

Do UD Trees Match Mention Spans in Coreference Annotations?

Martin Popel, Zdeněk Žabokrtský, Anna Nedoluzhko, Michal Novák and Daniel Zeman
Charles University,

Faculty of Mathematics and Physics, Institute of Formal and Applied Linguistics
{popel, zabokrtsky, nedoluzhko, mnovak, zeman}@ufal.mff.cuni.cz

Abstract

One can find dozens of data resources for various languages in which coreference – a relation between two or more expressions that refer to the same real-world entity – is manually annotated. One could also assume that such expressions usually constitute syntactically meaningful units; however, mention spans have been annotated simply by delimiting token intervals in most coreference projects, i.e., independently of any syntactic representation. We argue that it could be advantageous to make syntactic and coreference annotations convergent in the long term. We present a pilot empirical study focused on matches and mismatches between hand-annotated linear mention spans and automatically parsed syntactic trees that follow Universal Dependencies conventions. The study covers 9 datasets for 8 different languages.

1 Introduction

The relation of coreference holds between expressions in a text which (co-)refer to the same real-world entity or event. The dominating practice in coreference-related language resources is that the individual referring expressions (often called *mentions*) are represented simply as linearly delimited sequences of tokens (often called *mention spans*) within the original sentences, without using syntactic representation of the sentences.

We argue that integrating coreference annotation with syntactic annotation, more specifically with dependency trees, would be beneficial in the long term from various linguistic and computational perspectives, especially if we hypothesize that:

1. mentions are not just unconstrained subsequences of tokens, but mostly correspond to syntactically meaningful units,
2. certain types of coreference relations are manifested primarily by syntactic means (such as reflexive and relative constructions),

3. the syntactic head of a mention is more important than the rest of the mention, as it imposes agreement in congruent categories (gender, number, person) along coreference relations in languages with richer morphology,
4. there are linguistic phenomena that are relevant – but at the same time difficult to handle – both for dependency syntax and for coreference annotation schemes (such as coordination and other paratactic relations, zeros, named entities, and numerous tokenization subtleties), and it seems natural to use the same annotation conventions for both.

In this paper, we aim to bring empirical evidence for supporting the first hypothesis. We show that mention spans usually correspond to subtrees of dependency trees quite straightforwardly; and if they do not, it can be explained by errors in the dependency trees or by mention span errors in most cases. This also shows a useful side effect of our line of thought: a joint view on coreference and syntactic annotations could be used for detection of annotation inconsistencies in both.

There are dozens of coreference-related resources annotated for various languages under highly diverse schemes. In our study, we use 9 coreference datasets for 8 different languages selected from CorefUD 0.1 (Nedoluzhko et al., 2021), which is a collection of coreference datasets unified under a common scheme. Mention spans in all 9 datasets result from manual annotation. Dependency trees available in the collection follow the Universal Dependencies (UD) scheme (Nivre et al., 2020) and result from manual annotation in one case and from automatic parsing with UDPipe (Straka and Straková, 2017) in the 8 remaining cases. In all cases, the dependency trees came into existence independently of the coreference annotation.

2 Related work

The idea of considering coreference and syntactic information together was quite popular in the last two decades of the 20th century, generally accepted in the Meaning-Text theory (Mel’čuk, 1981) or in the Functional-Generative Description (Sgall et al., 1986). Coreference is also one of the main concepts underlying binding phenomena in generative syntax (Chomsky, 1993).

However, with the advent of large-scale annotated corpora, coreference and syntax have somewhat diverged. The syntax-aware annotation of coreference demands for manual syntactic annotation, which is very expensive and not always feasible. As a result, coreference relations in most existing large-scale annotated resources are marked on raw texts, textual spans being defined as co-referring mentions, see, e.g. Hinrichs et al. (2005); Uryupina et al. (2020); Hendrickx et al. (2008); Désoyer et al. (2016); Landragin (2016); Bourgonje and Stede (2020); Guillou et al. (2014); Lapshinova-Koltunski et al. (2018); Žitkus and Butkienė (2018); Toldova et al. (2014). Some of these datasets (Hendrickx et al., 2008; Toldova et al., 2014) label syntactic heads of the mentions.

