

Annotation schemes of various language treebanks usually suffer substantial alterations because each of them rely on different grammatical traditions. These discrepancies cause severe obstacles when comparing parsing results across languages. One of the latest attempts to unify the cross-linguistic framework under a single annotation scheme is the Universal Dependencies project (McDonald et al., 2013). This initiative represents syntax based on word-to-word dependencies, i.e. Dependency Grammar, developed by Lucien Tesnière (Tesnière, 1959).

In this thesis we present the first publicly available syntactic treebank for Buryat, annotated according to the UD formalism. Buryat is a minority language of an endangered status, it is the second official and national language of the Republic of Buryatia, one of Russian regions in the south of Siberia. Since it is a low-resourced language, it has not received proper attention from the viewpoint of language technology. Our Buryat treebank contains manually annotated POS tags and UD trees for 418 sentences from public domains. First, we introduce the annotation guidelines, provided with morphologically and syntactically analyzed examples. Furthermore, we report parsing experiments of a state-of-the-art pipeline UDPipe. We train and test different language models, which vary according to the number of features. In conclusion, we suggest orientations for future research.

Keywords: Universal Dependencies, Buryat, Treebank, UDPipe