Language is only sparsely compositional

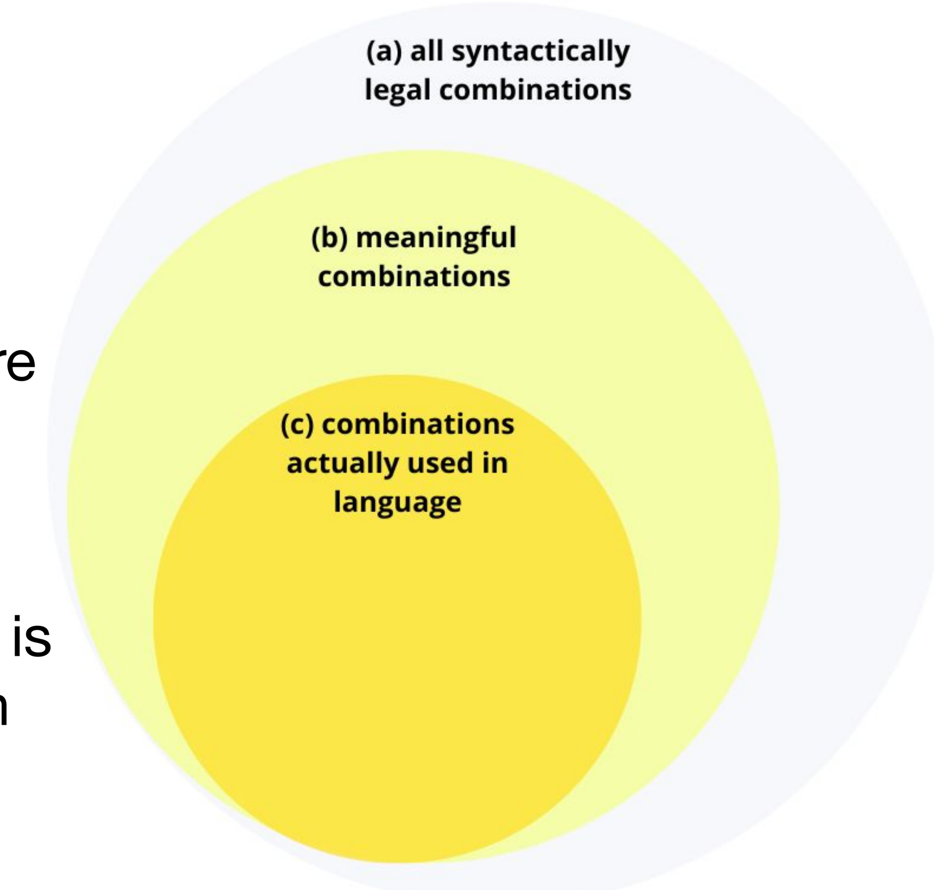
Aalok **Sathe**[1,2], George **Flint**[3], Evelina **Fedorenko**[1]*, Noga **Zaslavsky**[4]*
[1]MIT [2]Brown University [3]UC Berkeley [4]NYU *equal senior contribution

