Measuring bilingual lexical semantics with word-embedding models

Jingyuan "Selena" She[1,2], Gabrielle Ma[1], & Benjamin Zinszer[1]

1. Swarthmore College, 2. Haverford College

Introduction

Bilingualism & lexical semantics

Second language learners gradually adapt L2 word meanings to be more similar to native-speakers' judgments of **word-word similarity**
(Dong, Gui, & MacWhinney, 2005)

… but bilinguals do not finally maintain two separate, native-like sets of word meanings in L1 and L2. Bilingual lexical semantics show **cross-language convergence**
(Ameel et al., 2005)

L2 word meanings follow a **years-long trajectory** shaped by **both L1 and L2** lexical semantic knowledge
(Zinszer et al., 2014)

Translation equivalence & inequivalence

Neural Translation: Word-word similarity measured in brain response patterns of native Chinese and English speakers are also *highly correlated between languages*
(Zinszer et al., 2016, *JoCN*; Yang et al., 2017)

This finding suggests more similarity between languages' neural representations than native speakers report in behavioral measures. However, unexplained variance in brain response patterns is *not* well-explained by these behavioral ratings. (Zinszer et al., 2016, *Cog Sci Society*)

Can word-embedding models explain semantic differences between L1 and L2?

Computational models of lexical semantics based on word embeddings have proven very powerful for explaining behavioral *and* neural data (e.g., Word2Vec; Mikelov et al., 2008). They may provide a useful model for quantifying translation inequivalence by explaining the differences in behavioral and brain data.

Question: Do word-embedding models like Word2Vec and GloVe correlate differently to speakers' ratings of word similarity based on their bilingual status & language history?

- Do English native speakers' and Chinese-English bilinguals' word-word ratings in English differ?
- Do our models' representations of meaning capture these differences in meaning?

Method

28 words clustered in six semantic categories, selected for:

1. Concrete, visualizable, & relatively familiar object names
2. Include complicated translation relationships between American English and Mandarin Chinese within the semantic category (ex: table & desk, chair & sofa, purse & bag)

Human ratings of semantic similarity

53 native English speakers and 22 Chinese-English bilinguals recruited from the PSYC 001, FindingFive, and Amazon Mturk subject pools.

Procedure:

1. Participants self-reported proficiency and history for each language they speak or understand
2. Participants rated similarity of meaning between two words (28 total words, 406 combinations):
   5 = same word
   1 = not similar at all

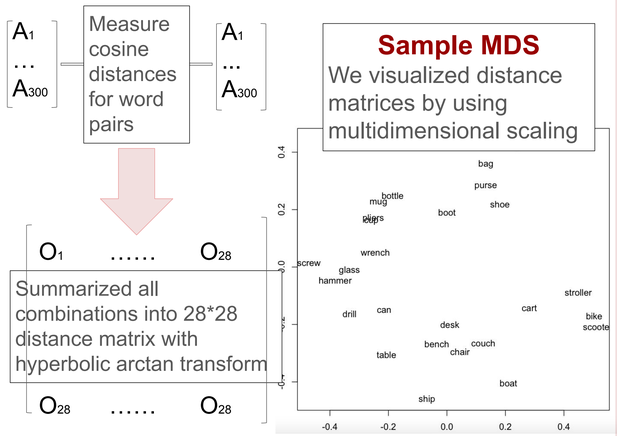
Excluded:

- Other bilingual participants
- Did not respond to >=80% of the word pairs

Analyses:
Took average score of each word pair, used mean ratings and hyperbolic arctan transformation to create distance matrix of word similarities

