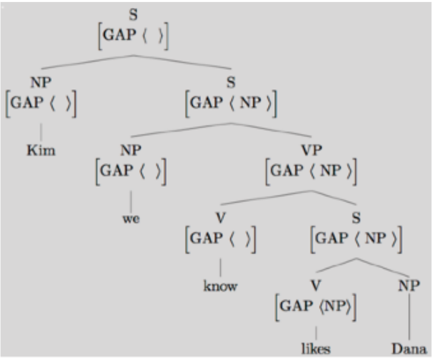
# Predictive processing in human language comprehension

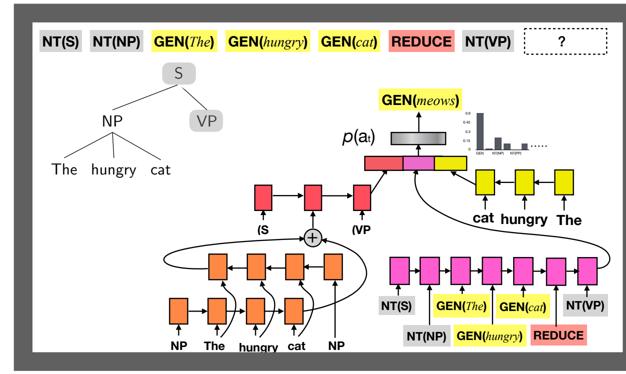
Roger Levy

9.19/9.190: Computational Psycholinguistics
November 8, 2023

# Triangulating on a model of human(-like) language




**Theory of linguistic knowledge**

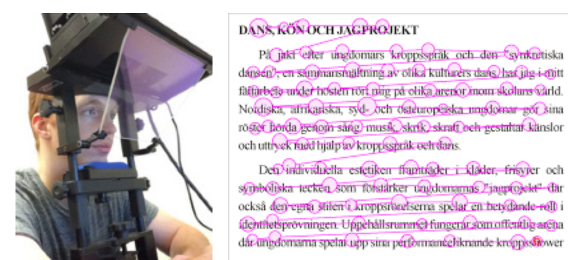



**Computational Models**

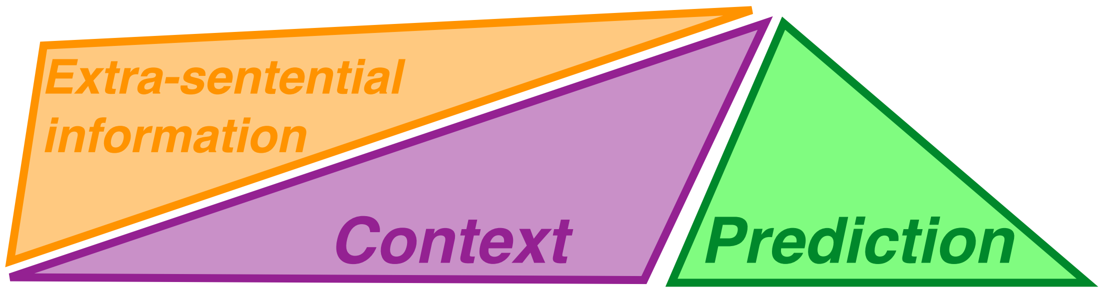
**Human(-like) linguistic knowledge and use**

**Language Datasets**

**Psychological Experimentation**



# Expectations in incremental comprehension




***Previous Input*** ***Current Input***

*These expectations from* ***diverse contextual cues*** *affect human language processing* ***extremely quickly***

- Syntactic:

  *Jamie was clearly intimidated… by* [source]

- Phonological knowledge:

  *Terry ate an… apple/orange/ice cream cone*

  *Terry ate a… nectarine/banana/sandwich*

- Semantic & situational knowledge:

  *The children went outside to…play*

  *The squirrel stored some nuts in the…~~statue~~*

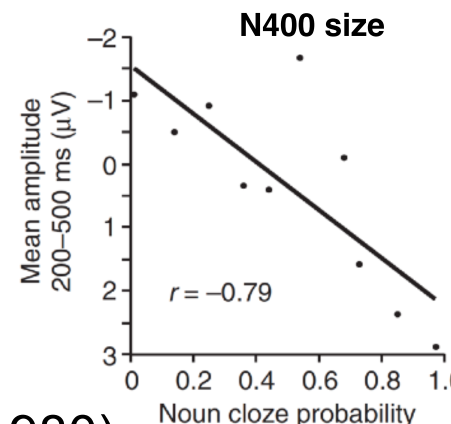
  *tree*

# Surprisal as an index of real-time processing load

- Let a word's difficulty be its *surprisal* given its context:

$$\text{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\text{CONTEXT})}$$
$$\left[ \approx \log \frac{1}{P(w_i|w_{1\ldots i-1})} \right]$$

- Captures the *expectation* intuition: the more we expect an event, the easier it is to process
  - Brains are prediction engines!



*(Federmeier & Kutas, 1999)*

- Predictable words are:
  - read faster (Ehrlich & Rayner, 1981)
  - have distinctive EEG responses (Kutas & Hillyard 1980)

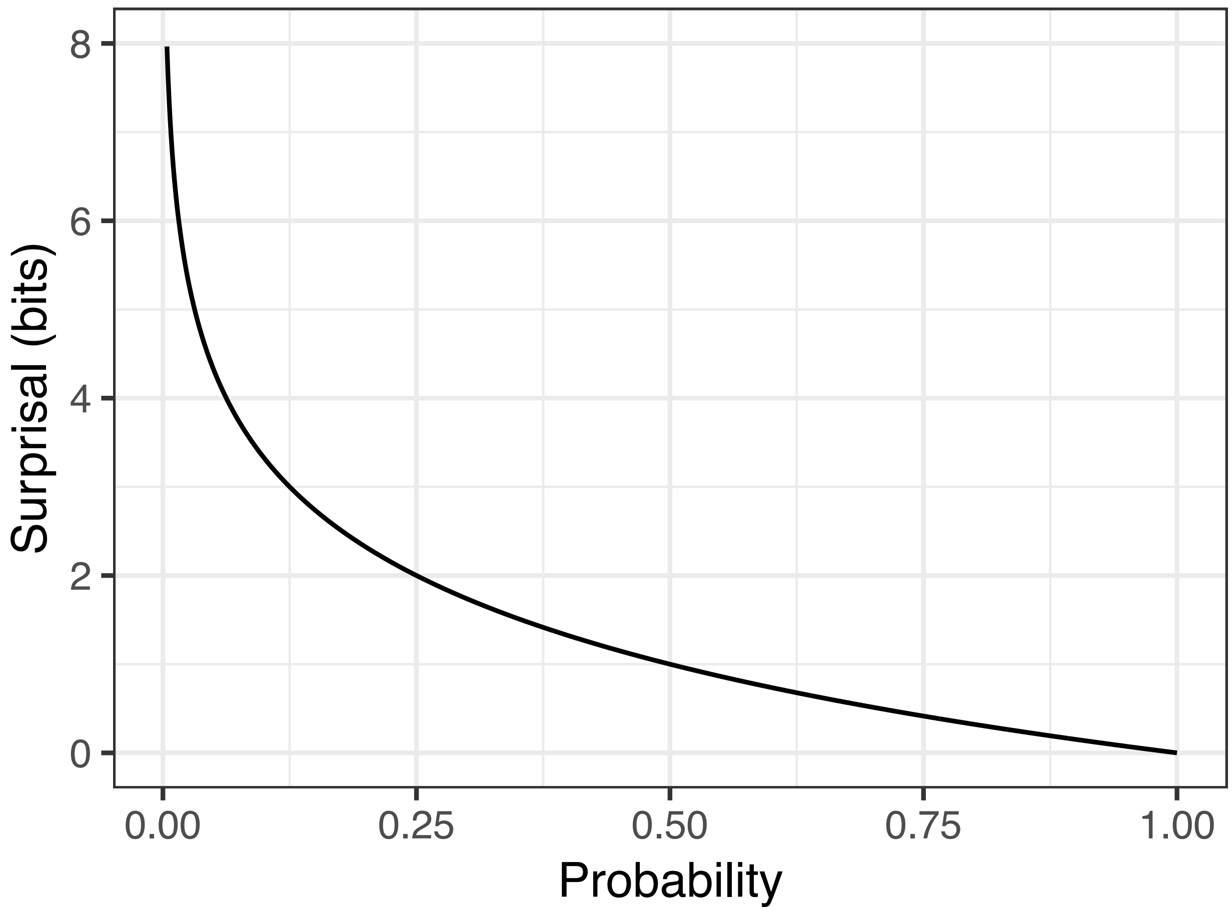
- with a language model that captures syntactic structure, we can get GRAMMATICAL EXPECTATIONS

