The Application of Computer-Aided Under-Resourced Language Translation for Malay into Kadazandusun

ABSTRACT

A computer-aided language translation using a Machine translation (MT) is an application performed by computers (machines) that translates one natural language to another. There are many online language translation tools, but thus far none offers a sequence of text translations for the under-resourced Kadazandusun language. Although there are web-based and mobile applications of Kadazandusun dictionaries available, the systems do not translate more than one word. Hence, this paper aims to present the discussion of the preliminary translation of Malay to Kadazandusun. The basic word-to-word with dictionary alignment translation based on Direct Machine Translation (DMT) is selected to begin the exploration of the translation domain where DMT is one of the earliest translation methods which relies on the word-to-word approach (sequence-to-sequence model). This paper aims to investigate the under-resourced language and the task of translating from the Malay language to the Kadazandusun language or vice versa. This paper presents the application and the process as well as the results of the system according to the basic Kadazandusun word arrangement (Verb-Subject-Object) and its translation quality using the Bilingual Evaluation Understudy (BLEU) score. Several phases are involved during the process, including data collection (word pair translation), preprocessing, text selection, translation procedures, and performance evaluation. The preliminary language translation approach is proven to be capable of producing up to 0.5 BLEU scores which indicate that the translation is readable, however, requires post-editing for better comprehension. The findings are significant for the quality of the under-resourced language translation and as a starting point for other machine translation methodologies such as statistical or deep learning-based translation.