Towards Adding Arabic to CorefUD

Dima Taji and Daniel Zeman
Charles University
Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics (ÚFAL)
Prague, Czechia
{taji, zeman}@ufal.mff.cuni.cz

Abstract

Training models that can perform well on various NLP tasks requires large amounts of data, which becomes even more apparent with more nuanced tasks such as anaphora and coreference resolution. This paper presents the automatic creation of an Arabic CorefUD dataset through the automatic conversion of the existing gold-annotated OntoNotes.

1 Introduction

Coreference resolution is the linguistic task of clustering the different noun phrases that refer to the same entity within a text (Nedoluzhko et al., 2022; Zheng et al., 2011; Elango, 2005). For example, in the sentence "*As Gregor Samsa awoke one morning from uneasy dreams he found himself transformed in his bed into a gigantic insect.*", the mentions *Gregor Samsa*, *he*, *himself*, and *his* all refer to the same entity in the real world.

Recognizing these mentions and clustering them has been shown to improve the performance of different NLP tasks, particularly the tasks that require constructing the meaning of the text, such as opinion target identification (Jakob and Gurevych, 2010), machine translation (Luong and Popescu-Belis, 2016; Miculicich Werlen and Popescu-Belis, 2017), and machine reading comprehension (Huang et al., 2022). Research has also shown that the applications of coreference resolution can extend to tasks in other fields, such as building an ontology in the biomedical domain (Ashury Tahan et al., 2024), improving diversity in rankings (Zhu et al., 2007), and weather forecasting (Belz, 2007).

However, datasets, even for the same language, vary greatly in their format, the phenomena covered, and the way they are annotated. As a result, different state-of-the-art systems are evaluated on different datasets, and as such, evaluations are often not directly comparable (Kobayashi and Ng, 2020). Moreover, using data from different languages to train multilingual systems cannot be achieved without extensive preprocessing of the data to harmonize it.

Multiple efforts have been made to create multilingual coreference-annotated corpora and harmonize existing corpora to follow a unified scheme. Of all these efforts, which will be discussed in Section 2, we are interested in adding an Arabic dataset to the CorefUD (Nedoluzhko et al., 2021, 2022) corpus, following the approach that has been designed to convert the English OntoNotes (Weischedel et al., 2013) dataset to CorefUD.

Due to the morphologically-rich nature of Arabic, we had to modify the existing conversion approach to account for the presence of zero mentions, in addition to converting the annotated data following the same approach used for the conversion of the English data. These decisions are presented in Section 3.

Finally, Section 4 outlines our future plans and concludes.

2 Literature Review

In this section, we present previous efforts that have been made pertaining to the work we are describing in this paper.

2.1 Multilingual Coreference and Anaphora Corpora

Multilingual corpora annotated with coreference and anaphora information are an intuitive solution for creating multilingual and language-agnostic systems. Some of these corpora are limited to languages that are of the same family, such as AnCora (Recasens and Martí, 2010) for Spanish and Catalan, PAWS (Nedoluzhko et al., 2018) for Czech, English, Polish, and Russian, ParCor (Guillou et al., 2014) and ParCorFull (Lapshinova-Koltunski et al., 2022) for English and German, PCEDT (Nedoluzhko et al., 2016) for Czech and English,

Proceedings of the Eighth Workshop on Computational Models of Reference, Anaphora and Coreference (CRAC 2025), pages 70–76
Suzhou, China, November 9, 2025.

Wino-X (Emelin and Sennrich, 2021) for English, German, French, and Russian. Others include distant families such as TransMuCoRes (Mishra et al., 2024) for English and 31 South Asian languages, MMC (Zheng et al., 2023) for English, Chinese, and Farsi, and OntoNotes (Weischedel et al., 2011) for English, Chinese, and Arabic.

The issue with this approach is that every corpus follows its own scheme and has a unique set of definitions and annotation guidelines, making the process of adding new languages a costly endeavor in terms of time, effort, and monetary cost. On the other hand, there is a potential for combining these corpora if their schemes and annotations were harmonized, as discussed next.

