In the present work we study semi-automatic evaluation techniques of machine translation (MT) systems based on comparison of the MT system's out-put to human translations of the same text. Various metrics were proposed in the past years, ranging from metrics using only unigram comparison to metrics that try to take advantage of additional syntactic or semantic information. The main goal of this thesis is to compare these metrics with respect to their correlation with human judgments and to propose the most suitable ones for evaluation of MT systems with Czech as target language. An implementation of a tool that computes the MT metrics is part of this work.