking the anaphors to their antecedents. Also, in both cases we can have split-antecedent anaphors that refer to multiple antecedents—in fact, split antecedent reference is the norm for discourse deixis.
- ▶ In CONLL-UA, discourse deixis is specified in the separate `Discourse_Deixis` layer, but using the exact same attributes as the `Identity` column of the ‘exploded’ format.
- ▶ This representation enables using standard coreference metrics to evaluate discourse deixis— and given that our new scorer provides a way to incorporate split-antecedents into the standard metrics, split antecedent discourse deixis can be handled as well.
- ▶ This is exactly how the UA scorer evaluates discourse deixis: it computes the same MUC, B³, CEAF, CONLL, BLANC and LEA metrics as for identity anaphora.

Scoring Zero Anaphora Resolution

- ▶ In both CONLL-UA and COREFUD, zeros are represented using the UD standard of empty nodes, in which the first column (ID, word index) is indicated using the decimal numbers.
 - ▶ For instance, if we have a zero anaphora right after a token whose ID is 5, we index the zero with 5.1 instead of 6 used for a normal token.
- ▶ The scorer identifies the zeros by the decimal indexing and has the option to include zeros in the evaluation.
- ▶ When zeros are included in the evaluation, again we need to align them between the key and response.
- ▶ Currently, alignment is based on the position of the zeros– i.e. zeros are aligned if they are located in the same position in the sentences.

Shared Tasks

- ▶ The CODI/CRAC 2021 and 2022 Shared Tasks on Anaphora Resolution in Dialogue
- ▶ The CRAC 2022 and 2023 Shared Tasks on Multilingual Coreference Resolution

Conclusions

- ▶ Phase 1 of the Universal Anaphora initiative has achieved most of its initial goals, including the development of markup formats suitable to encode the anaphoric phenomena in its coverage in multiple languages, and of a scorer that can be used to evaluate models carrying out more complex forms of anaphoric interpretation.
- ▶ Phase 2 is now beginning, with no less ambitious objectives.
 - ▶ Further develop the markup in order to cover more aspects of anaphoric interpretation, such as ambiguity and quasi-coreference;
 - ▶ further specify the methods for marking in deictic reference in visual contexts and several as yet poorly understood aspects of anaphoric reference in dialogue.
 - ▶ Start attempting developing common guidelines, as done in UD, ideally in collaboration with the linguistic community.
 - ▶ Finally, and as importantly, we hope to expand our community to include more researchers, from the computational and the linguistic fields.

Thank You