2.2 Harmonized Schemes

The first effort to a corpus that harmonizes schemes began with the SemEval 2010 shared task on coreference resolution in multiple languages (Recasens et al., 2009), which extracted coreferences from a number of datasets with varying schemes, and represented them in a CoNLL-like format. The MMAX tool (Müller and Strube, 2022) introduced an XML format that can be used to annotate anaphora as well as other linguistic phenomena, and was used for multiple corpora. However, the annotation approaches and the annotated attributes varied greatly between projects. Universal Anaphora (Poesio et al., 2024) proposes a markup scheme for encoding anaphoric information to facilitate the creation of a collection of corpora using the same scheme. Similarly, CorefUD (Nedoluzhko et al., 2021, 2022) addresses the challenges posed by varying data formats and annotation guidelines in existing coreference corpora by creating a unified scheme and format for coreference annotation, facilitating cross-lingual research and development in anaphora and coreference resolution.

Nevertheless, these efforts focus more on unifying the underlying file format, while there is no work being done on the harmonization of linguistic content.

2.3 Arabic Coreference Corpora

Although, as far as we are aware, OntoNotes 5.0 (Weischedel et al., 2013) is the only current multilingual corpus that includes Arabic, there are several coreference corpora for Arabic alone.

Abolohom and Omar (2015) and Abolohom and Omar (2017) use the Quranic corpus, annotated with antecedent references of pronouns. However, since the linguistic structure of Quranic Arabic (QA) is quite distinct from Modern Standard Arabic (MSA), the transfer of the models' knowledge from QA to MSA cannot be directly compared to the experiment results presented in both papers, where their models were evaluated on QA.

Others created their own corpora, which have been used for a limited number of models, such as Mezghani et al. (2009) and Abdul-Mageed (2011).

However, the corpus that made the most sense to be our starting point was OntoNotes (Weischedel et al., 2011, 2013). Since the Arabic portion of the corpus has been used in numerous efforts (Pradhan et al., 2012; Li, 2012; Pradhan et al., 2013; Aloraini et al., 2020; Min, 2021; Aloraini et al., 2022), that indicates that (1) the corpus is popular enough so the momentum of using it to create and test models could be transferred to our new format, and (2) evaluating new systems created with our converted corpus against existing systems would be easy. On the other hand, the downside is that OntoNotes cannot be redistributed freely, which unfortunately affects accessibility of derived works.

2.4 CorefUD

Inspired by the progress achieved by standardizing the labels and annotation guidelines of morphosyntactic labels brought on by Universal Dependencies (Nivre et al., 2020), Nedoluzhko et al. (2022) introduced CorefUD, a collection of corpora with a harmonized scheme that would unify and standardize the annotation of anaphoric and coreference relations.

CorefUD has proved to be beneficial, especially for languages with small training data sets (Pražák et al., 2021; Chai and Strube, 2023), and has been used in four shared tasks focusing on systems for multilingiual coreference resolutions (Žabokrtský et al., 2022, 2023; Novák et al., 2024, 2025). Additionally, the CorefUD format is being used to produce new corpora (Dyer et al., 2024; Jørgensen and Kåsen, 2024). All of these efforts indicate that following this format has the potential to further propel research in the area of anaphora and coreference resolution.

3 Data and Conversion

For this experiment, we used the OntoNotes 5 Arabic dataset (Weischedel et al., 2013). The data set comprises 599 articles with approximately 400K tokens. Since the data is entirely from the Penn

Arabic Treebank, it only contains news articles.

3.1 OntoNotes Annotations and Labels

The annotations contained in OntoNotes are organized in layers, namely treebank, proposition, word sense, ontology, coreference, and named entity. For the purpose of our conversion, and following the approach used by Nedoluzhko et al. (2022), we require the treebank and coreference layers only.

The treebank layer consists of the syntactic annotations of the sentences. These annotations are the parses that are provided by the LDC for the data in the PATB part 3 – v3.1 (Maamouri et al., 2004). This layer is relevant for us because it contains the zero nodes that are discussed in Section 3.2.

