# A Collaborative Grammatical Annotation Tool for Agglutinative Languages

March 2023

Grammatically annotated sets of sentences serve automation by computational natural language processing (NLP) tools and act as a reference for a language's syntactic, morphological and dependency structures. Tools leveraging computational processing of natural language need datasets with accurate and consistent linguistic annotations. Even though these annotations are being done by experienced linguists, they still require lots of labor and time. Such tools exist for this purpose but are mostly designed with analytic languages in mind. Annotation of agglutinative languages involves extensive morphological feature annotation per token and splitting of lemmas that their analytic counterparts don't need. Thus, their annotation warrants a different design interface to accommodate the differences in their annotations. Because tools exist mostly for analytic languages, high-quality datasets exist for mostly those languages as well, with agglutinative languages, like Turkish, being low-resource in the field. In order for agglutinative languages to accumulate such resources, we need tools capable of this task. In this paper, we describe the requirements, design and use cases of a collaborative annotation tool that is designed with morphologically rich languages in mind from the ground-up. This tool uses, as a foundation, the ideas of a previous annotation tool that had been developed to undertake the annotation of a Turkish treebank. The feedback provided by this treebank's annotators has been instrumental in the requirements gathered for this new tool. As such, our tool's main goals have been to facilitate creation of quality datasets with consistently accurate annotations, improve the efficiency of the process while making sure the annotators have a smooth user experience, and provide a collaborative space for annotators to make use of each other's annotations and provide feedback to one another. We provide the tool as an open-source resource accessible to everyone in the computational linguistic community. We also provide our evaluations, discussions, future directions and possible use cases of the tool. The tool has already found real-life use and we expect more uses of it to come in the future. We hope this conference will disseminate our work further and increase interactions between computational and linguistic folks.