For some other datasets, syntactic annotation exists but it was created independently of coreference annotation. This is the case of GUM for English (Zeldes, 2017) or SzegedKoref, a coreferentially annotated corpus of Hungarian, which is a part of the Szeged Treebank (Csendes et al., 2005). Coreference in OntoNotes (Weischedel et al., 2011) is a special case, as potentially coreferential mentions have been first excerpted from constituency trees and then annotated independently.

To the best of our knowledge, there are only two large-scale coreference-annotated datasets where syntax is closely linked to coreference relations. In AnCora-CO (Recasens and Martí, 2010), co-referring mentions are nodes in constituency trees, and in the Prague Dependency corpora (Hajič et al., 2020; Nedoluzhko et al., 2016; Mikulová et al., 2017), coreference relations are annotated directly between syntactic heads in dependency trees and mention spans are implicitly defined as subtrees of the heads.

Finkel and Manning (2009) deal with issues similar to our work and have developed a model that performs named entity recognition and constituency parsing consistently, guaranteeing that named entity spans do not conflict with the phrasal spans of

the parse tree.

As for coreference resolution systems, some earlier algorithms took syntactic information into account, see e.g. Hobb’s naive approaches to pronoun resolution (Hobbs, 1978), Carter’s shallow processing approach (Carter, 1986) or fully symbolic Lappin and Leass’ algorithms for resolving third person pronouns and traversing syntactic trees (Lappin and Leass, 1994). Morpho-syntactic features were later largely used in statistical approaches (e.g., Ng and Cardie, 2002; Bergsma and Lin, 2006; Clark and Manning, 2015), especially for morphologically rich languages (e.g. Novák, 2017). With the advent of neural networks and contextual embeddings for coreference resolution (e.g., Lee et al., 2018; Joshi et al., 2019; Wu et al., 2020), the explicit treatment of morpho-syntax has practically vanished, even for the related task of mention detection. Such models are able to encode syntactic aspects implicitly, as shown by e.g., Hewitt and Manning (2019) and Limisiewicz et al. (2020).

3 Data selection

We draw our empirical observations about correspondences between manually annotated mention spans and manually or automatically produced dependency trees from CorefUD 0.1 (Nedoluzhko et al., 2021), the biggest collection of coreference datasets converted to a harmonized scheme.

CorefUD 0.1 contains 17 datasets for 11 languages. We excluded CorefUD datasets that are much smaller than the rest (English and German data originated in the ParCorFull project (Lapshinova-Koltunski et al., 2018)), and also datasets in which mentions and syntactic structures correspond by design. In the latter case, coreference annotations made use either of constituency trees – an English dataset from OntoNotes (Weischedel et al., 2011), and Spanish and Catalan datasets from the AnCora project (Recasens and Martí, 2010)), or of dependency trees – a Czech dataset from the Prague Dependency Treebank (Hajič et al., 2020), and English and Czech datasets from the Prague Czech-English Dependency Treebank (Nedoluzhko et al., 2016).

The selection resulted in 9 datasets, for which we use their CorefUD labels: (1) English-GUM: Georgetown Multilayer Corpus (Zeldes, 2017) (the only resource with hand-annotated dependency trees); (2) French-Democrat: Democrat (Landragin, 2016); (3) German-PotsdamCC: Potsdam Com-

mentary Corpus (Bourgonje and Stede, 2020); (4) Hungarian-SzegedKoref: SzegedKoref: Hungarian Coreference Corpus (Vincze et al., 2018); (5) Lithuanian-LCC: Lithuanian Coreference Corpus (Žitkus and Butkienė, 2018); (6) Polish-PCC: Polish Coreference Corpus (Ogrodniczuk et al., 2013); (7) Russian-RuCor: RuCor: Russian Coreference Corpus (Toldova et al., 2014); (8) Dutch-COREA: Coreference Corpus and Resolution System for Dutch (Hendrickx et al., 2008); (9) English-ARRAU: The ARRAU Corpus of Anaphoric Information (Uryupina et al., 2020).¹