For our current purposes, this layer contains the most relevant information. The annotations in this layer connect names, nominal references, and pronouns that refer to the same entity, marking them as coreferents. Similarly, verbs and their equivalent noun phrases are also marked as coreferents. These annotations can span multiple sentences as long as they occur in the same document. Appositions are also marked in this layer. In the Arabic OntoNotes dataset, only 447 articles are annotated for coreference, making a total of 319K annotated tokens.

The OntoNotes tags in the *Coreference* part are the ones that currently appear in our converted files. They include two types; *IDENT* and *APPOS*, and two subtypes; *HEAD* and *ATTRIB*.

- *IDENT* denoting any nominal mentions of the same entity. This is reflected in our dataset by giving the entities the same IDs, without any further elaboration on tags.
- *APPOS* denoting the initial nominal phrase when combined with the *HEAD*, or the referent when combined with the *ATTRIB*.

Figure 1 shows an excerpt from OntoNotes that illustrates the use of these labels.

3.2 Zeros

In pro-drop languages such as Arabic, subject pronouns can be omitted; these omitted subjects are called zero pronouns (Aloraini et al., 2024). However, even when these pronouns are dropped, they can still be part of a coreference chain. As such, in order to identify all the coreference occurrences in a text, zeros must be identified and inserted in their appropriate locations. Additionally, not all zeros need to be part of a coreference chain, and making this distinction is another task that a coreference resolution model needs to learn.

```
<COREF ID="64" TYPE="IDENT">
    <COREF ID="66" TYPE="APPOS"
    SUBTYPE="ATTRIB">
وزير الدفاع wzyr AldfAς 'Minister of Defense'
    </COREF>
    <COREF ID="66" TYPE="APPOS"
    SUBTYPE="HEAD">
    أنجلو رييس Ânjlw ryys 'Angelo Reyes'
    </COREF>
</COREF>
```

Figure 1: An example of OntoNotes' Coreference annotation showing the tags that are used to identify the nominal mentions. There are two mentions of the same entity connected with an `APPOS` relation and labeled as its `HEAD` and `ATTRIB`, respectively. Additionally, the whole apposition is labeled as a mention in an `IDENT`(ity) coreference relation, whereas `ID="64"` links it to other mentions of that entity elsewhere in the document (not shown here). Arabic transliteration follows the Habash-Soudi-Buckwalter scheme (Habash et al., 2007).

As the existing conversion approach is based on English, which does not contain zeros, the generated output does not cover this linguistic phenomenon. Fortunately, the PATB annotations included within OntoNotes contain zero nodes, and the subsequent coreference annotations in Arabic OntoNotes take the zeros into consideration.

Per the PATB annotation guidelines (Maamouri et al., 2009), there are five types of zero nodes, four of which appear in the data included in OntoNotes:

- *ICH tag* denoting discontinued constituents, when something interrupts the sentence, without affecting its syntax.
- *T tag* for subjects preceding the verb.
- **0* tag* indicating the existence of a null complementizer or zero WH-pronoun.
- * *tag* indicating the object of a passive verb, the subject of a nominal verb, or an omitted subject of a verb.
- **?* tag* denoting ellipses, which do not appear in the OntoNotes data.

Of the mentioned types of zero node tags that appear in our corpus, the nodes with the * tag are the only ones that represent a coreference relation. The other tags indicate relations that can be realized using Deep UD (Droganova and Zeman, 2019).

ID	Token	Coreference Annotations
3	وزير *wzyr* 'Minister'	Entity=(0001@ann@nw@ar@on_d1sec0c20–1(0001@ann@nw@ar@on_d1sec0c2–1-ATTRIB
4	الدفاع *AldfAς* 'Defense'	Entity=0001@ann@nw@ar@on_d1sec0c2)
5	أنجلو *Ânjlw* 'Angelo'	Entity=(0001@ann@nw@ar@on_d1sec0c2–1-appos:1,HEAD
6	ريييس *ryys* 'Reyes'	Entity=0001@ann@nw@ar@on_d1sec0c2)0001@ann@nw@ar@on_d1sec0c20)

Table 1: A segment of the generated CorefUD annotations corresponding to the example shown in Figure 1.

