

This work describes a machine learning approach for checking the part-of-speech annotation, and presents its implementation – a system called MissTagger. The checking procedure covers both error detection and error correction. MissTagger employs a simplified instance-based learning algorithm where the words in the text are recognized as instances. Part-of-speech tags of context of static length are selected as features, no lexical information is included. The words whose tags comprises this context are chosen based either on a linear or on a dependency-tree structure of the sentence. Two languages are examined in the experiments for evaluation, Czech and English.