

In this work we present the method of semi-automatic training of the probabilistic translation dictionary using large automatically annotated parallel corpora. According to the study of translation errors and the role of translation dictionary within the TectoMt translatio system in general we propose models of various complexity. These basic models were combined to hierarchical models that were designed to reduce impact of the sparse data problem. Various extensions were implemented to deal with common lexical errors. The dictionary along with extensions was compared to the former approach on test data and the results show improved translation quality.