



Word alignments were generated using fast_align tool (Dyer et al., 2013) in the symmetric intersection mode. Tokenization and other preprocessing were performed using cdec tool kit. Stanford Segmenter (Tseng et al., 2005) was used to preprocess the Chinese corpora.