

Keynote Talk II

Multimodal Simultaneous Machine Translation

Lucia Specia
Imperial College London, England

Abstract

Simultaneous machine translation (SiMT) aims to translate a continuous input text stream into another language with the lowest latency and highest quality possible. Therefore, translation has to start with an incomplete source text, which is read progressively, creating the need for anticipation. In this talk I will present work where we seek to understand whether the addition of visual information can compensate for the missing source context. We analyse the impact of different multimodal approaches and visual features on state-of-the-art SiMT frameworks, including fixed and dynamic policy approaches using reinforcement learning. Our results show that visual context is helpful and that visually-grounded models based on explicit object region information perform the best. Our qualitative analysis illustrates cases where only the multimodal systems are able to translate correctly from English into gender-marked languages, as well as deal with differences in word order, such as adjective-noun placement between English and French.