

THE EFFECTS OF DIACHRONIC LANGUAGE CHANGE ON SEMANTIC ORGANIZATION ACROSS THE LIFESPAN

Ellis Cain (ecain@ucmerced.edu) & Rachel Ryskin
University of California, Merced

ABSTRACT #166

BACKGROUND

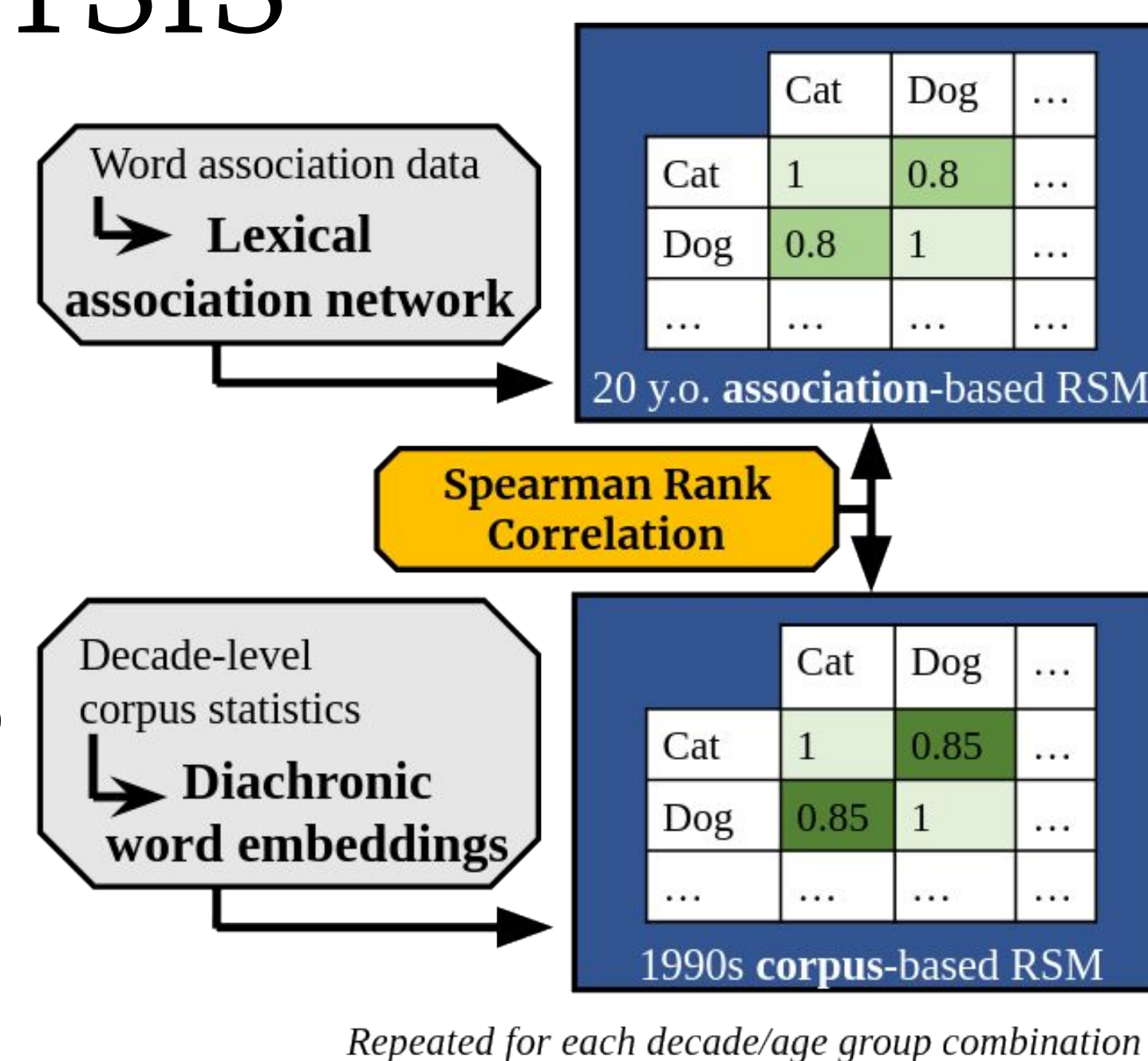
- **Organization of semantic space** reflects an integration of sensory-motor information, affective experience, and linguistic information (Vigliocco et al., 2009)
- **Distributional statistics** of word usage predicts similarity judgements from humans (Hill et al., 2015)
- Language usage and word meanings **change over time** (Michel et al., 2011; Hamilton et al., 2016)
- Organization of lexical association networks have been found to **change with age** (Dubossarsky et al., 2017)

