Additivity biases and classes of sampling algorithms for word predictability

Yutong Zhang and E. Matthew Husband
Faculty of Linguistics, Philology, and Phonetics, University of Oxford

FACULTY OF LINGUISTICS, PHILOLOGY AND PHONETICS

Word Predictability as Bayesian Inference

Language comprehension makes use of **word predictability**.

- E.g., *Jordan learned a lot about cars from his ...*
- The difficulty of comprehending a word is related to its predictability in context. (more probable to come next ➔ easier to process)

Comprehenders may be Bayesian about word predictability.

- Predictability reflects **Bayesian updating** of a prior probability distribution over lexical items, raising or lowering the probability of any lexical item given input [1].

However, exact Bayesian inference faces **resource limitations** [2].

- Calculating Bayes' Rule precisely can be computationally intractable, especially for large and complex lexical spaces.

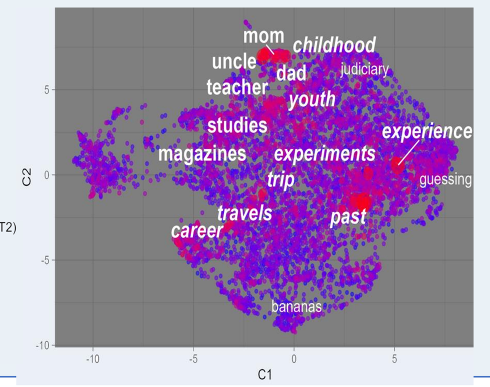
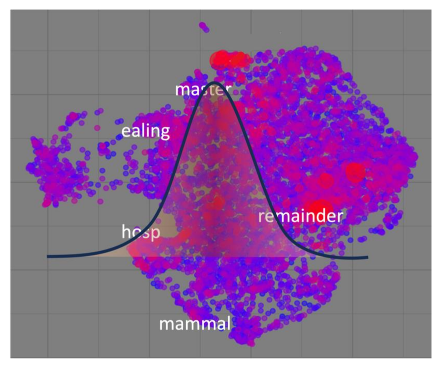
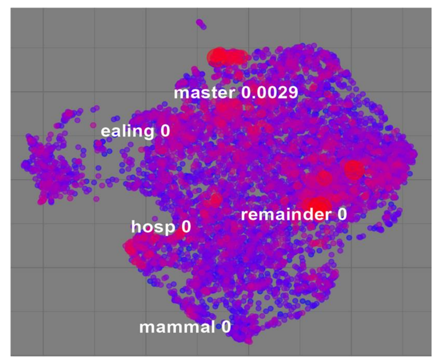
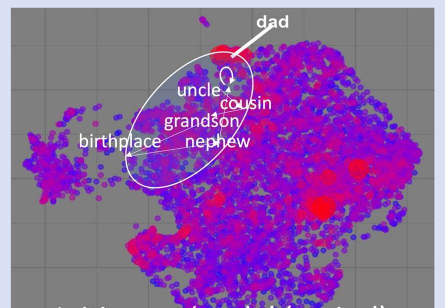
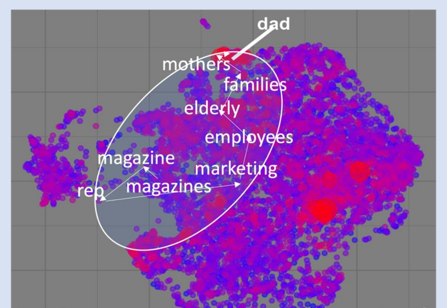
$$\underbrace{P(h|d)}_{\text{posterior probability}} = \frac{\overbrace{P(d|h)}^{\text{likelihood}} \cdot \overbrace{P(h)}^{\text{prior probability}}}{P(d)} = \frac{P(d|h)\cdot P(h)}{\underbrace{\sum_{h'} P(d|h')\cdot P(h')}_{\text{sum over all possible words}}}$$

a particular word h being the upcoming word given the observed context d

Question:

- How does language comprehension make use of Bayesian inference for word predictability under resource limitations?
  - **Sampling** approaches are proposed to approximate posterior distributions [3][4].
- Further, which **sampling algorithm** is deployed by comprehenders in this process?

