Modeling agreement attraction effects in vector space

Eva Neu[1], Maayan Keshev[2] and Brian Dillon[1]
[1]UMass Amherst, [2]The Hebrew University of Jerusalem
HSP2025

Our project in a nutshell

- Agreement attraction is sensitive to syntactic similarity.
- Retrieving the target of agreement can be modeled as recovering a vector representation of the lexical item bound to a vector representation of the subject position. Item vectors bound to similar positions are more likely to be misretrieved.
- We propose a method for computing position vectors such that higher cosine similarity between positions corresponds to higher rates of interference.

Agreement attraction

- In agreement attraction, the verb agrees not with the target of agreement but with an intervening distractor (e.g., Bock & Miller, 1991; Wagers et al., 2009):

(1) The key to the cabinets was/*were rusty.

- Interference rates are modulated by (among other factors) syntactic similarity (Arnett & Wagers, 2017; Franck et al., 2002; Van Dye, 2007; Van Dyke & McElree, 2011):

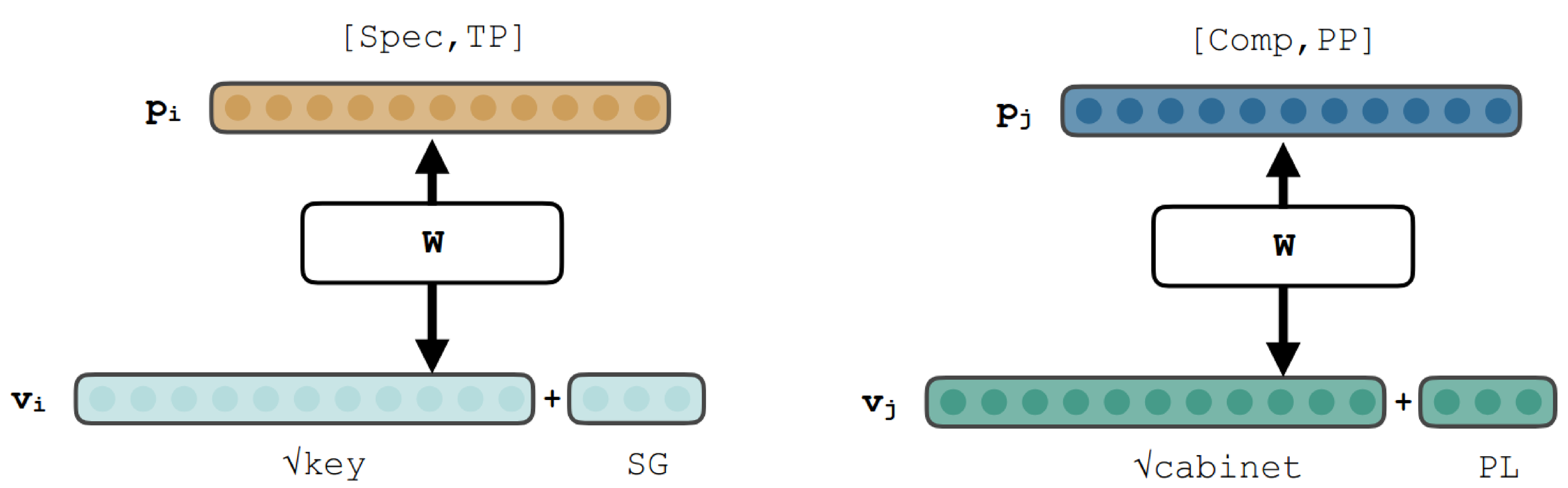
(2) The helicopter for the flight(s) over the canyon(s) is/*are rusty.

- More agreement from structurally closer distractor 'flights' than from 'canyons' – why?

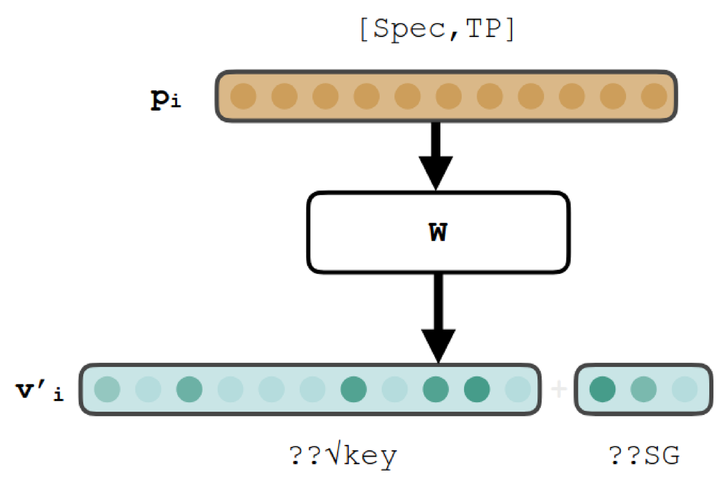
Vector-symbolic representation

- Long tradition of representing syntactic positions in vector space (e.g., Cho et al., 2020; Piantadosi et al., 2024; Plate, 1997; Smolensky, 1990; Smolensky et al., 2010)
- Our approach: sentences are encoded in working memory by binding lexical items to syntactic positions by means of a weight matrix (Keshev et al., 2024a)

(3) The key to the cabinets...