QUESTION

How do diachronic changes in language statistics relate to changes in semantic organization across the lifespan?

REPRESENTATIONAL SIMILARITY ANALYSIS

- RSA allows for comparison across data formats (Kriegeskorte et al., 2008)
- Representational Similarity Matrices (RSM) “represent” a specific data source, where each cell is the **similarity of two words**



Results

- A. Corpus-based RSMs were **most similar for adjacent decades**, but this **decreased as time between decades increased**
- B. Association-based RSMs had **lower correlation**, and had low internal reliability as well
- C. All of the association-based RSMs were **most similar to the latest decade** of corpus-based RSM

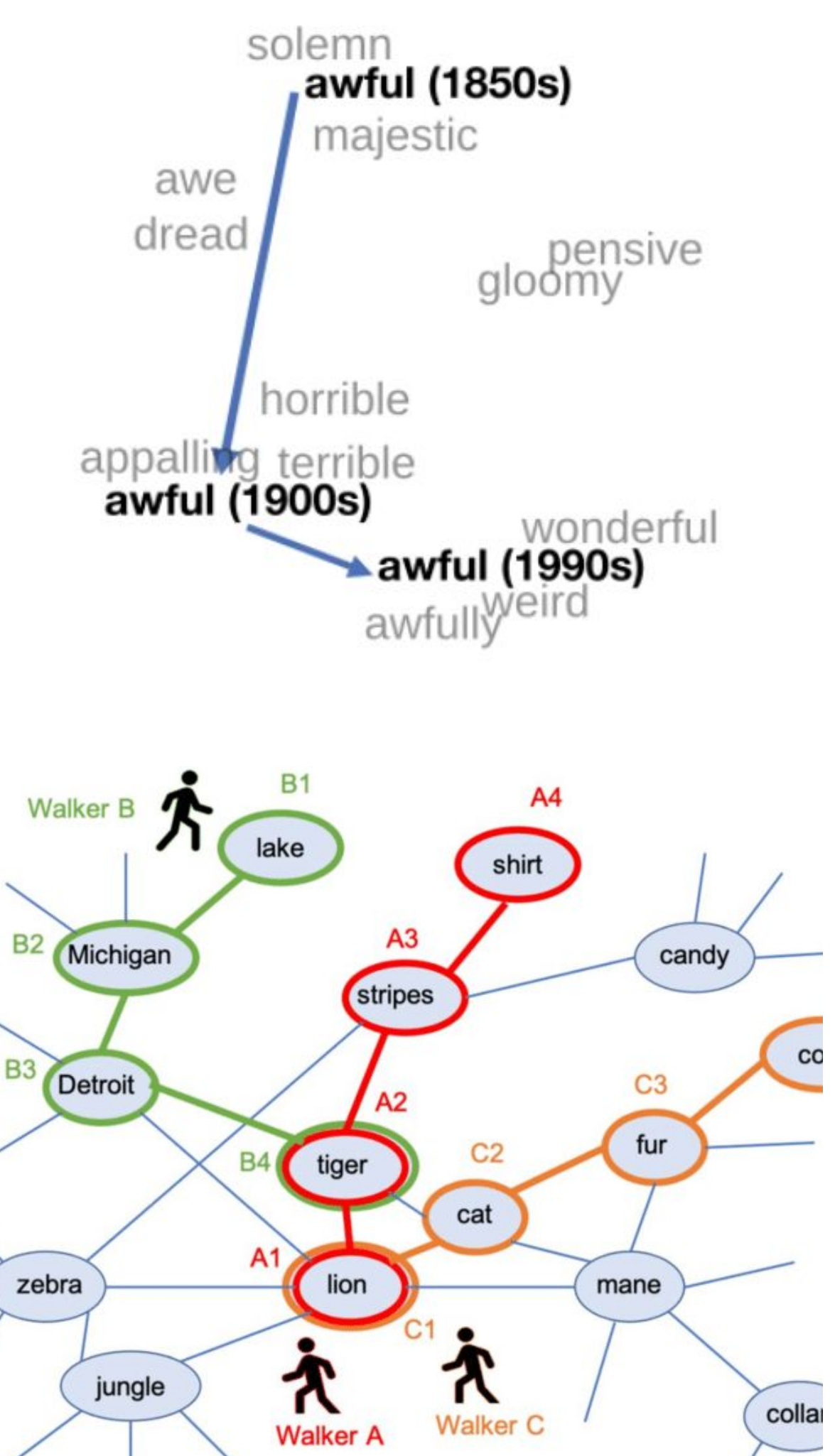
DATASETS

Corpus-based:

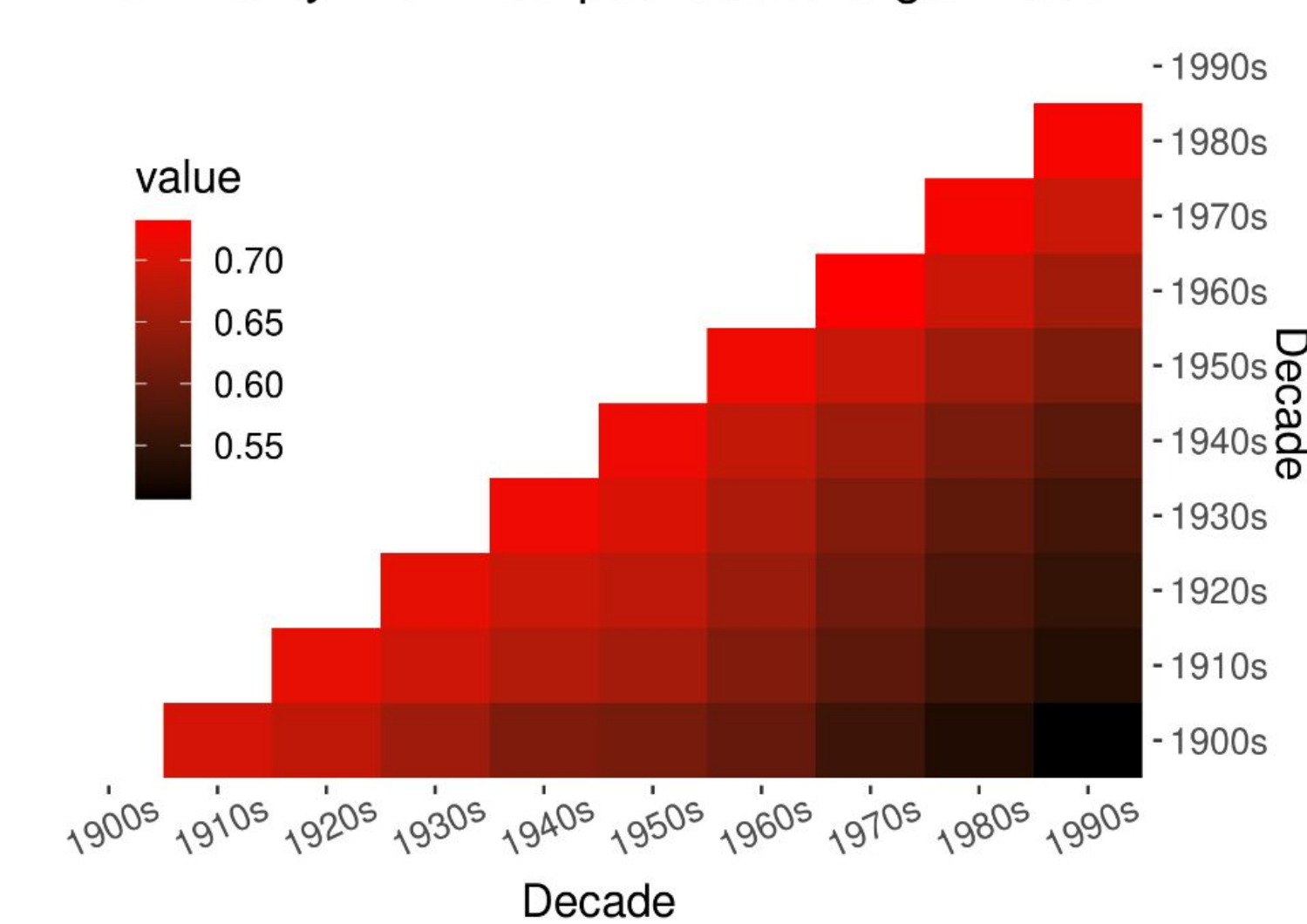
- *Histwords* (Hamilton et al., 2016)
- **Diachronic word embeddings** trained on decade-level subsets of the Google N-grams corpus
- Similarity evaluated as normalized dot-product