4 Matches and mismatches between mention spans and dependency trees

We took a random sample of sentences from the data described in Section 3 and manually examined instances where syntactic and coreference annotations do not match. The annotation work took about 5-8 hours per 100 examples, depending on the language. In cases where the language was understood by none of this work’s authors (Hungarian, Lithuanian, Dutch), we took advantage of public machine-translation services during the annotation process. In this section, we first describe how we understand matching coreference and syntax; then we discuss individual types of mismatches and their possible causes.

4.1 Correspondence criteria

Catena is defined as any connected subgraph of a dependency tree (Osborne et al., 2012).² This contrasts with the definition of (complete) **subtree**, which is a catena that spans a node and *all* its descendants in the tree. For example, in Figure 1, the nodes *American food* form a catena which is not a subtree because the nodes *and Japanese* are not included in it.

In this work, we focus on non-catena mentions. In future, we plan to analyze also catena-but-not-subtree mentions.³

¹COREA and ARRAU are taken from the non-public part of CorefUD 0.1, which are based on the original datasets in versions 1.0 and 2.1, respectively.

²Catenaes are also called *partial subtrees*, *chains* (Osborne et al., 2012) or *treelets* (Ding and Palmer, 2004).

³UD attaches most functional words as leaves, so e.g. in *for Bob*, *for* depends on *Bob*. Punctuation, prepositions, and conjunctions are a frequent cause of non-subtree mentions. In a preliminary experiment, we automatically filtered out such cases and observed many errors in parsing and mention span annotation in the remaining cases.

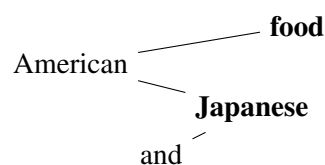


Figure 1: OK-COORD-ASCHILD example. An example of a non-catena mention caused by coordination as a child. The words in mention span are marked in bold.

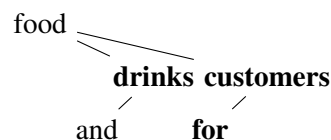


Figure 2: OK-COORD-ASPARENT example. An example of a non-catena mention caused by coordination as a parent. Mention span is marked in bold.

4.2 Non-catena mentions

Two main sources of non-catena mentions found in our data are errors in the dependency tree (WRONGTREE) and errors in annotation of the mention span (WRONGSPAN). These cases are further analyzed in Section 5. Here we describe the remaining types.

4.2.1 Coordination

There are many ways to annotate coordination in dependency treebanks (Popel et al., 2013). In UD, the first conjunct is the head and other conjuncts are attached to it (Nivre et al., 2020). Thus, when a mention span includes only one of the conjuncts plus other nodes, which either govern (Figure 1) or modify (Figure 2) the coordination, we can get non-catena mentions.⁴

4.2.2 Flat structures

Certain types of multi-word expressions such as names, foreign phrases, dates and complex numerals are annotated as flat structures, where all subsequent words in the expression are attached to the first one using the `flat` dependency relation.⁵ Usually, the whole multi-word expression is annotated as the mention, so it forms a catena. However, there are rare cases when the mention excludes the first word, e.g., *24 December 2016*. Such mentions

⁴We could also get discontinuous catena mentions, e.g., *American and Japanese food and drinks for customers* (mention highlighted in boldface). Furthermore, we could get discontinuous non-catena mentions, e.g., *food for employees or customers*.

