

Review of thesis proposal

Reviewer: David Mareček, ÚFAL MFF UK

Date: February 28, 2020

Thesis Title: Dependency Parsing beyond Simple Trees

Candidate: Kira Droganova

Supervisor: Dan Zeman, ÚFAL MFF UK

Proposal content

The goal of the thesis is to further enhance the Universal Dependencies (UD) annotations by deeper syntactic or semantic representations utilizing language universal rules or existing lexical resources available in their native frameworks such as valency lexicons. The first two sections introduce the problem and overview related work.

Kira's previous work is summarized in Section 3. She describes her analysis of elliptic constructions in UD. She also implemented data enrichments methods needed to train a parser that would reasonably deal with such constructions. Then, she introduces "Deep UD project" adding deep annotations derived from surface trees, e.g. identification of arguments in verbal predicates.

In Section 4, Kira proposes several other linguistic phenomena useful for deeper annotation which she wants to focus on in the future. The proposed structures are generally oriented graphs, which may contain cycles, so, in Section 5, she discusses possible parsers that could be used for parsing such representations. Section 6 summarizes downstream tasks suitable for the extrinsic evaluation of the parser.

Questions and comments

- 1) I would appreciate more technical details and examples in the proposal. For example, what is the format of the Enhanced UD in the data? Are there any additional columns expressing another arcs and labels? And what values are given into the two additional Deep UD columns DEEP:PRED and DEEP:ARGS? How works the algorithm for detecting annotation errors in ellipses? Is it rule-based? Can you show an example? It

would be nice if you could show at least some of these issues at the defence since not anyone is familiar with Enhanced UD.

- 2) Many of the downstream applications for the extrinsic evaluation proposed in Section 6 seems to be obsolete, for example, the TectoMT system from 2011 or Textual Entailment system from 2010 were substantially outperformed by recent end-to-end neural networks. Of course, such tasks can be used for comparison of different representations, but it would be better to work with tasks in which UD really helps. Do you know about any recent downstream applications that would benefit from the UD annotations compared to, for example, the fine-tuned BERT?
- 3) The proposed future enhancements are written very vaguely. Do you have any specific plans (rules or algorithms) that you plan to implement to add a new deeper phenomenon into UD? Could you show a specific example in your presentation?

Conclusion

Overall, I think Kira proved that she knows the area of Universal Dependencies very well. Her recent work was published in four conference papers. The proposal is written in very good English with very little mistakes and typos. As I am not an expert on the UD project, I do not know much about the novelty of her work nor about possible concurrent projects. However, I believe her supervisor knows about everything important. I think that if Kira is able to convert to UD at least several other phenomena she has proposed (even though I cannot guess how much work may be behind it) it will be sufficient for a good doctoral thesis. I recommend this proposal to be defended.