*(Hale, 2001, NAACL; Levy, 2008, Cognition)*

# *Quantifying* structure and surprise

- Hypothesis: a word's difficulty is its *surprisal* in context:

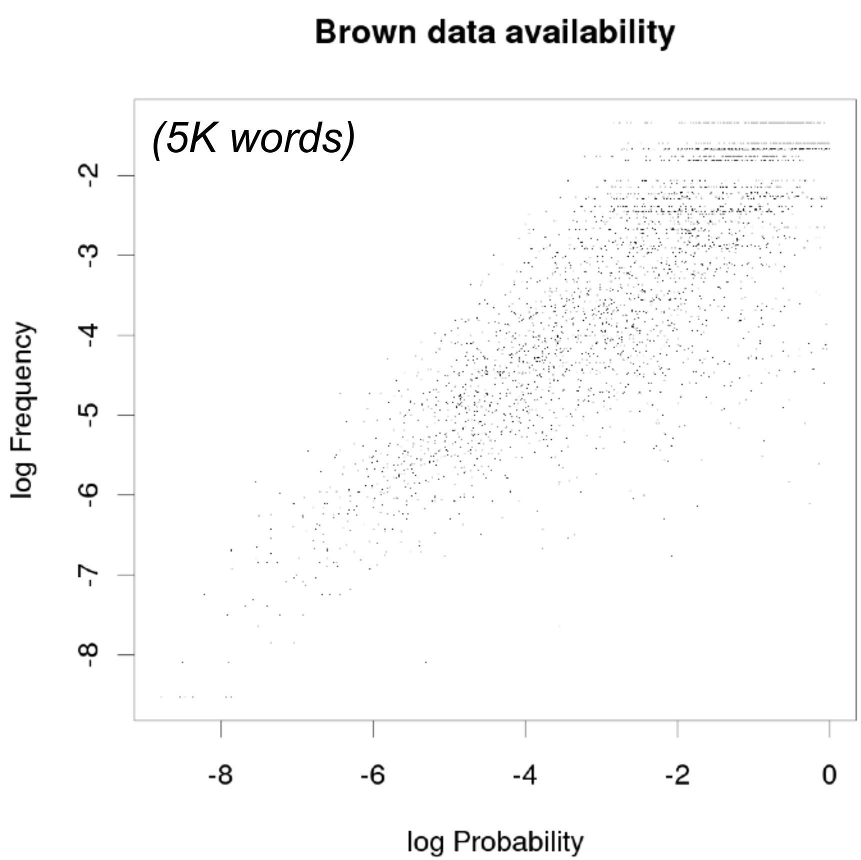
$$\mathbf{Surprisal}(w_i) \equiv \log \frac{1}{P(w_i|\mathbf{CONTEXT})}$$

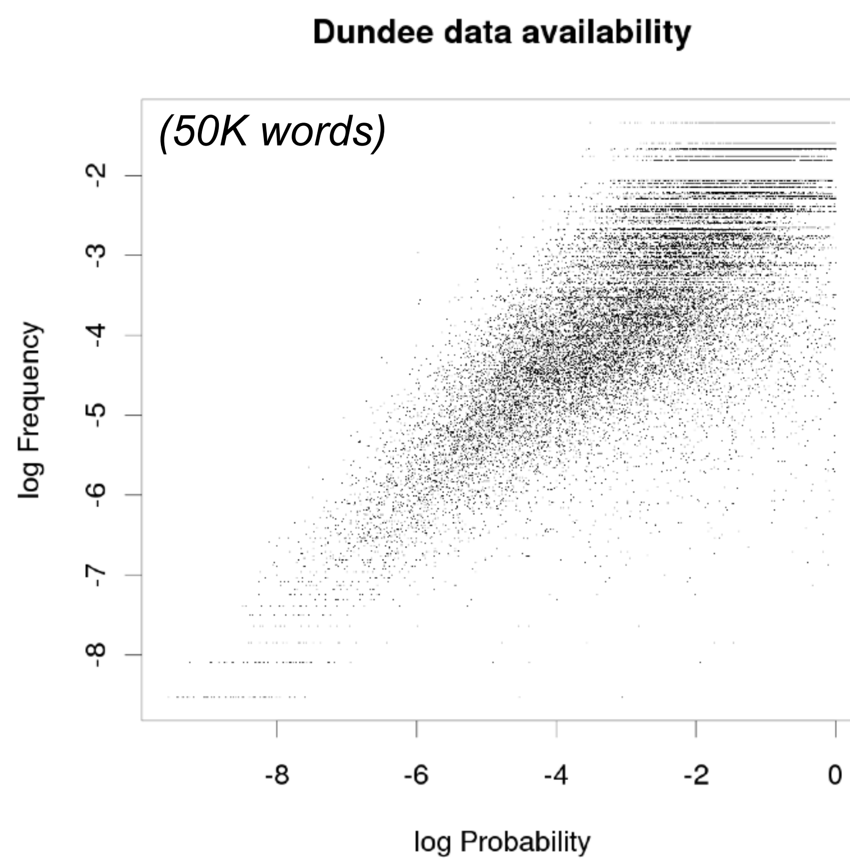



*(Shannon, 1948: a basic quantity from information theory!)*

# Estimating probability/time curve shape

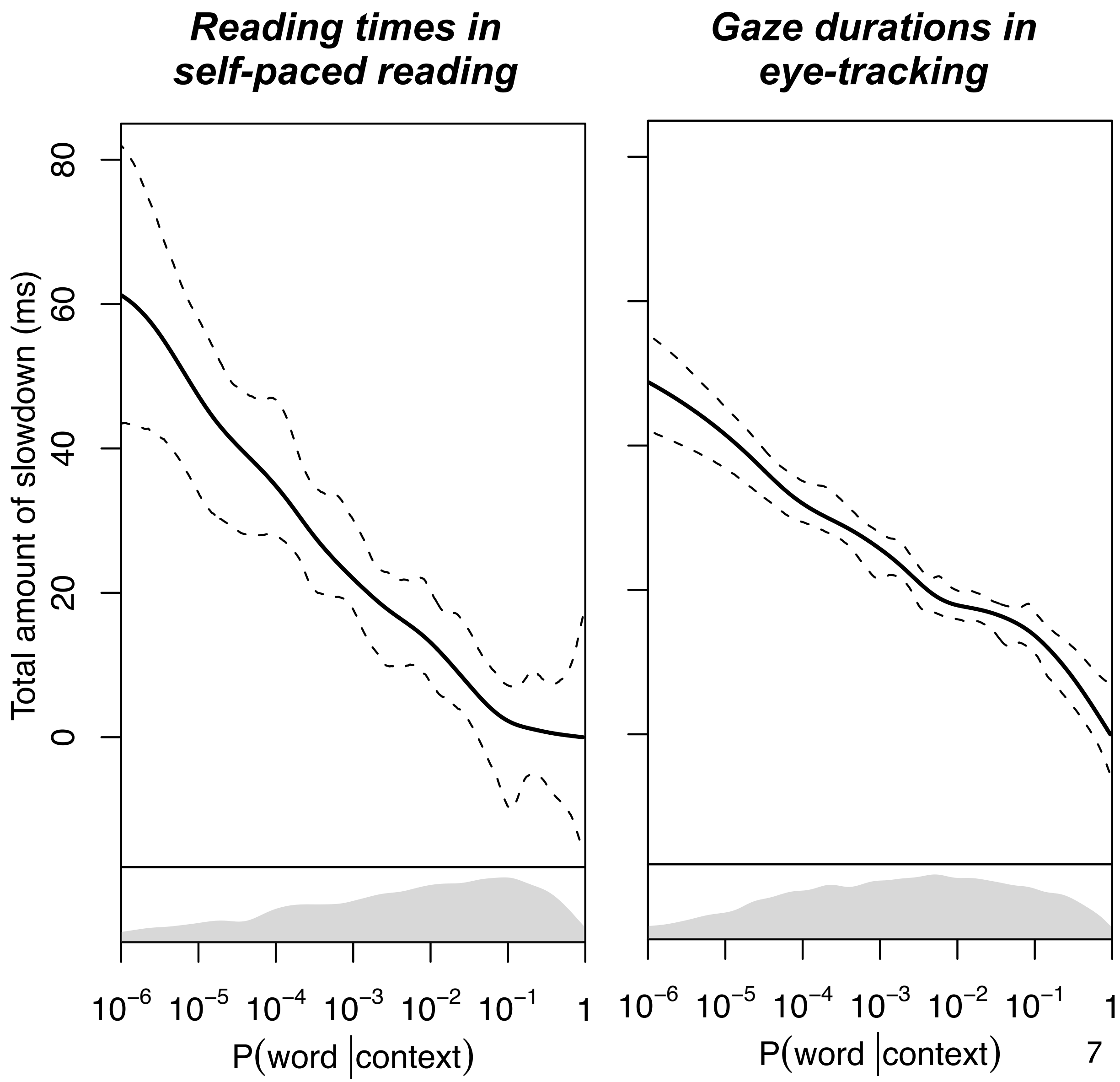
- As a proxy for "processing difficulty," reading time in two different methods: self-paced reading & eye-tracking
- Challenge: we need big data to estimate curve shape, but probability correlated with confounding variables



