

Natural language correction, a subfield of natural language processing (NLP), is the task of automatically correcting user errors in written texts. It includes, but is not limited to, grammatical error correction, spelling error correction and diacritics restoration. During the course of the work on this thesis, we witnessed a great advance in this field, with the emergence of new approaches to correct user errors, new datasets and also new evaluation metrics. This thesis presents, in the form of a dissertation by publication, our contributions to this field. As Czech is the primary language of the thesis author, special focus was devoted to improving natural language correction in Czech. The main contributions are (1) the creation of the Grammar Error Correction Corpus for Czech that comprises multiple sources of noisy texts such as essays or online discussion posts, evaluation of strong neural models on this dataset, and meta-evaluation of existing metrics, (2) the development of grammar error correction systems suited to scenarios in which only low amount of annotated data is available, and (3) the development of two state-of-the-art models and the creation of the new multilingual dataset comprising 12 languages for diacritics restoration.