Sampling Algorithms and Systematic Biases

Sampling algorithm	Procedure	Systematic Bias (with a limited number of samples)
Direct Sampling (requires global knowledge)	Sample in parallel from the posterior distribution by probability.	Stochastic behavior 
Importance Sampling (requires approximate global knowledge)	Sample in parallel from a convenient proposal distribution; Weight samples by the actual posterior probability.	Subadditivity (**Over-represents** typical samples, because samples with a higher posterior probability receive more weight.)  
Markov Chain Monte Carlo (MCMC) Sampling (requires local knowledge)	Initiate a chain with a sample; Propose a new sample, and then accept / reject that sample given its relative probability to the current sample in the chain.	Subadditivity and Superadditivity (**Over-represents /under-represent** typical examples when initiated with a typical/atypical example)  Initiate at 'uncle' (typical)  Initiate at 'magazines' (atypical)

Additivity Biases in Probability Judgment

Unpacking context impacts probability judgements [5].

Condition	Question
Packed	*I see a table. What is the probability that I also see an object starting with the letter C?*
Typical Unpacked	*I see a table. What is the probability that I also see a <u>chair, computer, curtain</u>, or any other object starting with the letter C?*
Atypical Unpacked	*I see a table. What is the probability that I also see a <u>cannon, cow, canoe</u>, or any other object starting with the letter C?*

- **Subadditivity**: unpacking to **typical** examples ➔ **increase** in probability judgment.
- **Superadditivity**: unpacking to **atypical** examples ➔ **decrease** in probability judgment.

★Question:

- Does language comprehension **show subadditivity and/or superadditivity** when unpacking contexts to examples?
- If so, do these biases diagnose **which class of sampling algorithms** is a more appropriate fit for word predictability?

Hypotheses & Predictions

Main Hypothesis: Unpacking context may impact word predictability and lead to systematic biases, which can be used to determine which class of sampling algorithms is deployed in language comprehension [5].

Bias	Direct Sampling	Importance Sampling	MCMC Sampling
Subadditivity		✓	✓
Superadditivity			✓

- If human language comprehension deploys direct sampling, there would be no bias.
- If deploying importance sampling, it would display only subadditivity.
- If deploying MCMC, it would display both subadditivity and superadditivity.

Predictions:

- This study aimed to explore how unpacking contexts to examples affects the **predictability of target words**.
- As reading times (RTs) are inversely related to word predictability [6], it was designed to measure **RTs to target words** in packed and unpacked conditions.

Bias	Unpacking Example	Word Predictability	Reading Time Prediction
Subadditivity	Typical	Increase	Faster than packed baseline
Superadditivity	Atypical	Decrease	Slower than packed baseline

Example Item

Condition	Sentence
Packed	Jordan learned a lot about cars from his **dad** during the last summer vacation. (target word: 0.45 cloze)
Typical Unpacked	Jordan learned a lot about cars from his <u>uncle</u> and his **dad** during the last summer vacation. (typical unpacking: 0.08 cloze)
Atypical Unpacked	Jordan learned a lot about cars from his <u>magazines</u> and his **dad** during the last summer vacation. (atypical unpacking: 0.01 cloze)