# Estimating probability/time curve shape

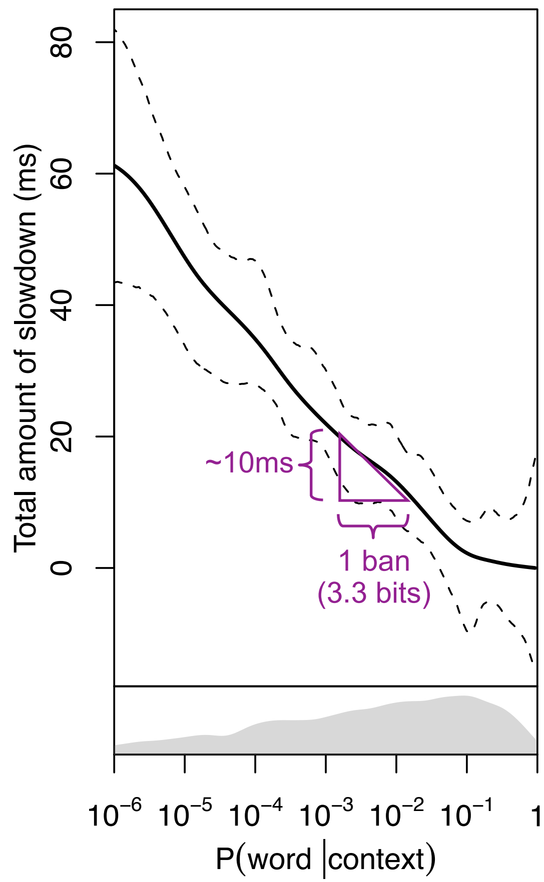
- Generalized additive model regression: total contribution of word (trigram) probability to RT near-linear over 6 orders of magnitude!




*(Smith & Levy, 2013)*

# Take-away: how long to process a word in context?

- On average, time ***linear in the word's log-probability***
- Methodologically: reading puts control in the comprehender's hands (and eyes!), allowing us to study processing difficulty through reading time



# A model system with incrementality, structure, and surprise

(                                        )

*The woman((who was)brought the sandwich from the kitchen)tripped.*

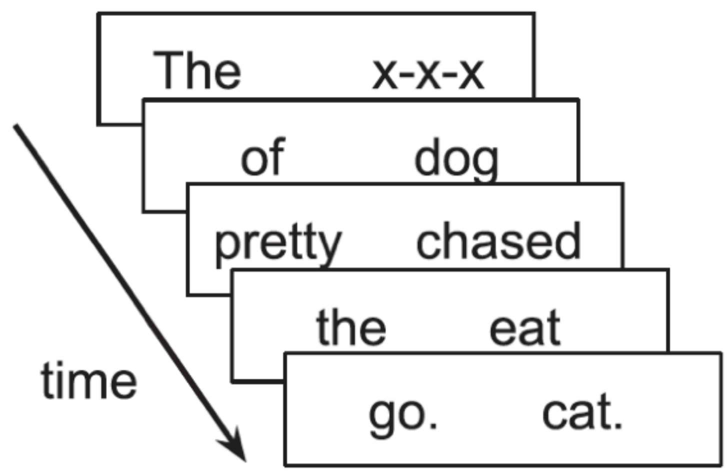
***The woman(given the sandwich from the kitchen)tripped.***

*The woman((who was)given the sandwich from the kitchen)tripped.*

| | Simple past | Past participle |
|---|---|---|
| **bring** | ***brought*** | ***brought*** |
| **give** | ***gave*** | ***given*** |

# Low-tech, crowd-sourceable reading

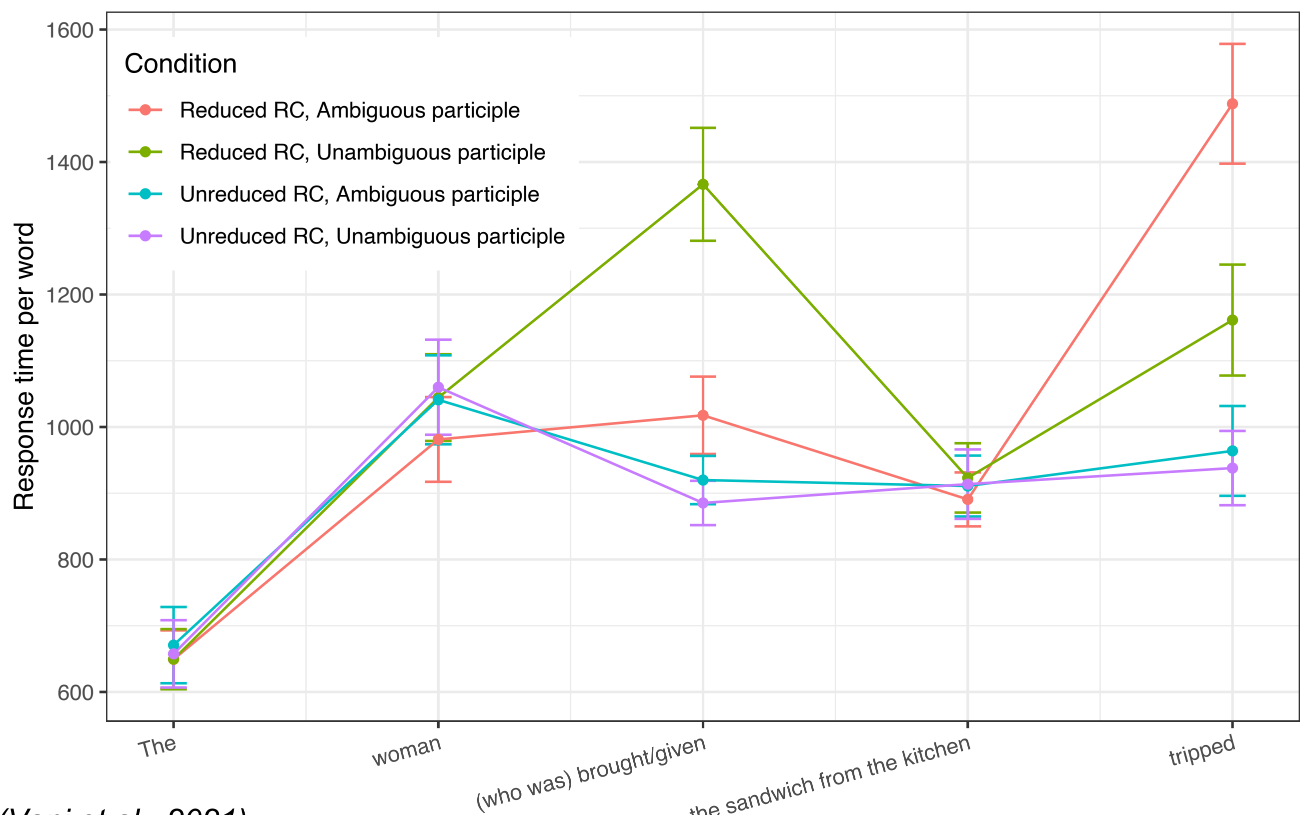
- The **maze task**
- Choose the word that fits given the preceding context