3.3 Output Annotation

Table 1 shows the coreference labels generated by our conversion process that correspond to the example shown in Figure 1. We can see that the tokens belonging to the phrases وزير الدفاع *wzyr AldfAς* 'Minister of Defense' and أنجلو ريييس *Ânjlw ryys* 'Angelo Reyes' are annotated with the same entity ID. We can also see that subtypes *ATTRIB* and *HEAD* have been maintained for the heads of each of the phrases.

Table 2 gives a general overview of the size and distribution of clusters and labels in our corpus. The number of unique coreference clusters, i.e. entities with the same Entity identifier, is 12,672, spanning 41,556 tokens. The average cluster size is 3.27, with cluster sizes spanning from 1 to 80 tokens per cluster.

We retained 92% of the zero nodes that appear in the original OntoNotes annotations during our conversion. As previously mention, these are the nodes of the type * which indicate the object of a passive verb, the subject of a nominal verb, or an omitted subject of a verb. The 8% of the zero nodes contained information that we can utilize to provide additional syntactic annotations for the Deep UD treebank (Droganova and Zeman, 2019).

It is worth noting that, according to the OntoNotes Release 5.0 document (Weischedel et al., 2013), the annotated coreferences were limited to only the intra-document occurrences. Nominal mentions, which would be marked with the *IDENT* label excluded all occurrences where the connection between entities can be derived from the use of copula or similar verbs. This is reflected in the rare appearance of this label in the Arabic OntoNotes corpus.

While we cannot directly redistribute the OntoNotes data, our code needed to reproduce the converted output from one's own copy of OntoNotes files will be publicly available.[1]

[1]The conversion code and documentation can be found under our GitHub repository https://tinyurl.com/arabic-corefud

Documents	447
Sentences	30,601
Tokens excluding zeros	299,362
Tokens including zeros	336,735
Tokens including retained zeros	309,631
Tokens part of a coreference cluster	41,556
Coreference clusters	12,672
Minimum cluster size	1
Maximum cluster size	80
Average cluster size	3.27
APPOS labels	1,789
IDENT labels	5
HEAD labels	1,749
ATTRIB labels	1,790

Table 2: Statistics of the converted corpus.

4 Conclusion and Future Work

In this paper, we presented our effort to add Arabic to the CorefUD collection of corpora. We described the decisions we made to modify the existing conversion process to accommodate phenomena that were not in the English corpus, namely the appearance of zero nodes.

Moving forward, we would like to test the quality of multilingual coreference resolution systems when trained on the entirety of CorefUD, including Arabic. Additionally, we plan to prepare a publicly available CorefUD dataset based on UD_Arabic-PADT (Taji et al., 2017). We believe this will be beneficial to furthering research in this area.

Acknowledgments

This work has been supported by the Charles University, project GA UK No. 190125, LINDAT/CLARIAHCZ (Project No. LM2023062) of the Ministry of Education, Youth, and Sports of the Czech Republic, and was partially supported by SVV project number 260 821.

References

Muhammad Abdul-Mageed. 2011. Automatic detection of Arabic non-anaphoric pronouns for improving anaphora resolution. *ACM Transactions on Asian Language Information Processing (TALIP)*, 10(1):1–11.

Abdullatif Abolohom and Nazlia Omar. 2015. A hybrid approach to pronominal anaphora resolution in Arabic. *Journal of Computer Science*, 11(5):764.

Abdullatif Abolohom and Nazlia Omar. 2017. A computational model for resolving Arabic anaphora using linguistic criteria. *Indian Journal of Science and Technology*, 10(3):1–6.

