Machine Translation Quality Estimation predicts quality scores for translations produced by Machine Translation systems based on source and output segments. Quality Estimation systems are usually trained in a supervised manner using training data that contains translation produced by one or more (other) Machine Translation systems. Therefore, the choice of training data for Machine Translation has an impact on how well the Quality Estimation system works.

This thesis studies the relationship between Machine Translation systems and sentence-level Quality Estimation systems. Using our definitions of Machine Translation system power and Quality Estimation system power, we conducted experiments that involve training Machine Translation and Quality Estimation systems of varying power. We presented Quality Estimation systems evaluation results on test sets of different domains and translated by Machine Translation systems of different power. We find that (i) Quality Estimation systems trained on translations of lower quality outperform Quality Estimation systems trained on translations of higher quality; (ii) evaluating high-quality Machine Translation systems is challenging for Quality Estimation systems of all powers; (iii) high-power Quality Estimation systems work better for out-of-domain distribution than low-power Quality Estimation systems.