⁵<https://universaldependencies.org/u/dep/flat.html>; lists are treated similarly

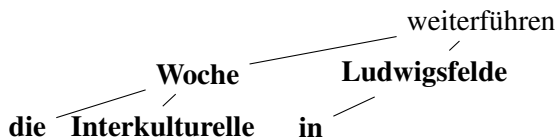


Figure 3: AMBIGUOUS example from German: *die Interkulturelle Woche in Ludwigsfelde weiterführen* “to continue the Intercultural Week in Ludwigsfelde.” The phrase *in Ludwigsfelde* can modify the verb or the noun *Woche* without a significant change in meaning. The parser attached it to the verb but the mention is annotated as if it modified the noun.

are not catenae and we mark them as OK-FLAT in our annotation in Section 5.

5 Evaluation

With the exception of English-GUM, in all other datasets considered in this study the syntactic annotation has been assigned by an automatic parser. Certain amount of parsing errors is thus to be expected and these errors may or may not cause the syntactic structure to diverge from the annotated mention spans. Table 1 shows that a significant portion of non-catenae mention spans are indeed caused by a wrong dependency tree. The confrontation of the syntactic structure with independent coreference annotation can thus be used as a signal for correction of the parsing errors.

There are unlimited ways how a parser can be wrong; nevertheless, a striking number of errors across languages falls to what could be described as a generalized PP attachment problem. A case-marked nominal (morphologically or with a preposition) should be analyzed as a modifier of the mention head but it is attached elsewhere, often to the predicate of the clause. We count as WRONGTREE only those instances where we are confident that the attachment is wrong; however, there are also AMBIGUOUS cases (Figure 3), where the alternative attachment is possible (without a clear shift in interpretation of the sentence). One could still expect that the tree structure will match the mention span, yet it is not clear whether it is the tree or the mention span what should be corrected.

Sometimes the mismatches revealed errors in mention span annotation. Such cases are less frequent than parsing errors, yet they are still considerable in some datasets (Lithuanian-LCC, German-PotsdamCC, English-GUM), and the syntactic annotation is helpful in identifying them. Both missing and extra words have been observed. Note that this

is not about by-design inclusion/omission of certain modifiers such as prepositions – these would affect the subtree condition but not the catena condition.

Our annotation focused only on non-catenae mentions, thus the percentage of WRONGSPAN and WRONGTREE reported in Table 1 should not be interpreted as an indicator of the amount of annotation errors in the whole dataset – catena mentions may also contain errors of both types.⁶

6 Conclusions and Future Work

We have investigated the correspondence of syntactic and coreferential annotation in a subset of CorefUD, a harmonized collection of coreference datasets with UD-style syntax. Based on a manually annotated sample, we have shown that most mention spans (86% to 98%, depending on dataset) correspond to catenae. More than a half of the mismatches can be attributed to errors in the syntactic structure, in some datasets there is also a significant number of errors in the mention spans. An automatic confrontation of independently annotated syntax and coreference can thus be used to find and correct annotation errors on either side. A relatively small proportion of the mismatches is justified by the UD guidelines for specific syntactic constructions (coordinations and flat structures).

Even if our findings are intrinsic with respect to CorefUD, the observation that certain elements of syntactic and coreference annotations tend to be strongly correlated even if created independently, suggest that the syntactic viewpoint should be taken into account for annotation of coreference in order to increase its consistency. The decision whether a particular mismatch has been incurred by the parse tree or by coreference annotation must be based on the parser’s quality and assumed coreference quality for individual resources. Further inspection of the resources may help to design heuristics to make the decision automatically. Otherwise, the corrections would have to be made manually.

Another natural future extension of this work is to look at mentions that are catenae but not com-

⁶It is possible that a catena mention contains a parsing error and in the correct tree, the span would not be catena anymore. Similarly, there may be a catena mention with wrongly annotated span and the correct span would not form a catena in the tree. Furthermore, both cases may coincide and after correcting both the tree and span, the mention may be catena again (as pointed out by a reviewer of this paper). We have not noticed any cases described in this footnote in the data, but for a proper investigation of these phenomena, we would need to annotate all mentions (including subtree mentions).