Abdulrahman Aloraini, Sameer Pradhan, and Massimo Poesio. 2022. Joint coreference resolution for zeros and non-zeros in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 11–21, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Abdulrahman Aloraini, Juntao Yu, Wateen Aliady, and Massimo Poesio. 2024. A survey of coreference and zeros resolution for Arabic. *ACM Transactions on Asian and Low-Resource Language Information Processing*.

Abdulrahman Aloraini, Juntao Yu, and Massimo Poesio. 2020. Neural coreference resolution for Arabic. In *Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 99–110, Barcelona, Spain (online). Association for Computational Linguistics.

Shir Ashury Tahan, Amir David Nissan Cohen, Nadav Cohen, Yoram Louzoun, and Yoav Goldberg. 2024. Data-driven coreference-based ontology building. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 14290–14300, Miami, Florida, USA. Association for Computational Linguistics.

Anja Belz. 2007. Probabilistic generation of weather forecast texts. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 164–171, Rochester, New York. Association for Computational Linguistics.

Haixia Chai and Michael Strube. 2023. Investigating multilingual coreference resolution by universal annotations. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 10010–10024, Singapore. Association for Computational Linguistics.

Kira Droganova and Daniel Zeman. 2019. Towards deep Universal Dependencies. In *Proceedings of the Fifth International Conference on Dependency Linguistics (Depling, SyntaxFest 2019)*, pages 144–152, Paris, France. Association for Computational Linguistics.

Andrew Dyer, Ruveyda Betul Bahceci, Maryam Rajestari, Andreas Rouvalis, Aarushi Singhal, Yuliya Stodolinska, Syahidah Asma Umniyati, and Helena Rodrigues Menezes de Oliveira Vaz. 2024. A multilingual parallel corpus for coreference resolution and information status in the literary domain. In *Proceedings of the 22nd Workshop on Treebanks and Linguistic Theories (TLT 2024)*, pages 55–64, Hamburg,Germany. Association for Computational Linguistics.

Pradheep Elango. 2005. Coreference resolution: A survey. *University of Wisconsin, Madison, WI*, 1(12):12.

Denis Emelin and Rico Sennrich. 2021. Wino-X: Multilingual Winograd schemas for commonsense reasoning and coreference resolution. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 8517–8532, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Liane Guillou, Christian Hardmeier, Aaron Smith, Jörg Tiedemann, and Bonnie Webber. 2014. ParCor 1.0: A parallel pronoun-coreference corpus to support statistical MT. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3191–3198, Reykjavik, Iceland. European Language Resources Association (ELRA).

Nizar Habash, Abdelhadi Soudi, and Tim Buckwalter. 2007. On Arabic Transliteration. In A. van den Bosch and A. Soudi, editors, *Arabic Computational Morphology: Knowledge-based and Empirical Methods*. Springer.

Baorong Huang, Zhuosheng Zhang, and Hai Zhao. 2022. Tracing origins: Coreference-aware machine reading comprehension. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1281–1292, Dublin, Ireland. Association for Computational Linguistics.

Niklas Jakob and Iryna Gurevych. 2010. Using anaphora resolution to improve opinion target identification in movie reviews. In *Proceedings of the ACL 2010 Conference Short Papers*, pages 263–268, Uppsala, Sweden. Association for Computational Linguistics.

Tollef Emil Jørgensen and Andre Kåsen. 2024. Aligning the Norwegian UD treebank with entity and coreference information. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 704–710, Torino, Italia. ELRA and ICCL.

Hideo Kobayashi and Vincent Ng. 2020. Bridging resolution: A survey of the state of the art. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 3708–3721, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Ekaterina Lapshinova-Koltunski, Pedro Augusto Ferreira, Elina Lartaud, and Christian Hardmeier. 2022. ParCorFull2.0: a parallel corpus annotated with full coreference. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 805–813, Marseille, France. European Language Resources Association.

Baoli Li. 2012. Learning to model multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 129–135, Jeju Island, Korea. Association for Computational Linguistics.

Ngoc Quang Luong and Andrei Popescu-Belis. 2016. Improving pronoun translation by modeling coreference uncertainty. In *Proceedings of the First Conference on Machine Translation: Volume 1, Research Papers*, pages 12–20, Berlin, Germany. Association for Computational Linguistics.