asathe@mit.edu
aloxatel

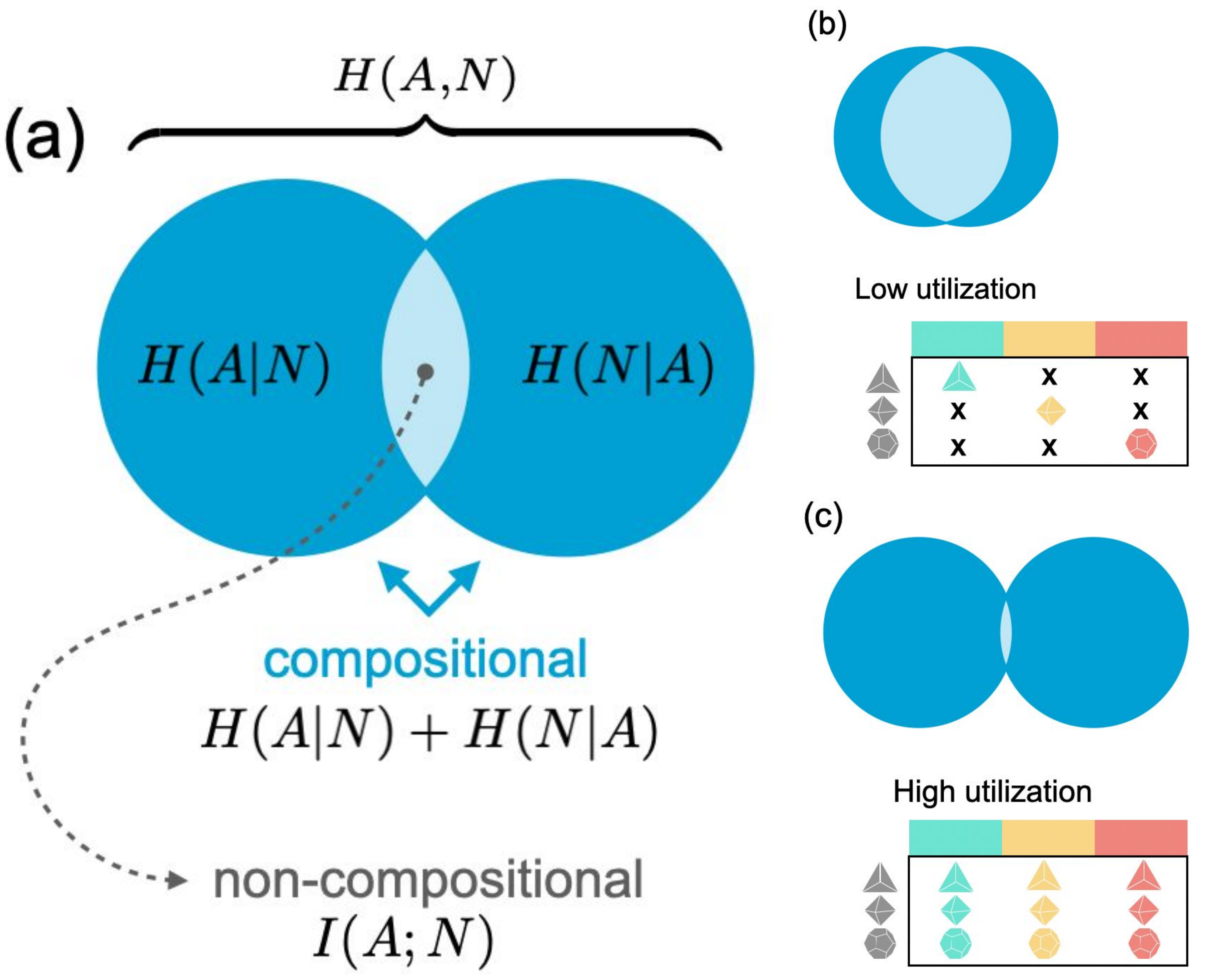
brain+cognitive sciences
COGSCI 2024 DYNAMICS OF COGNITION July 24-27 • Rotterdam

Introduction

- **Compositionality** is a key hallmark of language. However, most prior literature studies compositionality of phrases.
- **We ask**: how compositional is the space of Adjective-Noun combinations?



An information-theoretic measure for system-level utilized compositionality



Each circle corresponds to the entropy of the variables that can be composed, demonstrated here in the case of adjectives, *A*, and nouns, *N*. The colored area, corresponding to the joint entropy, consists of three parts: ***H(A|N)*** + ***H(N|A)*** + ***I(A;N)***. The light region reflects the non-compositional subspace, measured by *I(A;N)*, and the complementing darker region reflects the compositional subspace.

$$\mathrm{UC} = \frac{H(A,N) - I(A;N)}{H(A)+H(N)} = \frac{H(A)+H(N)-2I(A;N)}{H(A)+H(N)} = 1 - 2\frac{I(A;N)}{H(A)+H(N)}$$

Study 1: Corpus analyses

Materials

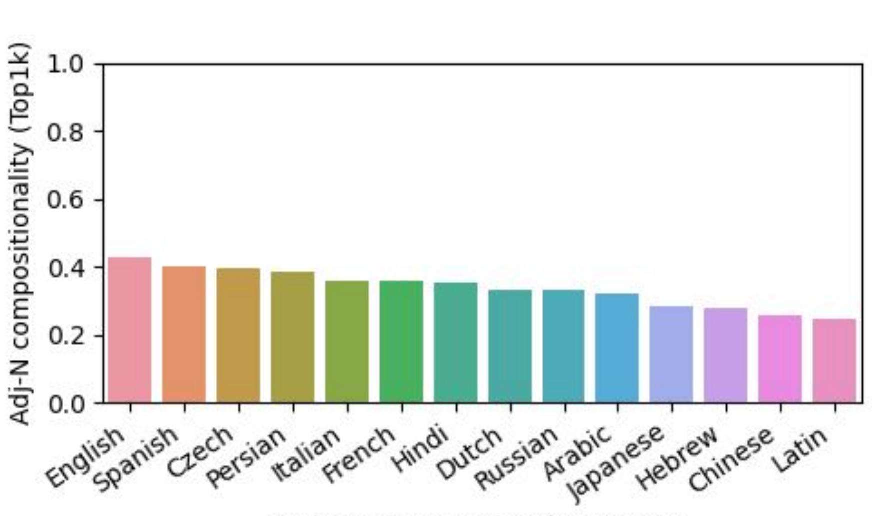
- Corpus of Contemporary American English (COCA): 385M+ words spanning 1991-2012; Universal Dependencies corpora for 14 languages