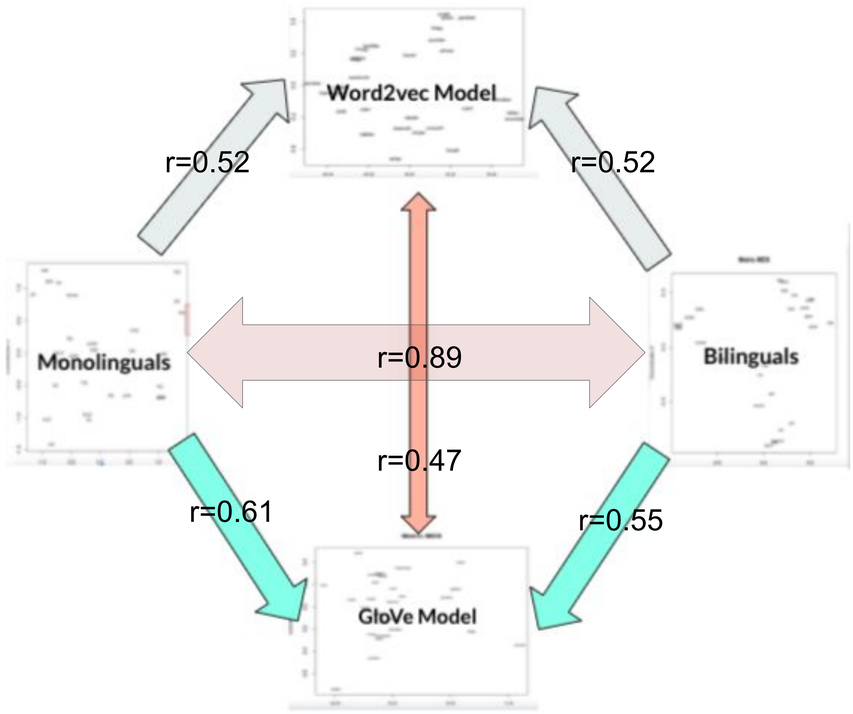
Word-embedding model of semantics

Word2Vec -> Google News pretrained English model,
GloVe -> 42B Common Crawl data.
Words are represented as 300d vectors in both models.



Sample MDS
We visualized distance matrices by using multidimensional scaling

Comparison of human ratings to the models (pairwise correlations)



Result

- L1 and L2 speakers both moderately correlated with both Word2Vec and GloVe models of meaning
- Chinese-English bilinguals were slightly less correlated to the GloVe model
- L1 and L2 speakers' ratings were *strongly* correlated
- Word2Vec and GloVe models are only moderately correlated with each other

Discussion

Human word-word similarity ratings

Based on previous studies, we expected participants' word-word similarity ratings to be only moderately correlated between L1 (native) speakers and L2 (Chinese-English bilingual) speakers of English.

Between-group correlation was **stronger than expected**

Word-embedding models

GloVe model was **moderately correlated with L1** English speakers, which was expected from high performance of this model in behavior and brain studies.

GloVe model was **moderately correlated with L2, but slightly less than L1.** This trend is consistent with previous findings about translation inequivalence, but a much smaller effect than expected.

No difference between L1 & L2 speakers' correlations with Word2Vec. This finding is unexpected!

Conclusion: Compared to Word2Vec, GloVe does a slightly better job capturing the differences in word similarity judgements between these groups, but neither model is a good representation for between-group lexical semantics.

Future directions

Human ratings: Finer-grained semantic differences using photographs of objects to estimate category boundaries.

Models: We will explore BERT, a newer embedding model with similar approaches and neural methods such as MVPA Analysis with EEG.

Acknowledgements

We are grateful to the students of the PSYC 001 course in Fall 2020 and Spring 2021 semesters for participating in this project.

We are grateful to the KINSC Summer Scholars Program and Swarthmore College Department of Psychology for support.

References

Ameel, E., Storms, G., Malt, B. C., & Sloman, S. A. (2005). How bilinguals solve the naming problem. *Journal of Memory and Language, 53*(1), 60-80.
Dong, Y., Gui, S., & MacWhinney, B. (2005). Shared and separate meanings in the bilingual mental lexicon. *Bilingualism, 8*(3), 221.
Yang, Y., Wang, J., Bailer, C., Cherkassky, V., & Just, M. A. (2017). Commonality of neural representations of sentences across languages: predicting brain activation during Portuguese sentence comprehension using an English-based model of brain function. *NeuroImage, 146*, 658-666.
Zinszer, B., Anderson, A. J., & Raizada, R. D. (2016). Chinese and English speakers' neural representations of word meaning offer a different picture of cross-language semantics than corpus and behavioral measures. In *Proceedings of the Cognitive Science Society*.
Zinszer, B. D., Anderson, A. J., Kang, O., Wheatley, T., & Raizada, R. D. (2016). Semantic structural alignment of neural representational spaces enables translation between English and Chinese words. *Journal of Cognitive Neuroscience, 28*(11), 1749-1759.
Zinszer, B. D., Malt, B. C., Ameel, E., & Li, P. (2014). Native-likeness in second language lexical categorization reflects individual language history and linguistic community norms. *Frontiers in Psychology, 5*, 1203.