Mohamed Maamouri, Ann Bies, Tim Buckwalter, and Wigdan Mekki. 2004. The penn arabic treebank: Building a large-scale annotated arabic corpus. In *NEMLAR conference on Arabic language resources and tools*, volume 27, pages 466–467. Cairo.

Mohamed Maamouri, Ann Bies, Sondos Krouna, Fatma Gaddeche, and Basma Bouziri. 2009. Penn Arabic Treebank Guidelines. *Linguistic Data Consortium*.

Souha Mezghani, Lamia Belguith, and Abdelmajid Ben Hamadou. 2009. Arabic anaphora resolution: Corpora annotation with coreferential links. *Int. Arab J. Inf. Technol.*, 6:480–488.

Lesly Miculicich Werlen and Andrei Popescu-Belis. 2017. Using coreference links to improve Spanish-to-English machine translation. In *Proceedings of the 2nd Workshop on Coreference Resolution Beyond OntoNotes (CORBON 2017)*, pages 30–40, Valencia, Spain. Association for Computational Linguistics.

Bonan Min. 2021. Exploring pre-trained transformers and bilingual transfer learning for Arabic coreference resolution. In *Proceedings of the Fourth Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 94–99, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Ritwik Mishra, Pooja Desur, Rajiv Ratn Shah, and Ponnurangam Kumaraguru. 2024. Multilingual coreference resolution in low-resource South Asian languages. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 11813–11826, Torino, Italia. ELRA and ICCL.

Mark-Christoph Müller and Michael Strube. 2022. Annotating anaphoric and bridging relations with mmax. In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue. September 1 - 2, 2001, Aalborg, Denmark*, page 6.

Anna Nedoluzhko, Michal Novák, Silvie Cinková, Marie Mikulová, and Jiří Mírovský. 2016. Coreference in Prague Czech-English Dependency Treebank. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 169–176, Portorož, Slovenia. European Language Resources Association (ELRA).

Anna Nedoluzhko, Michal Novák, and Maciej Ogrodniczuk. 2018. PAWS: A multi-lingual parallel treebank with anaphoric relations. In *Proceedings of the First Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 68–76, New Orleans, Louisiana. Association for Computational Linguistics.

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, Amir Zeldes, and Daniel Zeman. 2022. CorefUD 1.0: Coreference meets Universal Dependencies. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 4859–4872, Marseille, France. European Language Resources Association.

Anna Nedoluzhko, Michal Novák, Martin Popel, Zdeněk Žabokrtský, and Daniel Zeman. 2021. Is one head enough? mention heads in coreference annotations compared with UD-style heads. In *Proceedings of the Sixth International Conference on Dependency Linguistics (Depling, SyntaxFest 2021)*, pages 101–114, Sofia, Bulgaria. Association for Computational Linguistics.

Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher D. Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. Universal Dependencies v2: An evergrowing multilingual treebank collection. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4034–4043, Marseille, France. European Language Resources Association.

Michal Novák, Barbora Dohnalová, Miloslav Konopik, Anna Nedoluzhko, Martin Popel, Ondrej Prazak, Jakub Sido, Milan Straka, Zdeněk Žabokrtský, and Daniel Zeman. 2024. Findings of the third shared task on multilingual coreference resolution. In *Proceedings of the Seventh Workshop on Computational Models of Reference, Anaphora and Coreference*, pages 78–96, Miami. Association for Computational Linguistics.

Michal Novák, Miloslav Konopik, Anna Nedoluzhko, Martin Popel, Ondrej Prazak, Jakub Sido, Milan Straka, Zdeněk Žabokrtský, and Daniel Zeman. 2025. Findings of the fourth shared task on multilingual coreference resolution: Can llms dethrone traditional approaches? In *Proceedings of the Joint Sixth Workshop on Computational Approaches to Discourse (CODI) and Eigth Computational Models of Reference, Anaphora and Coreference (CRAC)*, Suzhou, China.