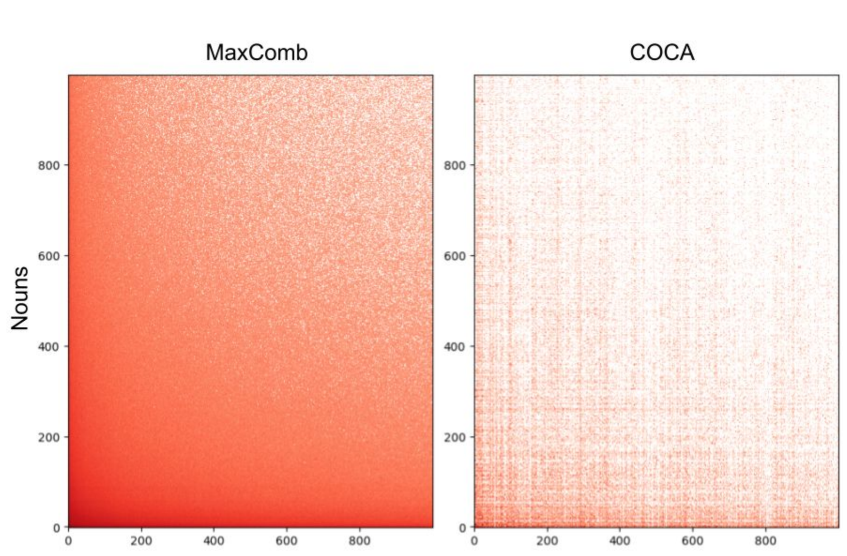
Methods

- Dependency-parse corpus; extract *amod* relations
- MaxComb: simulate maximal possible combinations respecting marginal *P(A), P(N)* as observed in COCA

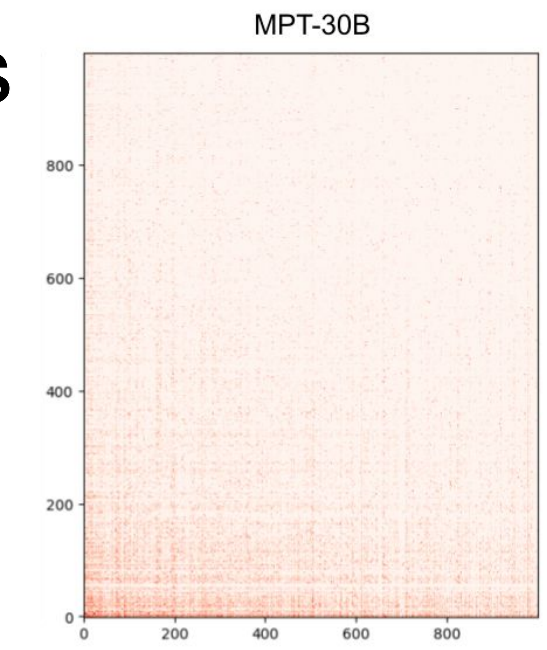
Results

- Across languages and corpora the utilized compositionality is substantially below the theoretical maximum of 1.