CorefUD dataset	non-catena		OK [%]		AMBIG. [%]	WRONG [%]	
	% of all	annotated	COORD	FLAT	TREE/SPAN	TREE	SPAN
English-GUM	1.5	100	10	9	15	52	22
French-Democrat	1.9	100	3	1	7	81	10
German-PotsdamCC	5.3	101	14	0	5	57	28
Hungarian-SzegedKoref	3.3	25	0	0	0	88	12
Lithuanian-LCC	4.6	50	0	0	0	76	64
Polish-PCC	13.5	100	0	2	2	95	1
Russian-RuCor	4.3	100	0	1	6	86	7
Dutch-COREA	5.8	51	0	0	0	88	18
English-ARRAU	13.1	100	4	0	4	88	4

Table 1: Distribution of different types of non-catena mentions. The “% of all” column indicates the percentage of non-catena mentions out of all mentions. “annotated” is the number of mentions manually annotated in our study. The remaining columns show percentages of the annotated non-catena mentions of a given type. In rare cases, a single mention can be annotated with multiple tags (e.g. WRONGSPAN and WRONGTREE), so the number of tags is sometimes higher than the number of all annotated mentions.

plete subtrees. Our preliminary experiments in this direction show that individual datasets in CorefUD have not been harmonized as to their mention span guidelines: e.g., some datasets include prepositions and relative clauses while others do not. Another closely related question is whether syntactic heads of mentions can be inferred automatically, and whether it would be sufficient to represent a mention by its head. On the other hand, extension towards more languages is limited by availability of harmonized data.

Acknowledgments

This work was supported by the Grants GA19-14534S and 20-16819X (LUSyD) of the Czech Science Foundation; LM2018101 (LINDAT/CLARIAH-CZ) of the Ministry of Education, Youth, and Sports of the Czech Republic; and EC/H2020/825303 (Bergamot) of the European Commission.

We thank three anonymous reviewers for their very insightful and useful comments.

References

Shane Bergsma and Dekang Lin. 2006. [Bootstrapping path-based pronoun resolution](#). In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 33–40, Sydney, Australia. Association for Computational Linguistics.

Peter Bourgonje and Manfred Stede. 2020. [The Pots-](#)

[dam commentary corpus 2.2: Extending annotations for shallow discourse parsing](#). In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1061–1066, Marseille, France. European Language Resources Association.

David Maclean Carter. 1986. [A shallow processing approach to anaphor resolution](#). Technical Report UCAM-CL-TR-88, University of Cambridge, Computer Laboratory, 15 JJ Thomson Avenue, Cambridge CB3 0FD, United Kingdom, phone +44 1223 763500.

Noam Chomsky. 1993. *Lectures on Government and Binding*. Studies in Generative Grammar (SGG). De Gruyter Mouton, Berlin, Germany.

Kevin Clark and Christopher D. Manning. 2015. [Entity-centric coreference resolution with model stacking](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1405–1415, Beijing, China. Association for Computational Linguistics.

Dóra Csendes, János Csirik, Tibor Gyimóthy, and András Kocsor. 2005. [The Szeged treebank](#). In *Proceedings of the 8th International Conference on Text, Speech and Dialogue, TSD’05*, page 123–131, Berlin, Heidelberg. Springer-Verlag.

Adèle Désoyer, Frédéric Landragin, Isabelle Tellier, Anaïs Lefeuvre, Jean-Yves Antoine, and Marco Dinarelli. 2016. [Coreference Resolution for French Oral Data: Machine Learning Experiments with ANCOR](#). In *17th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing’2016)*, Konya, Turkey.