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 Daniel Zeman. 2024. Universal anaphora: The first three years. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 17087–17100, Torino, Italia. ELRA and ICCL.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Hwee Tou Ng, Anders Björkelund, Olga Uryupina, Yuchen Zhang, and Zhi Zhong. 2013. Towards robust linguistic analysis using OntoNotes. In *Proceedings of the Seventeenth Conference on Computational Natural Language Learning*, pages 143–152, Sofia, Bulgaria. Association for Computational Linguistics.

Sameer Pradhan, Alessandro Moschitti, Nianwen Xue, Olga Uryupina, and Yuchen Zhang. 2012. CoNLL-2012 shared task: Modeling multilingual unrestricted coreference in OntoNotes. In *Joint Conference on EMNLP and CoNLL - Shared Task*, pages 1–40, Jeju Island, Korea. Association for Computational Linguistics.

Ondřej Pražák, Miloslav Konopík, and Jakub Sido. 2021. Multilingual coreference resolution with harmonized annotations. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1119–1123, Held Online. INCOMA Ltd.

Marta Recasens and M Antònia Martí. 2010. AnCora-CO: Coreferentially annotated corpora for Spanish and Catalan. *Language resources and evaluation*, 44:315–345.

Marta Recasens, Antonia Martí, Mariona Delor, Lluís Màrquez, and Emili Sapena. 2009. Semeval-2010 task 1. page 70.

Dima Taji, Nizar Habash, and Daniel Zeman. 2017. Universal Dependencies for Arabic. In *Proceedings of the Third Arabic Natural Language Processing Workshop*, pages 166–176, Valencia, Spain. Association for Computational Linguistics.

Ralph Weischedel, Hovy Eduard, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradhan, Lance Ramshaw, and Nianwen Xue. 2011. OntoNotes: A large training corpus for enhanced processing. *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*, pages 54–63.

Ralph Weischedel, Martha Palmer, Mitchell Marcus, Hovy Eduard, Sameer Pradhan, Lance Ramshaw, Nianwen Xue, Ann Taylor, Jeff Kaufman, Michelle Franchini, Mohammed El-Bachouti, Robert Belvin, and Ann Houston. 2013. OntoNotes Release 5.0 LDC2013T19.

Zdeněk Žabokrtský, Miloslav Konopik, Anna Nedoluzhko, Michal Novák, Maciej Ogrodniczuk, Martin Popel, Ondrej Prazak, Jakub Sido, and Daniel Zeman. 2023. Findings of the second shared task on multilingual coreference resolution. In *Proceedings of the CRAC 2023 Shared Task on Multilingual Coreference Resolution*, pages 1–18, Singapore. Association for Computational Linguistics.

Zdeněk Žabokrtský, Miloslav Konopík, Anna Nedoluzhko, Michal Novák, Maciej Ogrodniczuk, Martin Popel, Ondřej Pražák, Jakub Sido, Daniel Zeman, and Yilun Zhu. 2022. Findings of the shared task on multilingual coreference resolution. In *Proceedings of the CRAC 2022 Shared Task on Multilingual Coreference Resolution*, pages 1–17, Gyeongju, Republic of Korea. Association for Computational Linguistics.

Boyuan Zheng, Patrick Xia, Mahsa Yarmohammadi, and Benjamin Van Durme. 2023. Multilingual coreference resolution in multiparty dialogue. *Transactions of the Association for Computational Linguistics*, 11:922–940.

Jiaping Zheng, Wendy W Chapman, Rebecca S Crowley, and Guergana K Savova. 2011. Coreference resolution: A review of general methodologies and applications in the clinical domain. *Journal of biomedical informatics*, 44(6):1113–1122.

Xiaojin Zhu, Andrew Goldberg, Jurgen Van Gael, and David Andrzejewski. 2007. Improving diversity in ranking using absorbing random walks. In *Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics; Proceedings of the Main Conference*, pages 97–104, Rochester, New York. Association for Computational Linguistics.