

Keynote Abstract: Events on a Global Scale: Towards Language-Agnostic Event Extraction

Elizabeth Boschee

University of Southern California

CA, California, U.S.A.

boschee@isi.edu

Event extraction is a challenging and exciting task in the world of machine learning & natural language processing. The breadth of events of possible interest, the speed at which surrounding socio-political event contexts evolve, and the complexities involved in generating representative annotated data all contribute to this challenge. One particular dimension of difficulty is the intrinsically global nature of events: many downstream use cases for event extraction involve reporting not just in a few major languages but in a much broader context. The languages of interest for even a fixed task may still shift from day to day, e.g. when a disease emerges in an unexpected location.

Early approaches to multi-lingual event extraction (e.g. ACE) relied wholly on supervised data provided in each language of interest. Later approaches leveraged the success of machine translation to side-step the issue, simply translating foreign-language content to English and deploying English models on the result (often leaving some significant portion of the original content behind). Most recently, however, the community has begun to show significant progress applying zero-shot transfer techniques to the problem, developing models using supervised English data but decoding in a foreign language without translation, typically using embedding spaces specifically designed to capture multi-lingual semantic content.

In this talk I will discuss multiple dimensions of these promising new approaches and the linguistic representations that underlie them. I will compare them with approaches based on machine translation (as well as with models trained using in-language training data, where available), and discuss their strengths and weaknesses in different contexts, including the amount of English/foreign bitext available and the nature of the target event ontology. I will also discuss possible future directions with an eye to improving the quality of event extraction no matter its source around the globe.

***Bio:** Elizabeth Boschee is the Director of the Boston office of the University of Southern California's Information Sciences Institute and a Senior Supervising Computer Scientist in the Emerging Activities division. Her current efforts focus on cross-lingual information extraction, retrieval, and summarization, specifically targeting low or zero-resource settings, e.g. cross-lingual settings with <1M words of bitext or event extraction from non-English languages with only English training data. Prior to joining ISI, Boschee spent 17 years at BBN Technologies. As a Lead Scientist there, she was the chief architect of the BBN ACCENT event coder, the technology behind the W-ICEWS event data, which more than doubled the precision (while still increasing recall) of the previously deployed solution for CAMEO event coding.*