(Forster et al., 2009; Boyce et al., 2020)

# Incrementality, structure, and surprise

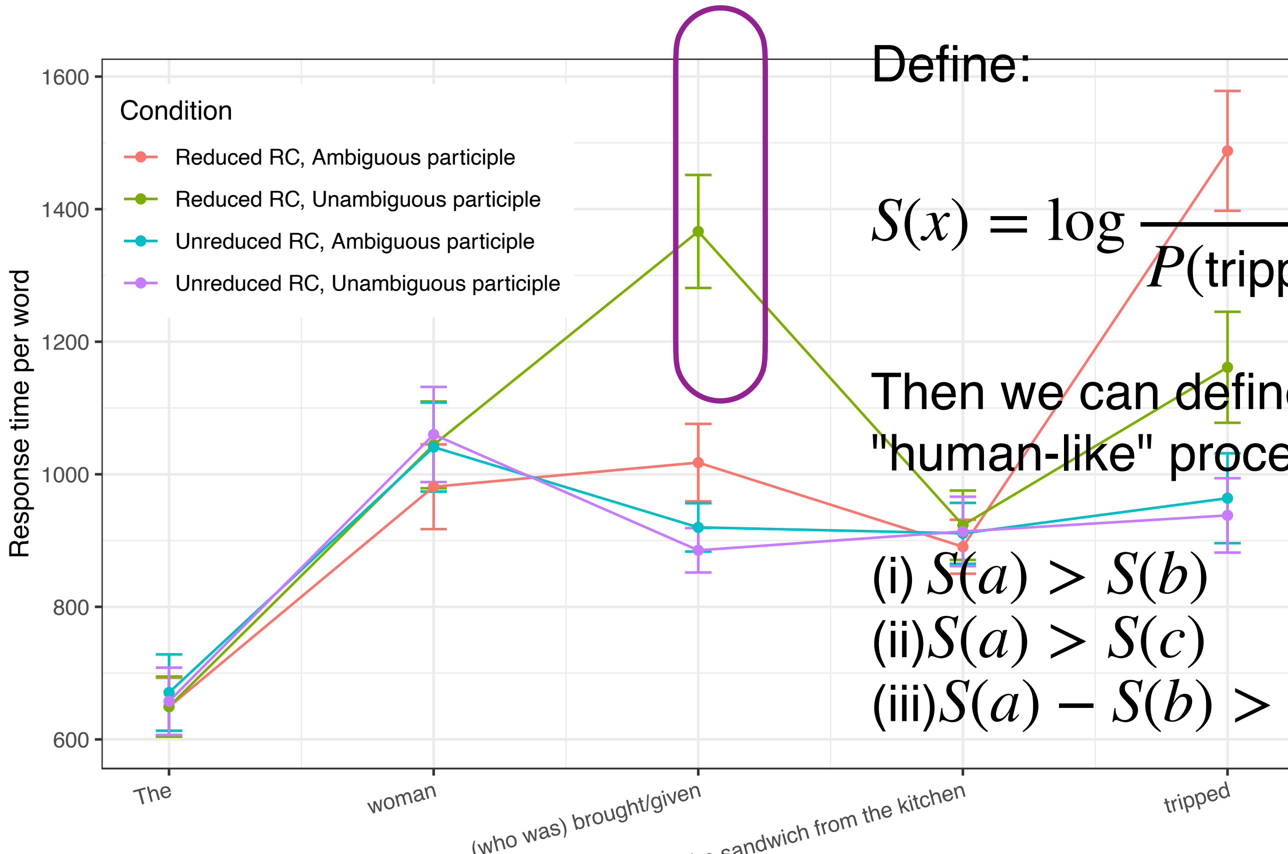
| | | Is the relative clause reduced? | Is the participle part-of-speech ambiguous? |
|---|---|---|---|
| ***The woman brought the sandwich from the kitchen tripped.*** | | + | + |
| ***The woman given the sandwich from the kitchen tripped.*** | | + | – |
| ***The woman who was brought the sandwich from the kitchen tripped.*** | | – | + |
| ***The woman who was given the sandwich from the kitchen tripped.*** | | – | – |




*(Vani et al., 2021)*

# Desiderata for human-like processing

| | | Is the relative clause reduced? | Is the participle part-of-speech ambiguous? |
|---|---|---|---|
| *The woman brought the sandwich from the kitchen tripped.* | a | + | + |
| *The woman given the sandwich from the kitchen tripped.* | b | + | – |
| *The woman who was brought the sandwich from the kitchen tripped.* | c | – | + |
| *The woman who was given the sandwich from the kitchen tripped.* | d | – | – |




Define:

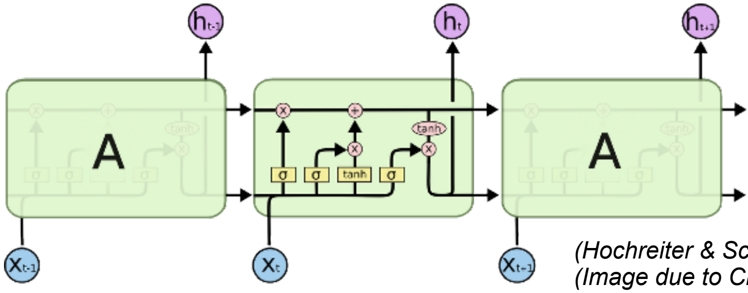
$$S(x) = \log \frac{1}{P(\text{tripped} \mid \text{Context}_x)}$$

Then we can define three criteria for "human-like" processing:

(i) $S(a) > S(b)$
(ii) $S(a) > S(c)$
(iii) $S(a) - S(b) > S(c) - S(d)$

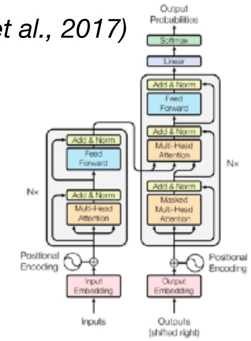
(Futrell et al., 2019; Hu et al., 2020; Gauthier et al., 2020; Vani et al., 2021)

