Abstract

This research will be focusing on making statistical machine translation for Minang – Indonesian language and seeing how well the machine translation results. The training and test data source is in the form of parallel corpus and monolingual corpus and the data comes from the Indonesian Traditional Minang language Wikipedia and the Indonesian news website. All configuration experiments were carried out using 600 lines of data. Two test scenarios were carried out, namely using scenarios based on monolingual corpus and parallel corpus. To see how much improvement will be seen by using automatic testing Bilingual Evaluation Understudy (BLEU). The test results on 6 configurations showed an increase in machine translation accuracy after the number of monolingual and parallel corpus was added. In the first scenario configurations 3 and 2 increased by 3.6%, configurations 2 and 1 increased by 2.59%. In the scenario of two configurations 5 and 4 it increases by 0.44%, configurations 4 and 1 increase by 0.06%. The results showed in this testing that the first scenario testing had more significant impact to the translation compared to the second scenario in the translated version. The lack of corpus sources is a problem in making statistical machine translators.

Keywords: minang language, indonesia language, statistical machine translation, bilingual evaluation understudy