Yuan Ding and Martha Palmer. 2004. Automatic learning of parallel dependency treelet pairs. In *Interna-*

- tional Conference on Natural Language Processing*, pages 233–243. Springer.
- Jenny Rose Finkel and Christopher D. Manning. 2009. Joint parsing and named entity recognition. In *Proceedings of Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 326–334. 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 9th International Conference on Language Resources and Evaluation (LREC-2014)*, Reykjavik, Iceland. European Language Resources Association (ELRA).
- Jan Hajič, Eduard Bejček, Jaroslava Hlaváčová, Marie Mikulová, Milan Straka, Jan Štěpánek, and Barbora Štěpánková. 2020. *Prague Dependency Treebank - Consolidated 1.0*. In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 5208–5218, Marseille, France. European Language Resources Association.
- Iris Hendrickx, Gosse Bouma, Frederik Coppens, Walter Daelemans, Veronique Hoste, Geert Kloosterman, Anne-Marie Mineur, Joeri Van Der Vloet, and Jean-Luc Verschelde. 2008. *A coreference corpus and resolution system for Dutch*. In *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco. European Language Resources Association (ELRA).
- John Hewitt and Christopher D. Manning. 2019. *A structural probe for finding syntax in word representations*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4129–4138, Minneapolis, Minnesota. Association for Computational Linguistics.
- Erhard W. Hinrichs, Sandra Kübler, and Karin Naumann. 2005. *A unified representation for morphological, syntactic, semantic, and referential annotations*. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky*, pages 13–20, Ann Arbor, Michigan. Association for Computational Linguistics.
- J.R. Hobbs. 1978. Resolving pronoun references. *Lingua* 44, pages 311–338.
- Mandar Joshi, Omer Levy, Luke Zettlemoyer, and Daniel Weld. 2019. *BERT for coreference resolution: Baselines and analysis*. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5803–5808, Hong Kong, China. Association for Computational Linguistics.
- Frédéric Landragin. 2016. *Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT)*. *Bulletin de l'Association Française pour l'Intelligence Artificielle*, 92:11–15.
- Shalom Lappin and Herbert J. Leass. 1994. *An algorithm for pronominal anaphora resolution*. *Computational Linguistics*, 20(4):535–561.
- Ekaterina Lapshinova-Koltunski, Christian Hardmeier, and Pauline Krielle. 2018. ParCorFull: a Parallel Corpus Annotated with Full Coreference. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Kenton Lee, Luheng He, and Luke Zettlemoyer. 2018. *Higher-order coreference resolution with coarse-to-fine inference*. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 687–692, New Orleans, Louisiana. Association for Computational Linguistics.
- Tomasz Limisiewicz, David Mareček, and Rudolf Rosa. 2020. *Universal Dependencies According to BERT: Both More Specific and More General*. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 2710–2722, Online. Association for Computational Linguistics.
- Igor A. Mel'čuk. 1981. *Meaning-text models: A recent trend in soviet linguistics*. *Annual Review of Anthropology*, 10:27–62.
- Marie Mikulová, Jiří Mírovský, Anna Nedoluzhko, Petr Pajas, Jan Štěpánek, and Jan Hajič. 2017. *PDTSC 2.0 - spoken corpus with rich multi-layer structural annotation*. In *Text, Speech, and Dialogue. TSD 2017*, volume 10415 of *Lecture Notes in Computer Science*, Cham, Switzerland. Springer.
- 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, Martin Popel, Zdeněk Žabokrtský, and Daniel Zeman. 2021. *Coreference meets Universal Dependencies – a pilot experiment on harmonizing coreference datasets for 11 languages*. Technical Report 66, ÚFAL MFF UK, Praha, Czechia.
- Vincent Ng and Claire Cardie. 2002. *Improving machine learning approaches to coreference resolution*. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 104–111, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