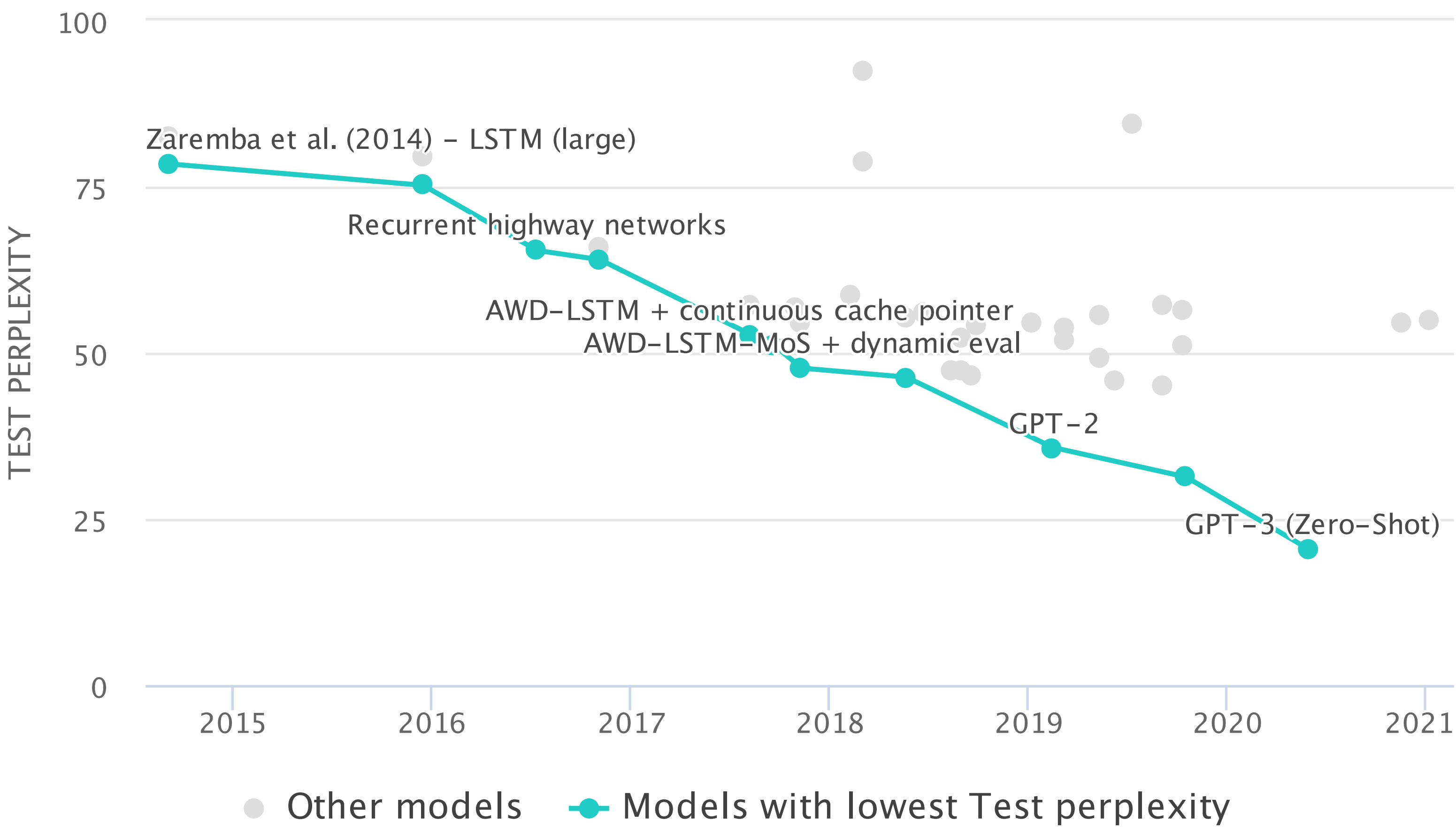
# Deep learning has revolutionized language modeling




(Hochreiter & Schmidhuber, 1997)
(Image due to Christopher Olah)

(Vaswani et al., 2017)

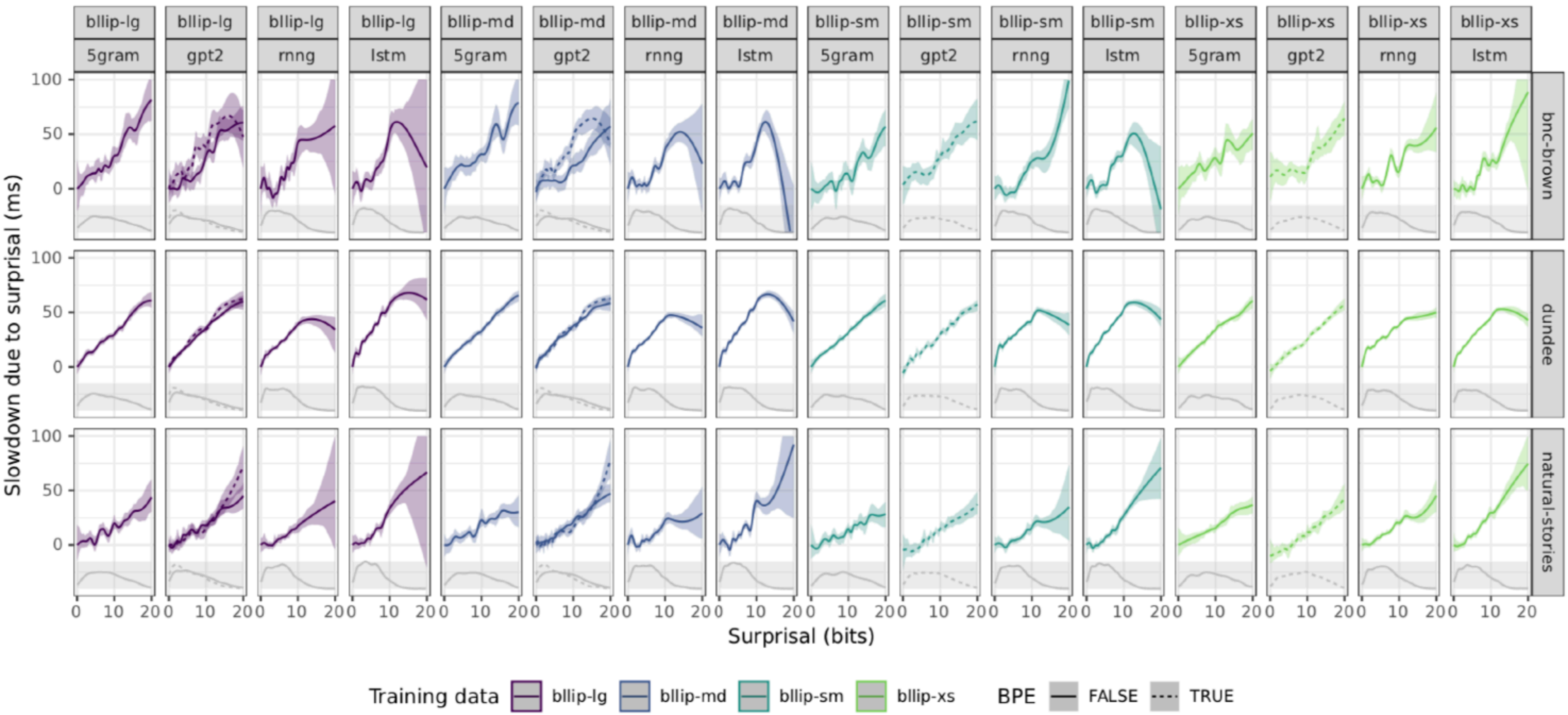





*https://paperswithcode.com/sota/language-modelling-on-penn-treebank-word*

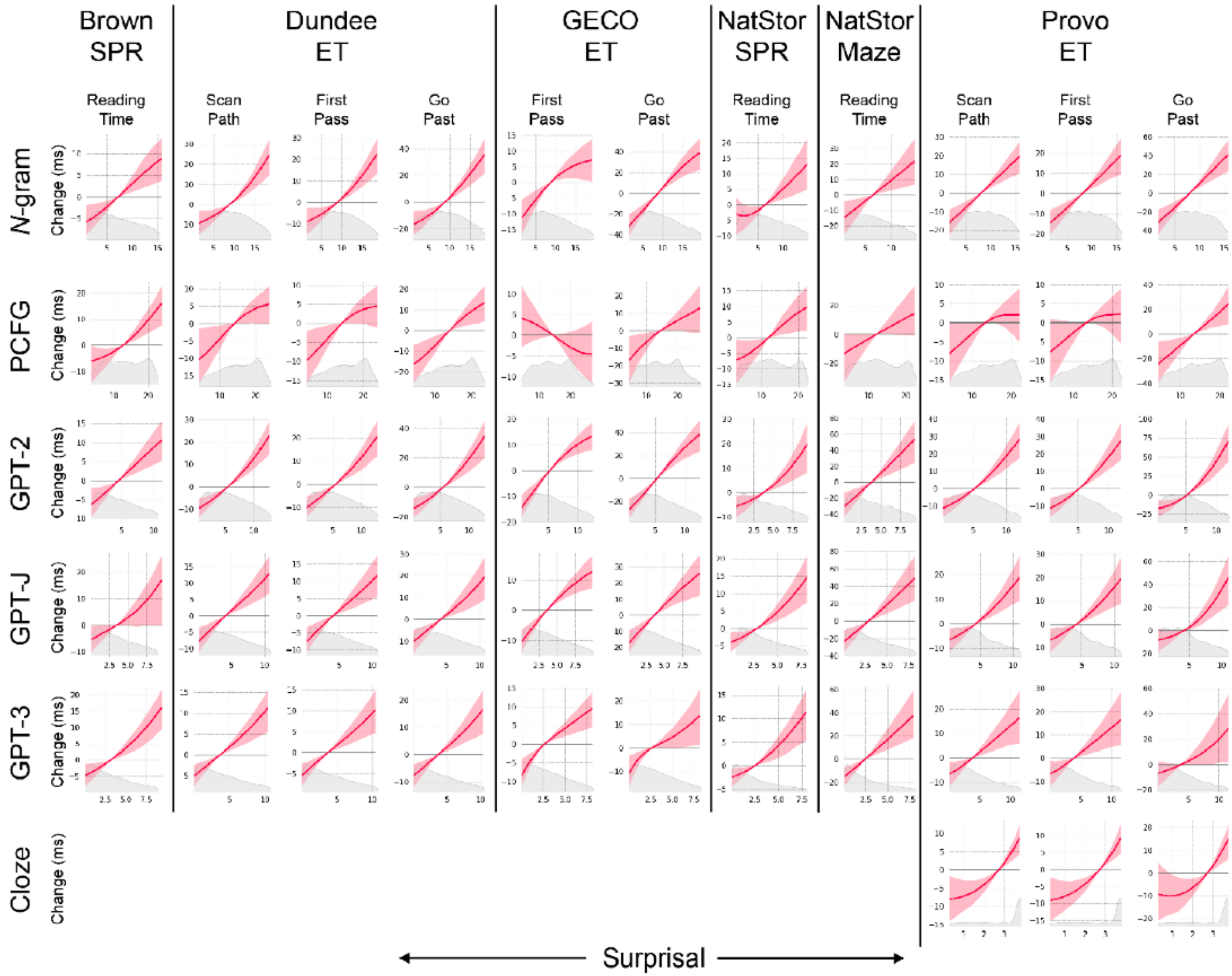
# *Quantitative* calibration to human processing

