

Automatic Word Alignment Error Analysis for Myanmar-English Translation

Nway Nway Han

nwaynwayhan@ucsm.edu.mm

University of Computer Studies, Mandalay, UCSM

Aye Thida

ayethida@ucsm.edu.mm

University of Computer Studies, Mandalay, UCSM

Word alignment is a key task for every innovative statistical machine translation (SMT) system. In this paper, we present and compare the results of three word alignment tools to analyze the errors in an Automatic Word Alignment (AWA) system for Myanmar-English translation. AWA system needs bilingual corpora to align the words in the parallel corpus. However, Myanmar is an inflected language and a language of scarce resource. For this reason, we developed a manually sentence aligned bilingual corpus which has three thousand sentence pairs for general domains. This paper is the first in error analysis for the automatic word alignment process between Myanmar and English languages. We investigate the errors and missed alignments by training Myanmar-English sentence pairs with GIZA++, Berkeley Aligner and fast_align. We also present the type of errors based on the language nature to find out the best way for Myanmar-English Automatic Word Alignment System. Experimental results of three alignment tools are compared with accuracy and alignment error rate in terms of F1 score and AER which against the manually annotated Gold Standard corpus.