

Priming Effects in Post-Editing Machine Translation

Silvia Hansen-Schirra (Johannes Gutenberg-Universität Mainz) hansenss@uni-mainz.de

Oliver Čulo (Johannes Gutenberg-Universität Mainz) culo@uni-mainz.de

Jean (Johannes Gutenberg-Universität Mainz) Nitzke nitzke@uni-mainz.de

Due to the growing demand for translational services in a globalized market, the use of so-called post-editing machine translation (PEMT) has moved into focus as a more efficient and cost-effective method of translation. In order to facilitate and support high quality human translation PEMT has been integrated in existing TM systems and CAT tools. This integration into translational environments optimizes the translation workflow and minimizes the post-editing effort. However, questions like the quality of post-edited texts and potential influence from machine translation (MT) on the final product as well as on the cognitive processes during the post-editing process are currently in the focus of research. Recently, a study by Bangalore and colleagues indicated that post-edited texts strongly exhibit syntactic priming from machine translated texts on the syntactic level. In other words, syntactic structures proposed by an MT system are generally taken over, with very little variation, as compared to human translations of the same texts.

In the present paper, we depart from the assumption that post-edited texts will also exhibit features different from those of human-translated texts on the lexical level. We investigate a corpus of English-German post-edited texts including eye-tracking and key-logging to show how post-editors are influenced by the MT output on the lexical level. For instance, we classify unidiomatic renderings in a number of post-edited versions as well as lexical inconsistencies, which were introduced by the MT system and left uncorrected by half of the post-editors in our experiment. On the basis of eye-tracking and key-logging data, we will discuss to what extent the post-editors' cognitive processes are primed by the MT output.

In addition, we will use the perplexity coefficient borrowed from machine translation to measure the consistency of terms and terminological variation of term translation. We will present a schema of how to classify translation variants across different translation settings and sessions (post-editing vs. human translation) as different types of events in order to make them comparable in terms of consistency, and usable for perplexity calculation. On this basis priming effects in the post-editing data will be detected and discussed on the basis of empirical data.