

## Īā īāīīī īāōīāā āāōīīāōē÷āñēīē āōāīīāōē÷āñēīē ðāçīāōēē ñōāōīīā÷āōīūō òāēñōīā

Āāōīð Āðōāī Āēēōīðīāē÷ Āīāðāāā  
07.09.2014 ā.

A  
method is proposed for unsupervised morphosyntactic markup of old texts for which no exact grammar nor vocabulary may be known. The method employs building all possible mappings from text forms into grammemes and then reducing them using a loose context-free (CF) grammar. The forms are further lemmatized based on minimization of morphologic variation. The method has been tested on two old Lithuanian documents from the late 16th century by M. Dauksha and has proven to be rather efficient and accurate (up to 80 %).

andreev\_elmanuscript2014 (594.8 kB)