- The surprisal–RT relationship in naturalistic reading:




(Wilcox et al., 2020)

# *Quantitative* calibration to human processing




*(Shain et al., 2022)*

# Brain signatures of predictive processing

EEG

(Creator: Tim Sheerman-Case, CC-BY)

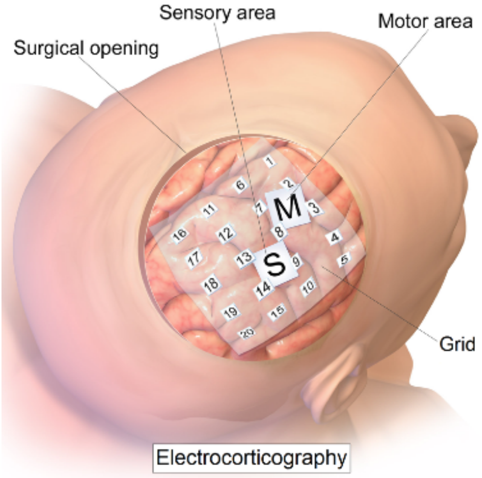
MEG


(Creator: J.M Eddings Jr, CC-BY-NC)

fMRI


(NIH Image Gallery, public domain)

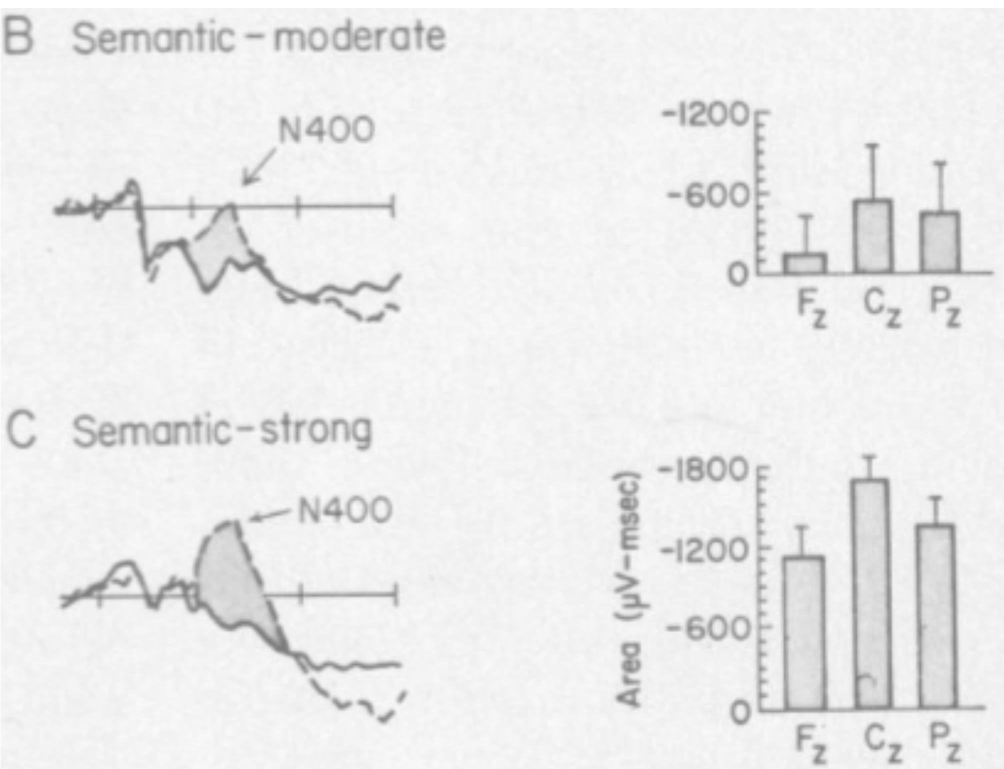
ECoG

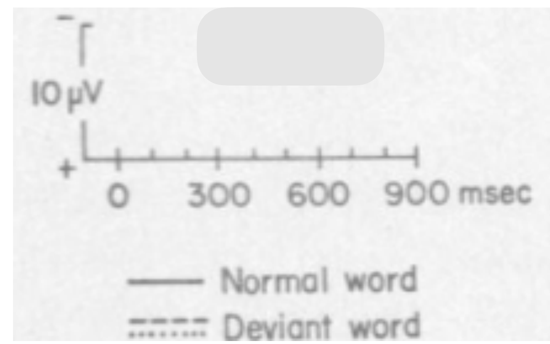



https://commons.wikimedia.org/wiki/File:Intracranial_electrode_grid_for_electrocorticography.png

# The N400 in language comprehension

- Differing degrees of semantic congruity:
  - He took a sip from the *drink*. (normal)
  - He took a sip from the *waterfall*. (moderate incongruity)
  - He took a sip from the *transmitter*. (strong incongruity)



