

# **Revealing Long-term Language Change with Subword-incorporated Word Embedding Models**

**Yang Xu**

San Diego State University, San Diego, California, United States

**Jiasheng Zhang**

The Pennsylvania State University, State College, Pennsylvania, United States

**David Reitter**

Penn State, University Park, Pennsylvania, United States

## **Abstract**

We propose an augmented word embedding model that better incorporates subword information with additional parameters that characterize the semantic weights of characters in composing words. Our model can reveal some interesting patterns of long-term change in Chinese language, which provides novel evidence and methodology that enriches existing theories in evolutionary linguistics. The resulting word vectors also has decent performance in NLP-related tasks.