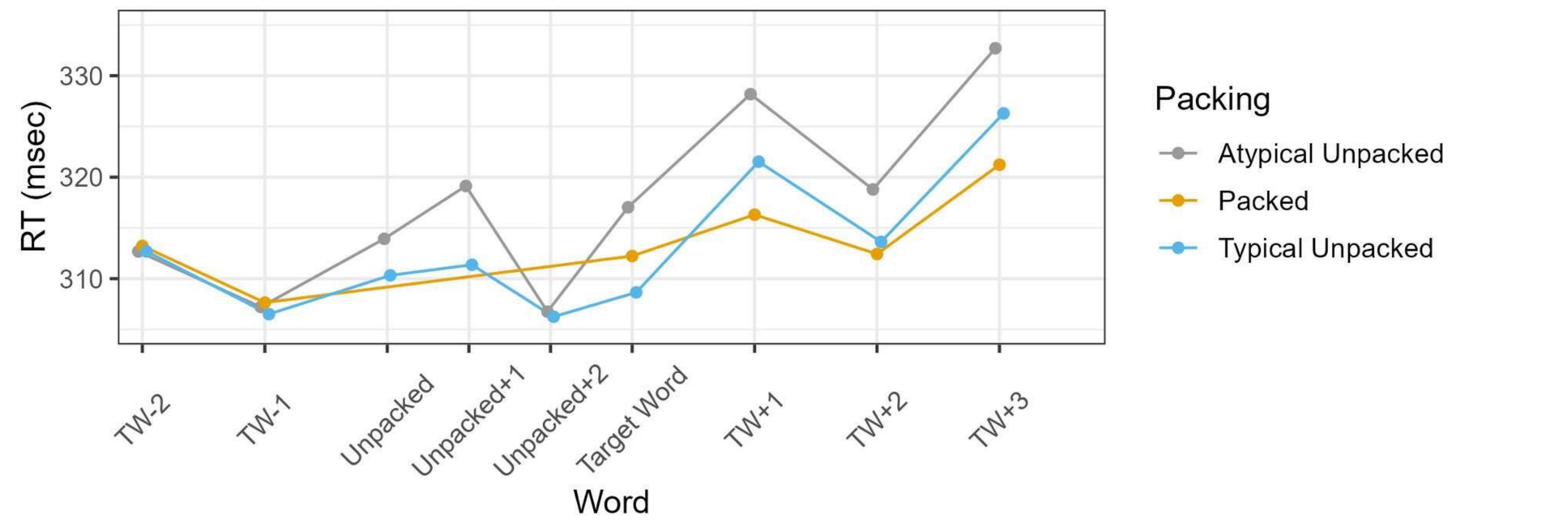
Experiment: Self-paced Reading

Design

- Written and hosted on PCIbex Farm.
- 60 Prolific-recruited native English speaker, 120 experimental items with 3 conditions.
- Contexts, target words and unpacking examples were pulled from [7]:
  - Target words: cloze range from 0.30 to 0.70
  - Typical unpacking examples: high cloze words
  - Atypical unpacking examples: low cloze words
- Sentence unfolds word by word.

Results

Model Estimated Reading Times (msec)



Linear Mixed Effects Model on Target Word

	Est	SE	t	Pr (> |t|)	
(Intercept)	312.64	12.99	24.07	<2e-16***	
Typical Unpacked	-7.98	3.17	-2.52	.014*	→ Subadditivity
Atypical Unpacked	8.80	3.25	2.70	.008**	→ Superadditivity

Conclusion & Discussion

- Comprehenders displayed **both subadditivity and superadditivity** for word predictability while reading sentences.
  - ➢ Consistent with the hypothesis that comprehenders deploy an **MCMC-class sampling algorithm** when predicting upcoming lexical items in sentence contexts under resource and time constraints.

General Discussion

The results are not simply about priming or surprisal:

- Word2Vec similarity between unpacking examples and target words did not interact with Typicality and did not improve model fit ($\chi2 = 1.203$; $p = .273$).
- GPT-2 surprisal of target word did not interact with Typicality and did not improve model fit ($\chi2 = 2.760$; $p = .252$).
- Typicality remained significant with Word2Vec/Surprisal as a covariate.

The biases we see during human language comprehension indicate:

- If sampling algorithms are appropriate models for cognition, **MCMC Sampling provides a better fit for word predictability** than Direct or Importance Sampling.
  - Only requires local knowledge
  - Makes appropriate use of prior context to directly anchor a sample chain
  - Computational efficient, even biologically plausible

MCMC is a general class of sampling algorithms, how these MCMC sampling algorithms are employed in language comprehension remains to be investigated.

Selected Reference

[1] Kuperberg, G. R., & Jaeger, T. F. (2015). What do we mean by prediction in language comprehension? Language, Cognition and Neuroscience, 31(1), 32–59. [2] Kwisthout, J., van Rooij, I. (2020). Computational Resource Demands of a Predictive Bayesian Brain. Computational Brain & Behavior, 3, 174–188. [3] Lévy, R., Reali, F., & Griffiths, T. L. (2008). Modeling the effects of memory on human online sentence processing with particle filters. Neural Information Processing Systems, 21, 937–944. [4] Hoover, J. L., Sonderegger, M., Piantadosi, S. T., & O'Donnell, T. J. (2023). The plausibility of sampling as an algorithmic theory of sentence processing. Open Mind, 1–42. [5] Dasgupta, I., Schulz, E., & Gershman, S. J. (2017). Where do hypotheses come from? Cognitive Psychology, 96, 1–25. [6] Staub, A. (2015). The Effect of lexical predictability on eye movements in Reading: Critical review and Theoretical interpretation. Language and Linguistics Compass, 9(8), 311–327. [7] Peelle, J. E., Miller, R. H., Rogers, C. S., Spehar, B., Sommers, M. S., & Van Engen, K. J. (2020). Completion norms for 3085 English sentence contexts. Behavior Research Methods, 52(4), 1795–1799.

Contact: yutong.zhang@ling-phil.ox.ac.uk