UniDive 2024 training school application

- *Title*: A Collaborative Grammatical Annotation Tool, BoAT
- *Name and affiliation*: Furkan Akkurt and Boğaziçi University, Turkey
- *Keywords*: annotation tool, collaboration, web application
- *Resource description*
    - BoAT [1] (Boğaziçi University Annotation Tool) is a collaborative grammatical annotation tool that is able to annotate all fields of UD treebanks. It started as my senior project during my undergraduate studies (Computer Science) where we started with requirements elicitation for the tool with an annotator of the UD Turkish BOUN treebank. However, I think the tool should be improved much more and / or some of its features should be integrated into other annotation tools.
        - It's a collaborative tool where multiple annotators can work on the same treebank and analyze each other's annotations.
        - It has a search functionality to query by all linguistic fields of UD treebanks (e.g. *POS*, *FEATS*).
            - As it's collaborative, by searching, one can find other people's annotations and learn / confirm what the annotation should be for a specific construction.
            - The tool allows search by multiple fields, so, for example, one can ask for annotations where there is a token with lemma *al* and category *ADJ*.
        - BoAT is designed first and foremost for agglutinative languages in mind. Treebanks of agglutinative languages need much morphological feature annotation and it's a very labor-heavy task. BoAT allows linguists to annotate by using only the keyboard so as not to disturb their flows by mouse - keyboard switches as sometimes there are very long sessions of annotation, which makes it harder to concentrate.
        - It has an autocomplete feature for the fields where there are UD documentation, such as morphological features.
        - It has a very preliminary inter-annotator agreement calculation but this should be, hopefully, improved and made into a much more useful calculation where the agreement could be calculated between each pair of annotators.
- *Explanation how the participation in the training school will be useful for the project*
    - I will be able to gauge what features would be good for further improvement of the tool, especially from the lecture on 'Corpus annotation infrastructure'. I have also ideas about new features and discussion in the school with real

annotators would feed the further requirements elicitation of the tool. Some features would be:

- Various types of inter-annotator agreement score calculation;
- Comparison of the current treebank with other existent treebanks of the language of the treebank at hand;
- Visualization of the current progress of the annotations;
- Automatic annotations by rules given by an annotator;
- Metadata annotation ability as some treebanks are using the metadata as a source of important information, such as language contact and geographic locations.

- I also want to further learn the features of Arborator and its Grew extension. I may also be able to help / discuss ideas with the team for further improvement of the Arborator-Grew tool, as I know they're still adding new features, such as the recent Git integration.

- *Open questions related to the project which could be addressed during the brainstorming hackathon*
  1. What parts of the treebank should the linguists be allowed to annotate?
  2. How much automatization can / should we do based on the language of the treebank being annotated and other existent treebanks of the language?
- *Project status*: The application is still being improved with new features, performance improvements and bug fixes.

References

1. github.com/furkanakkurt1335/boat