UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Revealing Long-term Language Change with Subword-incorporated WordEmbedding Models

Permalink
https://escholarship.org/uc/item/1818j7tg

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 41(0)

Authors
Xu, Yang
Zhang, Jiasheng
Reitter, David

Publication Date
2019

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

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.