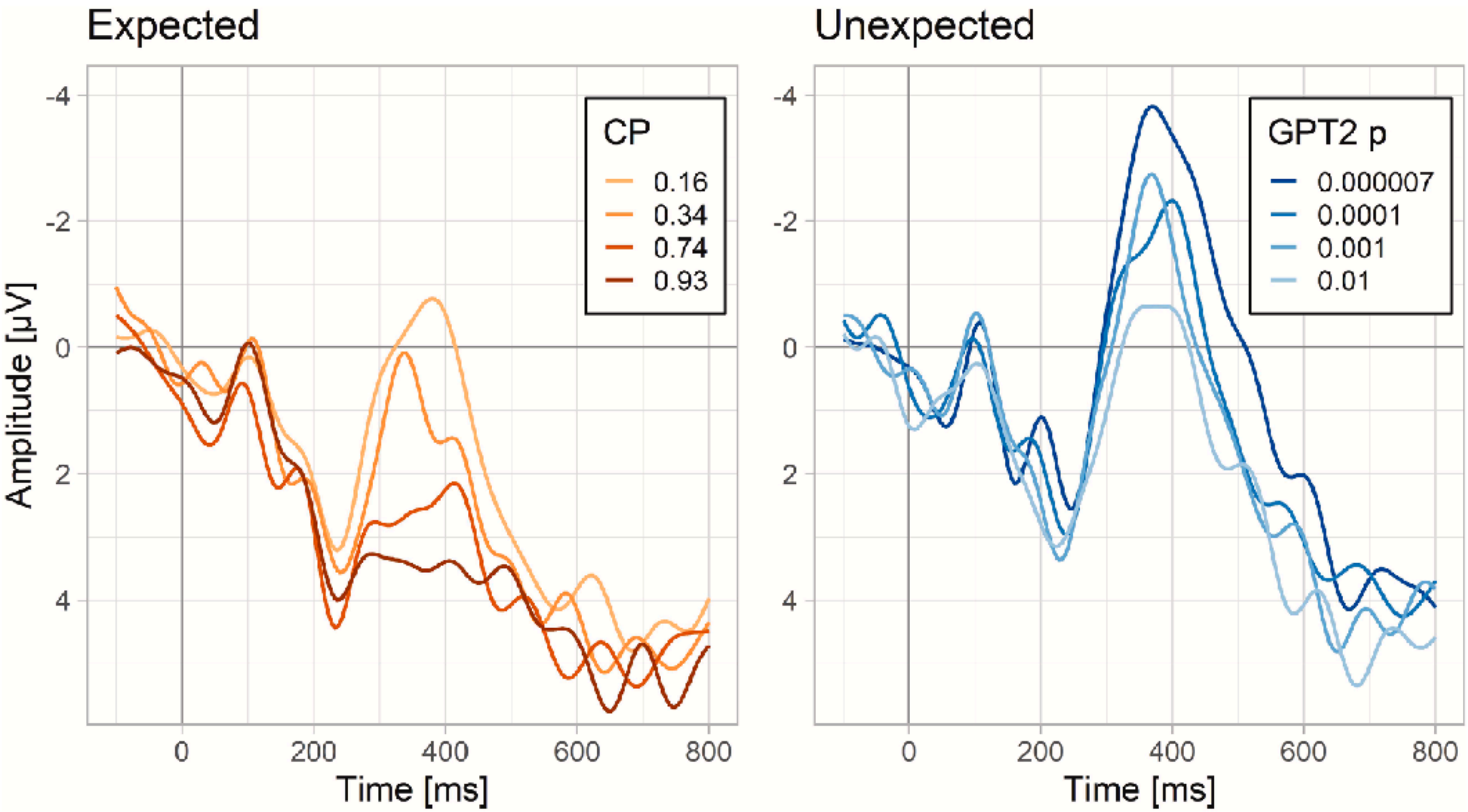




(Kutas & Hillyard, 1980, 1984)

# Word probability effects in the brain

**Weakly constraining** *Joy was too frightened to...* ***look*** ***move***

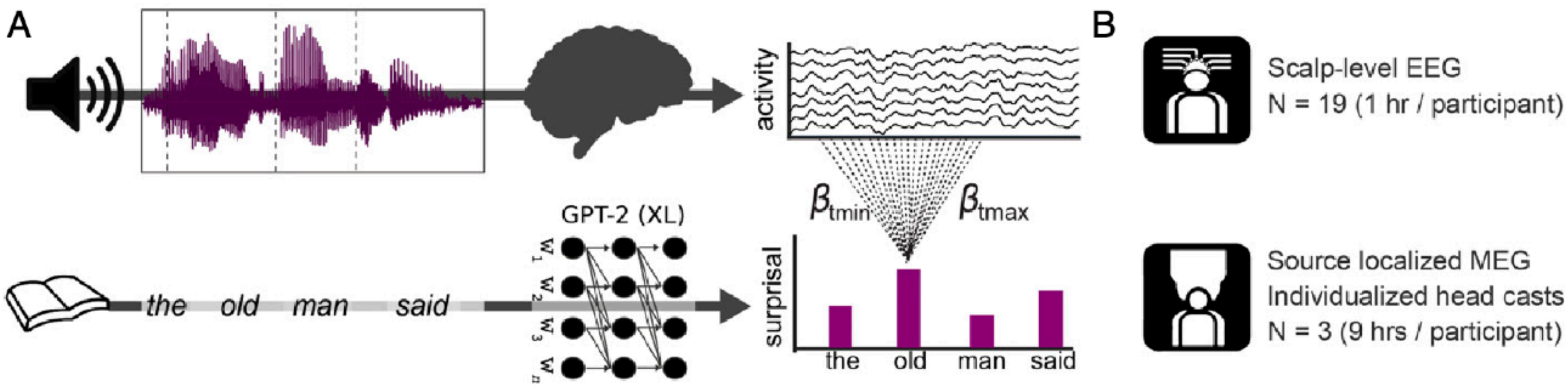
**Strongly constraining** *He brought her a pearl necklace for her...* ***collection*** ***birthday***



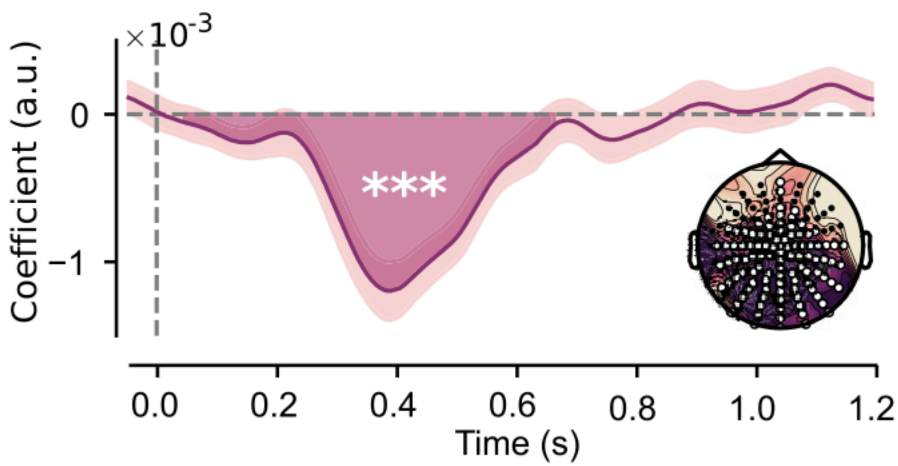

*(Original data: Federmeier et al., 2007; analysis: Szewczyk & Federmeier, 2022)*

# Surprisal effects in audiobook listening

- Analytic framework:

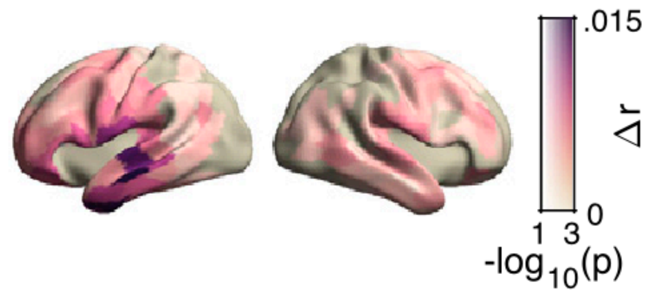



EEG results (temporal)



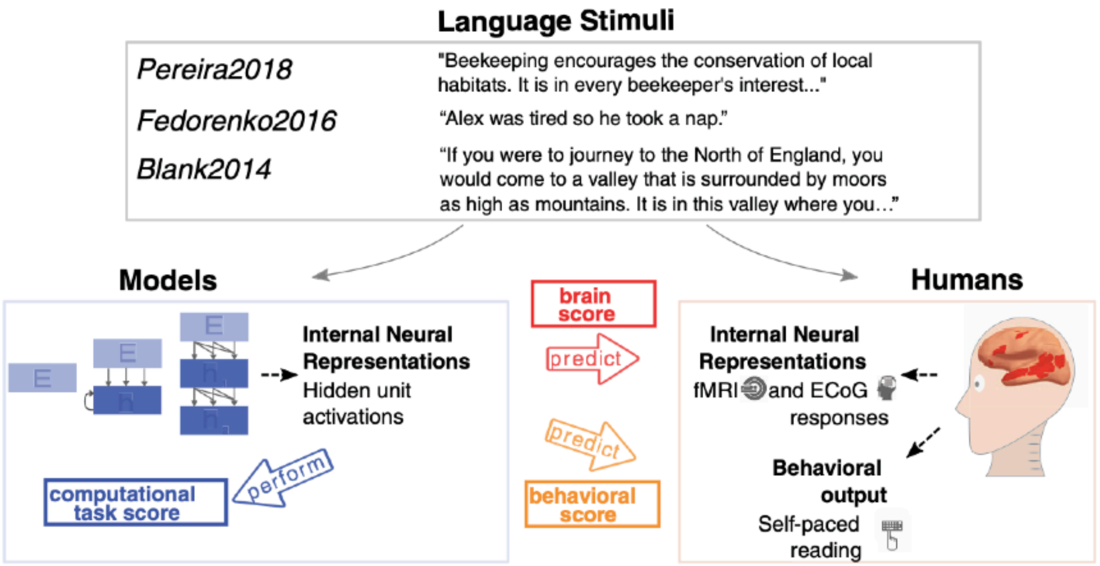

MEG results (temporal)