Study 2: Large language models



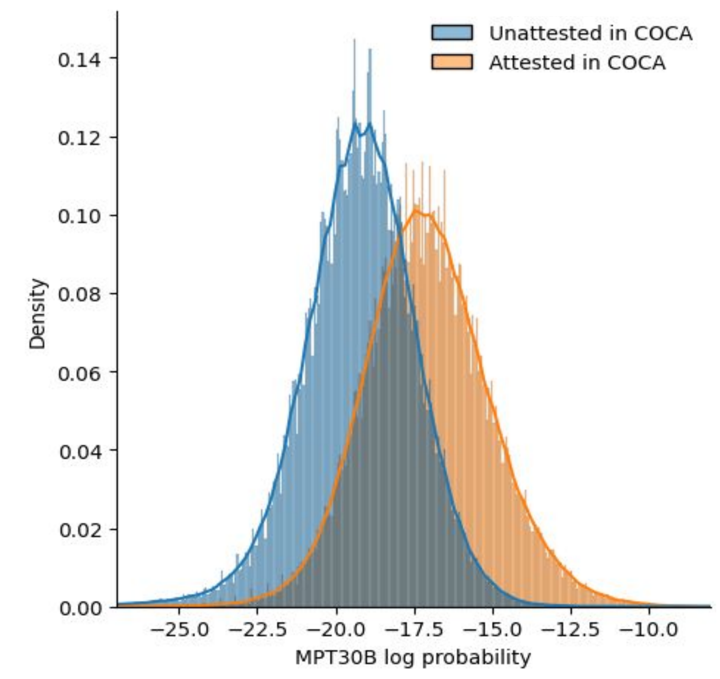
LLMs provide another window into language use and are exposed to much larger amounts of language

Methods

- We evaluated models ranging from 1.5B to 30B
- **P(N | "How much sense does this make: A")** Shorthand: $P_{LLM}(N|A)$
- Adjust using marginal $P_{COCA}(A)$ from COCA
- Evaluate on Top1k items (1M pairs) from COCA

Results

- LLM-probabilities over Adj-N pairs also indicate high sparsity
- LLM probabilities are correlated with corpus frequency, but paint a slightly different picture



Study 3: Human judgments

Is the sparsity we observe in Studies 1 & 2 a consequence of finite-sample effect?

Methods

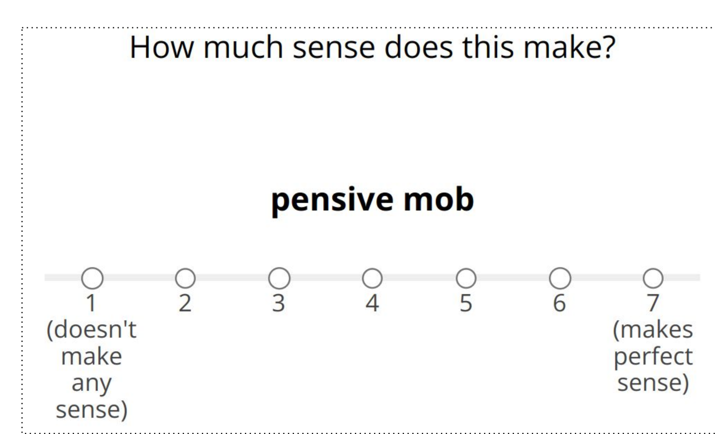
- Each participant was asked to rate 200 randomly-drawn adjective-noun pairs on a scale from 1 (doesn't make any sense) to 7 (makes perfect sense). Each pair received at least 8 ratings.

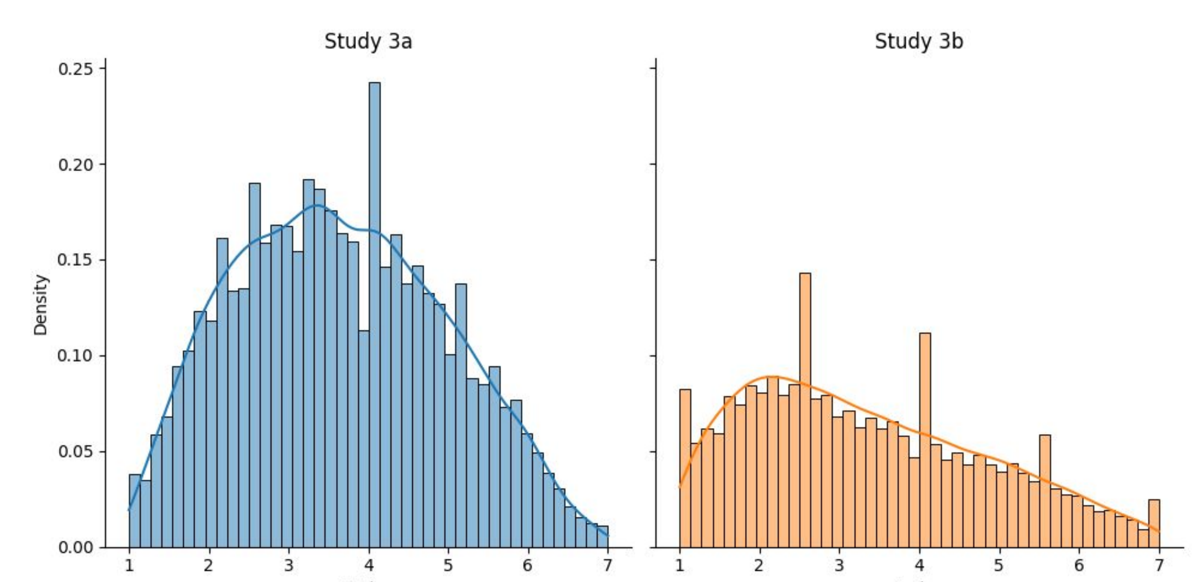
Study 3a: revisiting Vecchi et al. (2017)

