

Improving Coreference Resolution with Word Formation

Wei Zhao, Kevin Mathews and Haixia Chai

Heidelberg Institute for Theoretical Studies

Schloss-Wolfsbrunnenweg 35, 69118 Heidelberg, Germany

{wei.zhao, kevin.mathews, haixia.chai}@h-its.org

In natural language processing (NLP), coreference resolution, aiming for the identification of a collection of expressions referring to the same real-word entity, allows for a better understanding in natural language, and as such has received attention for long. Given the unprecedented success of contextualized language models such as BERT (Devlin et al., 2019), employing them in NLP tasks requiring deep semantic understanding has achieved extraordinary performances, as it is the case in coreference resolution (Joshi et al., 2019). However, previous work does not outline the current state of limitations in coreference resolution systems, such as where and why they cannot resolve coreference.

In discourse linguistics, word formation has a long-lasting connection with coreference resolution. Research has shown that employing syntactic function of word formation, such as pronominalizing compounds—a coreference-like approach, can decrease “readers” cognitive load, and accordingly, increase text coherence (SCHRODER, 1978; Dederding, 1983; Lipka, 1987; Eichinger, 1995; Schliez, 2004; Peschel, 2011). But previous studies are limited in scope to nominal compounds as the only pattern of word formation, and more prominently, they do not contribute to the current state of coreference resolution systems.

In this work, we aim for employing word formation to understand the deficiencies of coreference resolution systems and address them. To this end, we intend to assemble noun phrases from the largest coreference dataset CoNLL (Pradhan et al., 2012) and its extension (Chai et al., 2020), and relate the noun phrases to word formation patterns, such as compounds, word abbreviations, noisy words, metaphor and metonymy. Next, we will resolve coreference with the current state-of-the-art coreference resolution system (Joshi et al., 2020), and then analyze the impact of word for-

mation patterns on coreference resolution. Finally, we will propose a simple approach to improve the coreference resolution system with rewriting the noun phrases for which the corresponding coreference is challenging to resolve. In particular, we intend to mask these noun phrases and employ BERT to perform a fill-in-the-blank task, i.e., predicting masked words that fit best in the context. In the following, we elaborate on our approach.

Our Approach. First, we mask each target word, and then employ BERT to retrieve a shortlist of candidates pertaining to the target word, *v.i.z.*, retaining a set of words with the probabilities above a threshold. We illustrate how to choose per target word a substitution from shortlisted candidates:

- Novel word sense: we use cosine similarity to measure distances between synset (word sense) definitions of a target word and of each word in shortlisted candidates, based on their embeddings. We replace a target word with its nearest neighbor in the shortlist according to synset similarities. We acknowledge that this approach requiring a word sense disambiguation tool to identify word synset, which is prone to wrong identification.
- Novel words are not present in the BERT vocabulary and WordNet, thereby not associated with word embeddings and synset definitions. This means we cannot find per novel word a substitution, based on their semantic and synset similarities. To this end, we use the word with the highest probability in the vocabulary to replace a target word. Such substitutions, although contextually related, may be not semantically related to (or synonymous with) target words.

References

- Haixia Chai, Wei Zhao, Steffen Eger, and Michael Strube. 2020. [Evaluation of coreference resolution systems under adversarial attacks](#). In *Proceedings of the First Workshop on Computational Approaches to Discourse*, pages 154–159, Online. Association for Computational Linguistics.
- Hans-Martin Dederding. 1983. Wortbildung und text. zur textfunktion (tf) von nominalkomposita (nk).
- Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). 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 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.
- Ludwig M Eichinger. 1995. Wegweiser durch textwelten. wozu komplexe substantive gut sind.
- Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. 2020. [SpanBERT: Improving pre-training by representing and predicting spans](#). *Transactions of the Association for Computational Linguistics*, 8:64–77.
- Mandar Joshi, Omer Levy, Daniel S. Weld, and Luke Zettlemoyer. 2019. BERT for coreference resolution: Baselines and analysis. In *Empirical Methods in Natural Language Processing (EMNLP)*.
- Leonhard Lipka. 1987. Word-formation and text in english and german.
- Corinna Peschel. 2011. *Zum Zusammenhang von Wortneubildung und Textkonstitution*, volume 237. Walter de Gruyter.
- 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.
- Michael Schlienz. 2004. *Wortbildung und Text: eine Untersuchung textverknüpfender Wortbildungselemente*. Palm und Enke.
- Marianne SCHRODER. 1978. Über textverflechtende wortbildungselemente. *Deutsch als Fremdsprache. Zeitschrift für Theorie und Praxis des Deutschunterrichts für Ausländer Leipzig*, 15(2):85–92.