- Processing agreement means retrieving the item vector bound to subject position:

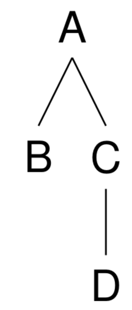


- Effect of distributed vector representation:
  - Single connection matrix, same units are activated to code different positions/items
  - Unless position vectors are orthogonal, item vector is not perfectly recovered
  - Items is in more similar positions are more likely to be misretrieved
- How can we systematically compute such position vectors?

Computing position vectors

Algorithm

- Step 1: Assign a constituency parse to English input sentence
- Step 2: Assign each node a base vector depending on its category
- Step 3: For each node, compute its position vector as the weighted sum of its own base vector and its mother's position vector (cf. TCM; Howard & Kahana, 2002)



Formula

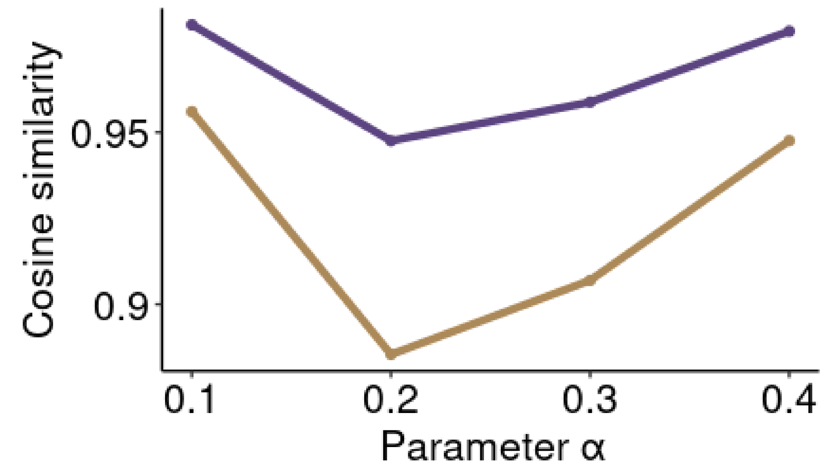
position vector (x) = α × base vector (x) + (1 α) × position vector (x's mother)

- Upshot: Position vectors contain category information of all dominating nodes such that more distant nodes make up a smaller part of the representation

Results: Cosine similarities

(4) The helicopter for the **flight** over the **canyon** was rusty.

- Structurally closer distractor ('flight') more similar to the target regardless of α
- Correctly predicts higher interference rates from 'flight' than from 'canyon'