- We used a random subset of 20,000 items chosen from the 27,000 items used by Vecchi et al (2017). The original set of items were designed to be unattested.
- *n*=1199 participants on Prolific (ages 18 to 85, mean: 34.1 ± 12.5)

Results

- Most unattested pairs are rated very low; finite-sample effect does not seem to be the main driving factor for sparsity



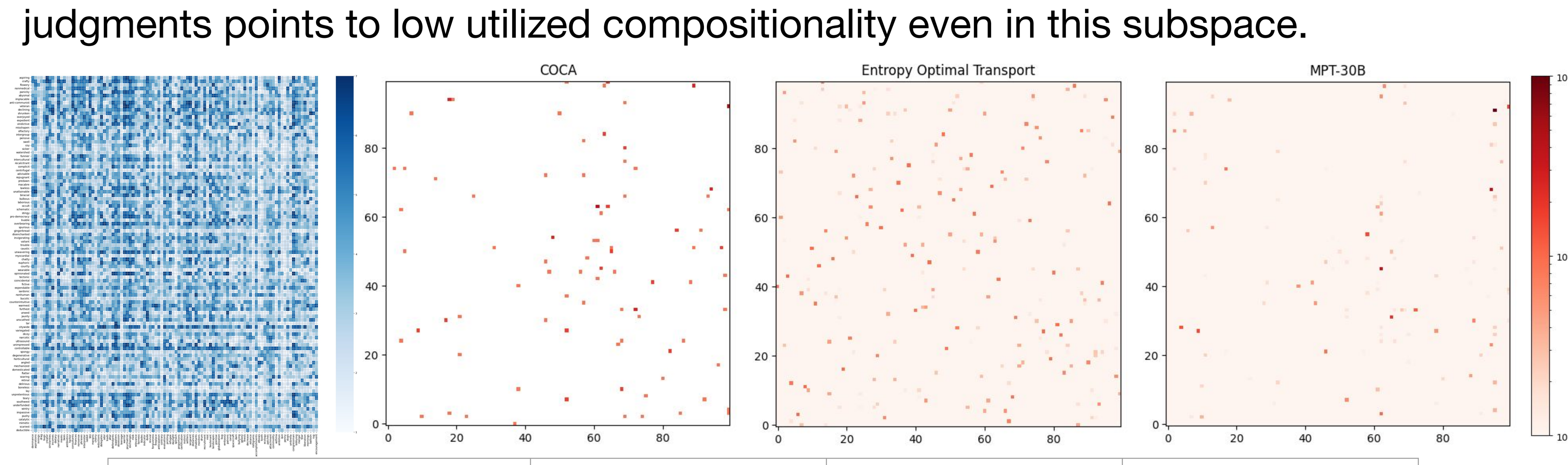


Study 3b: A challenging, measurable subdomain

- Top 4001–4100 Adj and N from COCA. We picked this range in order to sample at the edge of sensible and nonsensical.
- *n*=431 participants on Prolific (ages 18 to 81, mean: 38 ± 12.3)
- We obtained ratings on 100x100 = 10,000 combinations. Human judgments, however, do not form a probability distribution. We recover a joint distribution using Optimal Transport.
- We induce **γ ~ P(A,N)** subject to a cost function that maximizes alignment with human ratings, and is regularized to maximize joint entropy, respecting marginals.

Results

- Most items are rated nonsensical. A probability distribution recovered from judgments points to low utilized compositionality even in this subspace.



Source	COCA	EOT	MPT-30B
Utilized Compositionality	0.147	0.21	0.577

Conclusion

- Corpora and LLMs suggest low utilized compositionality in the Adj-N combination space
- To rule out finite-sample effect, we collected human judgments and found similar estimates
- Whereas a large number of combinations are possible, only a very small number is used in typical language

[1] Vecchi, E.M., Marelli, M., Zamparelli, R., & Baroni, M. (2017). Spicy adjectives and nominal donkeys: Capturing semantic deviance using compositionality in distributional spaces. Cognitive Science.