Association-based:

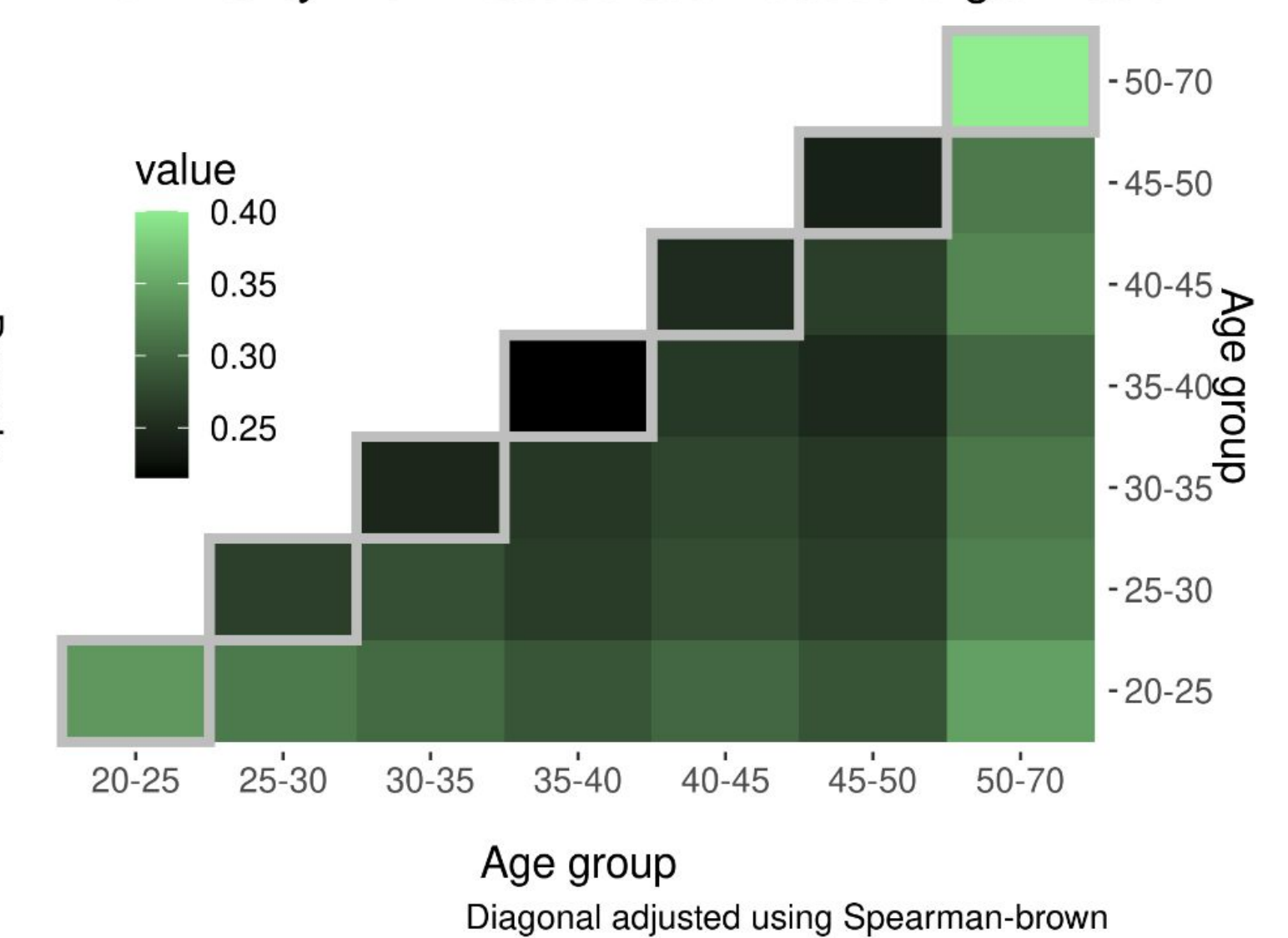
- *Small World of Words* (De Deyne et al., 2019)
- **Word association task**
- Responses used to generate **association networks**
- 88,722 participants (M = 36 yo, SD = 16 yo, female = 38%)
- Similarity evaluated as overlap of Katz random walks over association networks