- Joakim Nivre, Marie-Catherine de Marneffe, Filip Ginter, Jan Hajič, Christopher Manning, Sampo Pyysalo, Sebastian Schuster, Francis Tyers, and Daniel Zeman. 2020. [Universal Dependencies v2: An evergrowing multilingual treebank collection](#). In *Proceedings of the 12th International Conference on Language Resources and Evaluation (LREC 2020)*, pages 4034–4043, Marseille, France. European Language Resources Association.
- Michal Novák. 2017. Coreference Resolution System Not Only for Czech. In *Proceedings of the 17th conference ITAT 2017: Slovenskočeský NLP workshop (SloNLP 2017)*, pages 193–200, Praha, Czechia. CreateSpace Independent Publishing Platform.
- Maciej Ogrodniczuk, Katarzyna Glowńska, Mateusz Kopeć, Agata Savary, and Magdalena Zawistawska. 2013. [Polish coreference corpus](#). In *Human Language Technology. Challenges for Computer Science and Linguistics - 6th Language and Technology Conference, LTC 2013, Poznań, Poland, December 7-9, 2013. Revised Selected Papers*, volume 9561 of *Lecture Notes in Computer Science*, pages 215–226. Springer.
- Timothy Osborne, Michael Putnam, and Thomas Groß. 2012. Catenae: Introducing a novel unit of syntactic analysis. *Syntax*, 15(4):354–396.
- Martin Popel, David Mareček, Jan Štěpánek, Daniel Zeman, and Zdeněk Žabokrtský. 2013. [Coordination structures in dependency treebanks](#). In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 517–527, Sofia, Bulgaria. Association for Computational Linguistics.
- Marta Recasens and M. Antònia Martí. 2010. [AnCoraCO: Coreferentially Annotated Corpora for Spanish and Catalan](#). *Lang. Resour. Eval.*, 44(4):315–345.
- Petr Sgall, Eva Hajičová, and Jarmila Panevová. 1986. *The Meaning of the Sentence and Its Semantic and Pragmatic Aspects*. Academia/Reidel Publishing Company, Prague/Dordrecht.
- Milan Straka and Jana Straková. 2017. [Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe](#). In *Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies*, pages 88–99, Vancouver, Canada. Association for Computational Linguistics.
- S. Toldova, A. Roytberg, A. A. Ladygina, M. D. Vasilyeva, I. L. Azerkovich, M. Kurzukov, G. Sim, D. V. Gorshkov, A. Ivanova, A. Nedoluzhko, and Y. Grishina. 2014. Evaluating anaphora and coreference resolution for russian. In *Komp'juternaja lingvistika i intellektual'nye tehnologii. Po materialam ezhegodnoj Mezhdunarodnoj konferencii Dialog*, pages 681–695.
- Olga Uryupina, Ron Artstein, Antonella Bristot, Federica Cavicchio, Francesca Delogu, Kepa J. Rodriguez, and Massimo Poesio. 2020. [Annotating a broad range of anaphoric phenomena, in a variety of genres: the ARRAU Corpus](#). *Natural Language Engineering*, 26(1):95–128.
- Veronika Vincze, Klára Hegedűs, Alex Sliz-Nagy, and Richárd Farkas. 2018. [SzegedKoref: A Hungarian coreference corpus](#). In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA).
- Ralph Weischedel, Eduard Hovy, Mitchell Marcus, Martha Palmer, Robert Belvin, Sameer Pradhan, Lance Ramshaw, and Nianwen Xue. 2011. Ontonotes: A large training corpus for enhanced processing. In *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*, pages 54–63, New York. Springer-Verlag.
- Wei Wu, Fei Wang, Arianna Yuan, Fei Wu, and Jiwei Li. 2020. [CorefQA: Coreference resolution as query-based span prediction](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6953–6963, Online. Association for Computational Linguistics.
- Amir Zeldes. 2017. [The GUM Corpus: Creating Multilayer Resources in the Classroom](#). *Language Resources and Evaluation*, 51(3):581–612.
- Voldemaras Žitkus and Rita Butkienė. 2018. [Coreference annotation scheme and corpus for Lithuanian language](#). In *Fifth International Conference on Social Networks Analysis, Management and Security, SNAMS 2018, Valencia, Spain, October 15-18, 2018*, pages 243–250. IEEE.