Predicting error rates

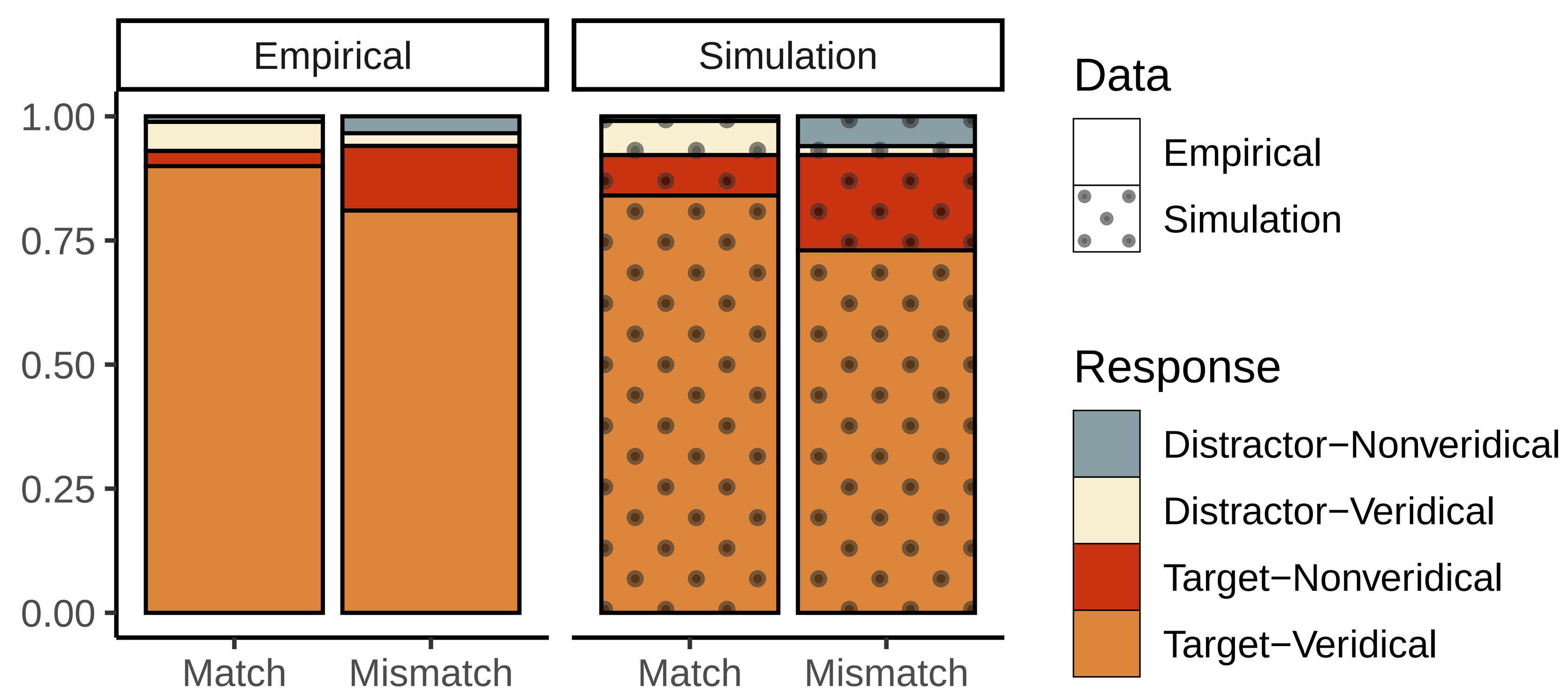
Strategy for turning cosine similarities into error rates

- Fit the model to empirical data for a single distractor using maximum likelihood estimation,
- Test the predictions of fitted model against held-out data for two distractors in different syntactic positions

Single distractor

- Empirical data: 4-AFC task for sentences with one distractor, either singular or plural (Keshev et al., 2024b)

(5) The apprentice of the chef/chefs worked diligently.
Who worked diligently?
The apprentice / the apprentices / the chef / the chefs