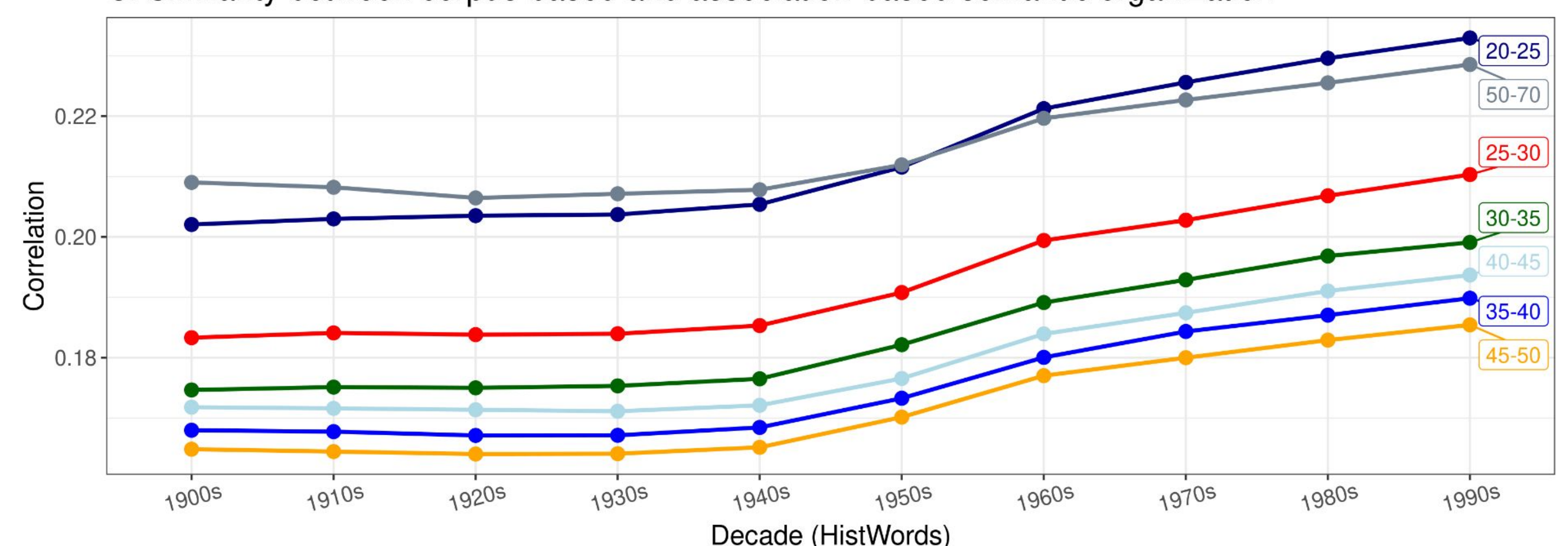
A. Similarity within corpus-based organization



