UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Language Models Show Within- and Cross-language Similarities in Concrete Noun Meaning, but not Differences Between L1 and L2 English Speakers

Permalink

https://escholarship.org/uc/item/4748z6vz

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 45(45)

Authors

She, Jingyuan Selena
Ma, Gabrielle
Wen, Ping
et al.

Publication Date

2023

Peer reviewed

Language Models Show Within- and Cross-language Similarities in Concrete Noun Meaning, but not Differences Between L1 and L2 English Speakers

Jingyuan She
Haverford College, Haverford, Pennsylvania, United States

Gabrielle Ma
Swarthmore College , Swarthmore, Pennsylvania, United States

Ping Wen
Haverford College, Haverford, Pennsylvania, United States

Benjamin Zinszer
Swarthmore College, Swarthmore, Pennsylvania, United States

Abstract

Monolingual and bilingual speakers of the same languages derive unique meanings for words, partly based on between-language differences in meaning of dictionary-translated words. Do language models also capture this variability between speakers? We compared several models of lexical semantic representation and their correspondences to a word-word meaning similarity rating task done by both L1 and L2 English speakers. We found most language models do not differently correlate with L1 vs. L2 English speakers. Further, these models exhibit more cross-language similarity between Mandarin and English representations than is supported by psycholinguistic research. Only GloVe and OpenAI's Davinci models more strongly correlated with L1 speakers than L2, but individual participants' similarity to these models did not relate to language history variables that might otherwise predict bilingual lexical semantic native-likeness. We concluded that language models are not yet reliable references for tracking lexical semantic learning and discuss future directions for computational and psycholinguistics.