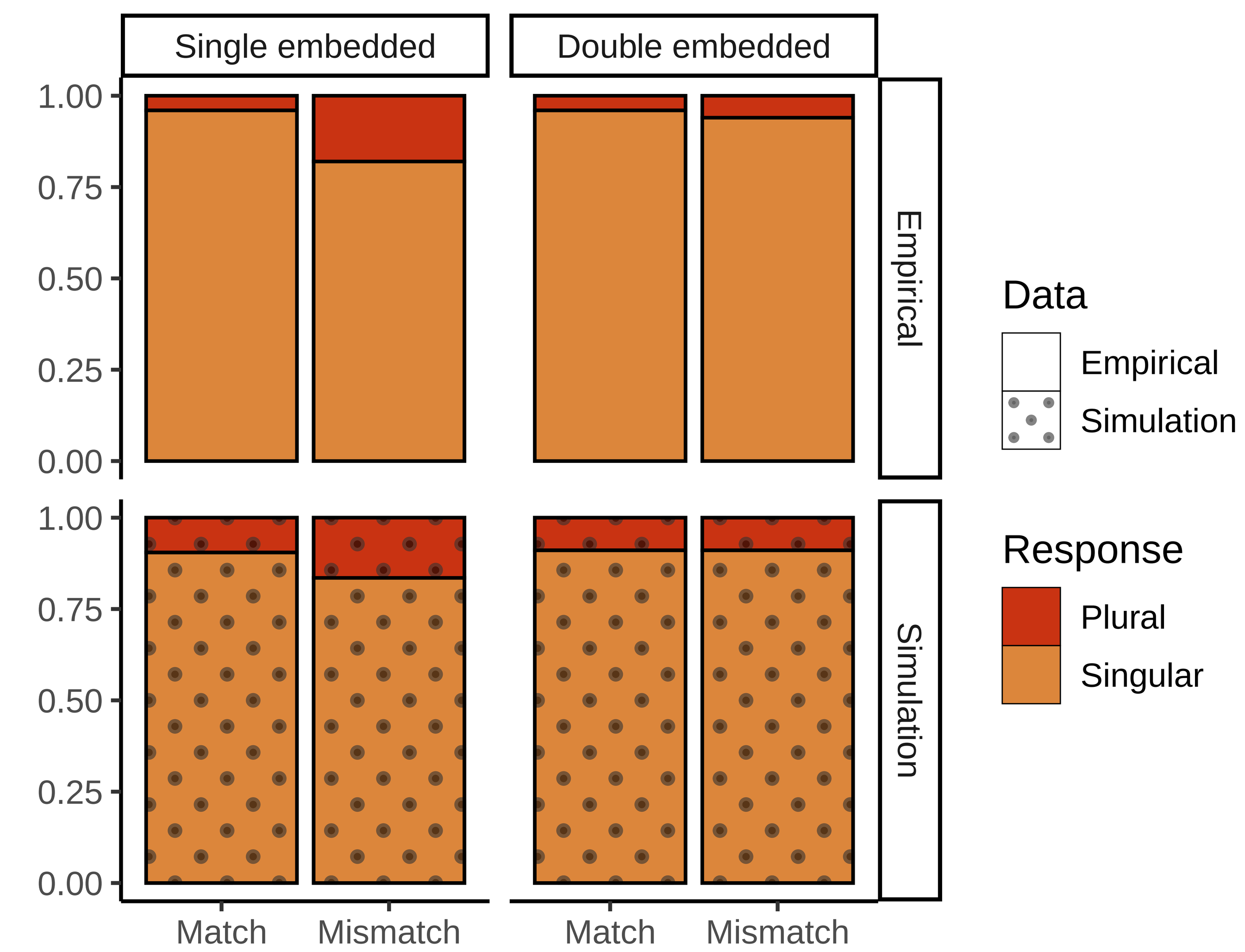


Predicting error rates (cont.)

Two distractors

- Empirical data: 2-AFC task for sentences with two distractors, one of which is plural (Keung & Staub, 2018)

(6) The helicopter for the flight(s) over the canyon(s)... is?/are?



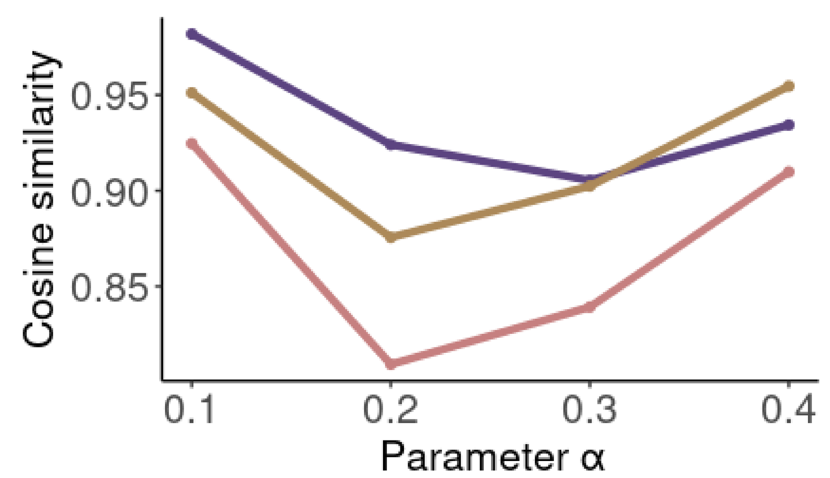
Discussion

Differences to percolation theories

- Cosine similarity is symmetric – predicts upwards and downwards agreement attraction
- Similarity measure does not reduce to node distance:

(7) The squirrel had a **nightmare** that the **chipmunk** ate the **nut**.

- 'Nightmare' is structurally closer but 'chipmunk' also has subject status
- For high α, the embedded subject is more similar to the target



Limitations and challenges

- No perfect quantitative fit to the test data
- Model predicts high cross-clausal similarities – should we implement clausal boundaries?

Possible extensions and modifications

- Dependency parse
- Non-orthogonal base vectors

Selected references and acknowledgments

Bock & Miller. (1991). Broken agreement. Franck et al. (2002). Attraction in sentence production: The role of syntactic structure. Keshev et al. (2024a). A working memory model of sentence processing as binding morphemes to syntactic positions. Keshev et al. (2024b). Feature distortion and memory updating. Keung & Staub. (2018). Variable agreement with coordinate subjects is not a form of agreement attraction. Many thanks to Samuel Amouyal, Mandy Cartner, Niki Koesterich, Aya Meltzer-Asscher and Stephanie Rich. This research was supported by the National Science Foundation (NSF BCS-1941485 to BD) and the Binational Science Foundation (NSF-BSF 2146798 to BD).

eneu@umass.edu
github.com/Eva-Neu/syntactic_vectors