B. Similarity within association-based organization

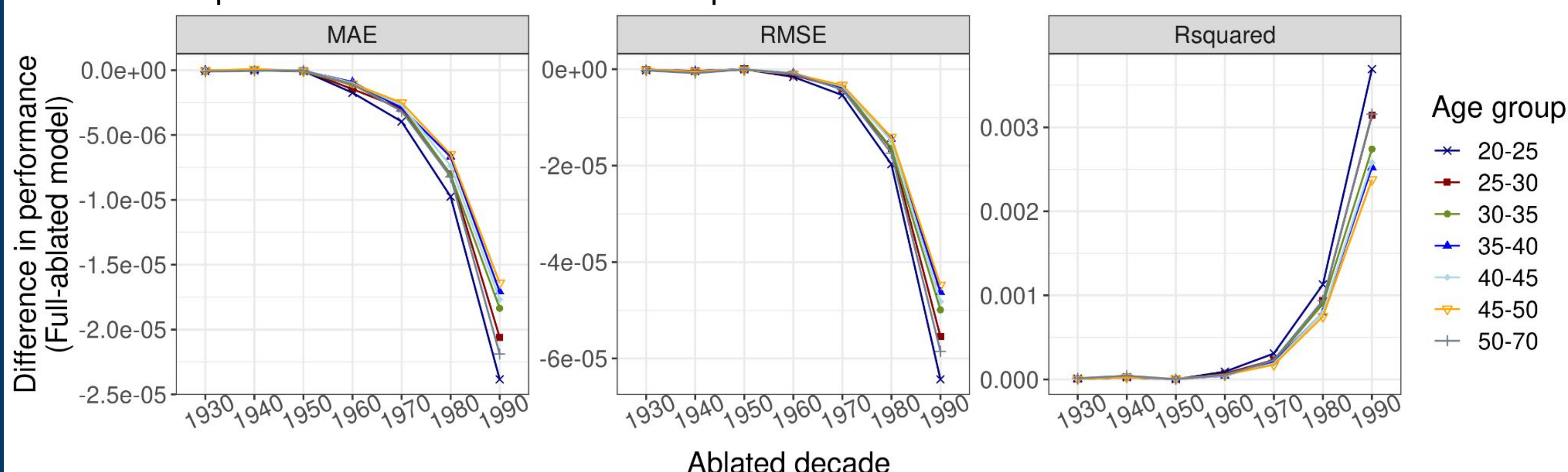


C. Similarity between corpus-based and association-based semantic organization



QUANTIFYING THE IMPACT OF DECADE-LEVEL ORGANIZATIONS

Impact of ablation on linear model performance



- Trained a linear model (5-fold CV) to **predict** each age cohort's RSM value as a **combination** of the corpus-based decade-level RSM values
- Ablated decades from the model to **quantify the impact of that decade**
- We compared the average difference in MAE, RMSE, R^2 between the full and ablated models

Results

- **More recent decades** have the **largest impact** on model performance, and **this effect is largest for the youngest cohort**

TAKEAWAYS

- The corpus-based semantic similarity spaces changed gradually over time, while there were no clear patterns regarding association-based similarity across the different age cohorts
- All age cohorts' semantic organizations were most similar to the 1990s corpus-based semantic organization, which is likely because it corresponds most closely to the linguistic environment of the participants' everyday life
- **Diachronic differences in the distributional properties of language usage seem to relate to changes in semantic organization, but the mechanism of meaning integration over time is still unclear (i.e., statistical learning with even weighting over the lifespan vs specific age range with more weight)**
- Stimuli in association dataset may be skewed towards words with unchanged meaning, due to frequency based selection
- Plan to run a similarity judgement study to supplement the analyses and address this limitation