

## **Statistical Machine Translation Using 5-grams Word Segmentation in Decoding**

**Aye Thida**  
**ayethida.royal@gmail.com**  
**University of Computer Studies,**  
**Mandalay, UCSM**

**Nway Nway Han**  
**nwaynwayhan@ucsm.edu.mm**  
**University of Computer Studies,**  
**Mandalay, UCSM**

**Sheinn Thawtar Oo**  
**sheinthawtaroo@ucsm.edu.mm**  
**University of Computer Studies,**  
**Mandalay, UCSM**

This paper presents UCSMNLP's submission to the WAT 2018 Translation Tasks focusing on the Myanmar-English translation for mixed domain tasks. In statistical machine translation (SMT), word segmentation is a necessary step to generate more fluent translation results. Myanmar is one of the low resource languages and many researches are supposed to develop word segmentation for Myanmar language. However, there are no public tools for Myanmar word segmentation. This paper addresses the problem of word segmentation for Myanmar language in SMT by developing syllable, 5-grams and longest matching word segmentation with monolingual lexicon. This paper describes the phrasebased statistical machine translation (PBSMT) with batch version of MIRA tuning using 5-grams word segmentation for English-Myanmar language pairs in both directions. The experimental results showed that the baseline SMT with 5-grams could outperform the baseline system in terms of BLEU, RIBES and AMFM scores.