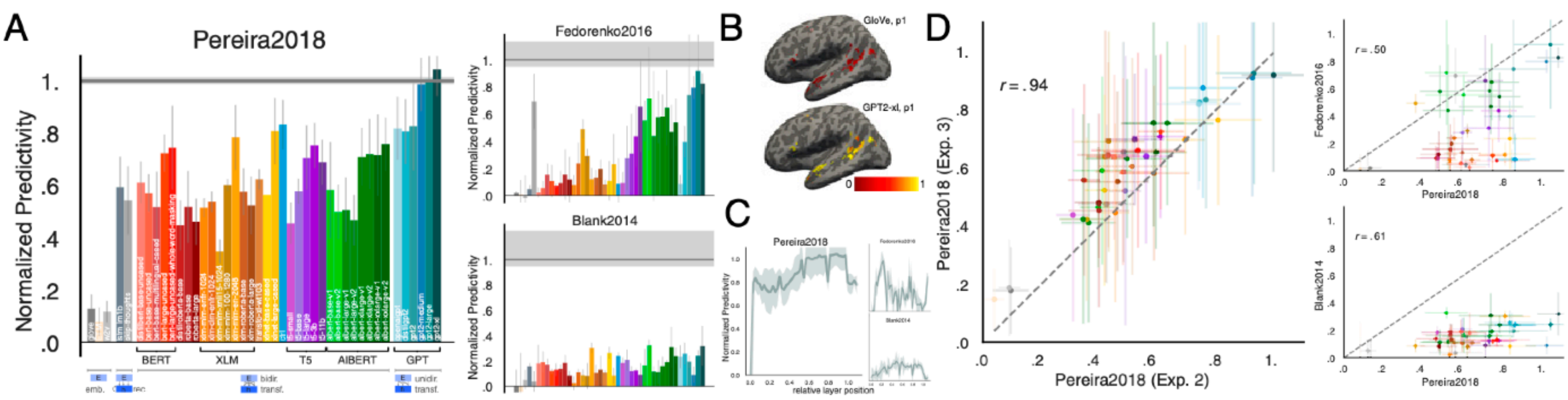
continuous prediction > baseline




(Heilbron et al., 2022)

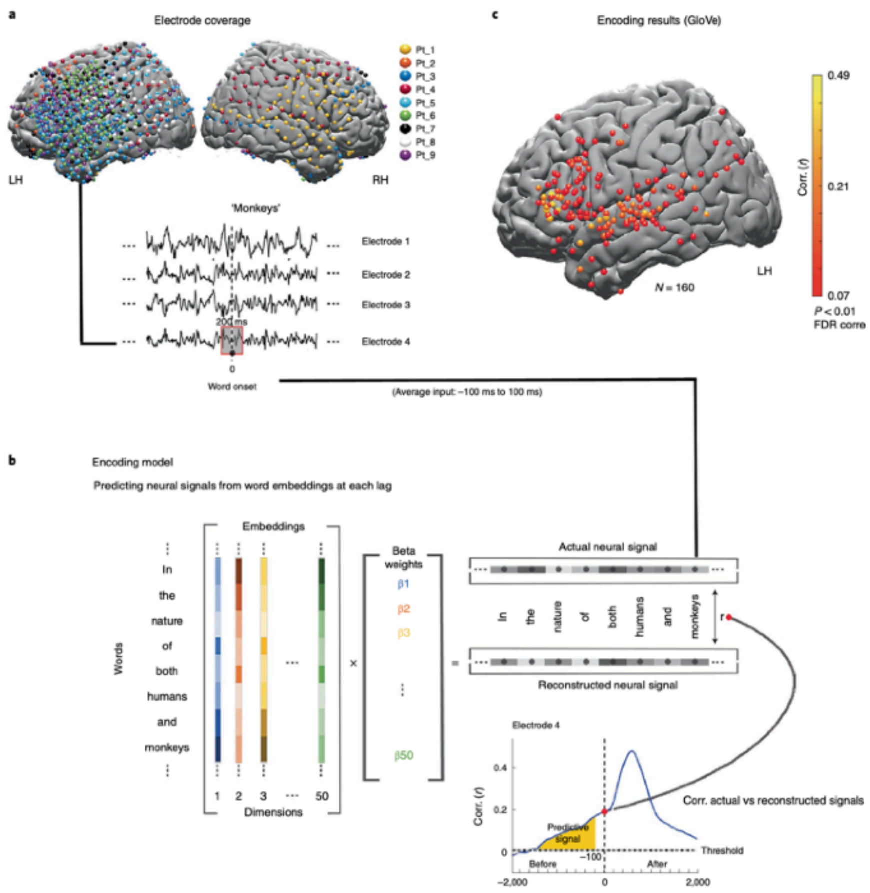
# Aligning neural network embeddings to brain responses

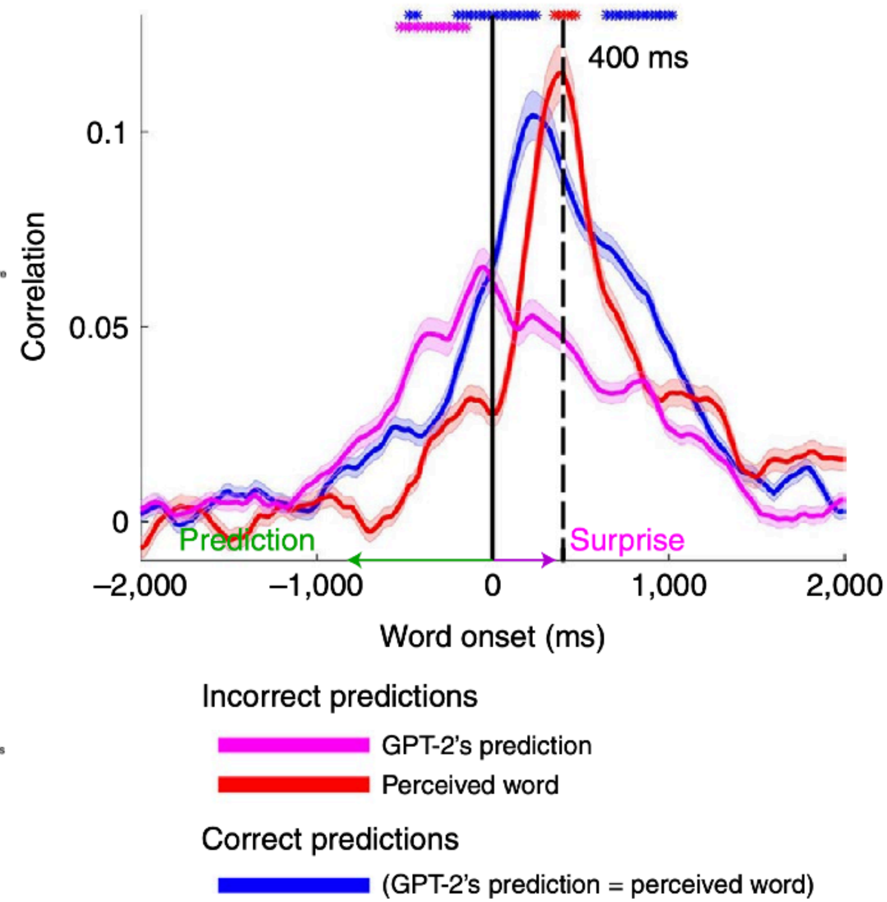



(Schrimpf et al., 2021)

# Prediction versus surprise in ECoG

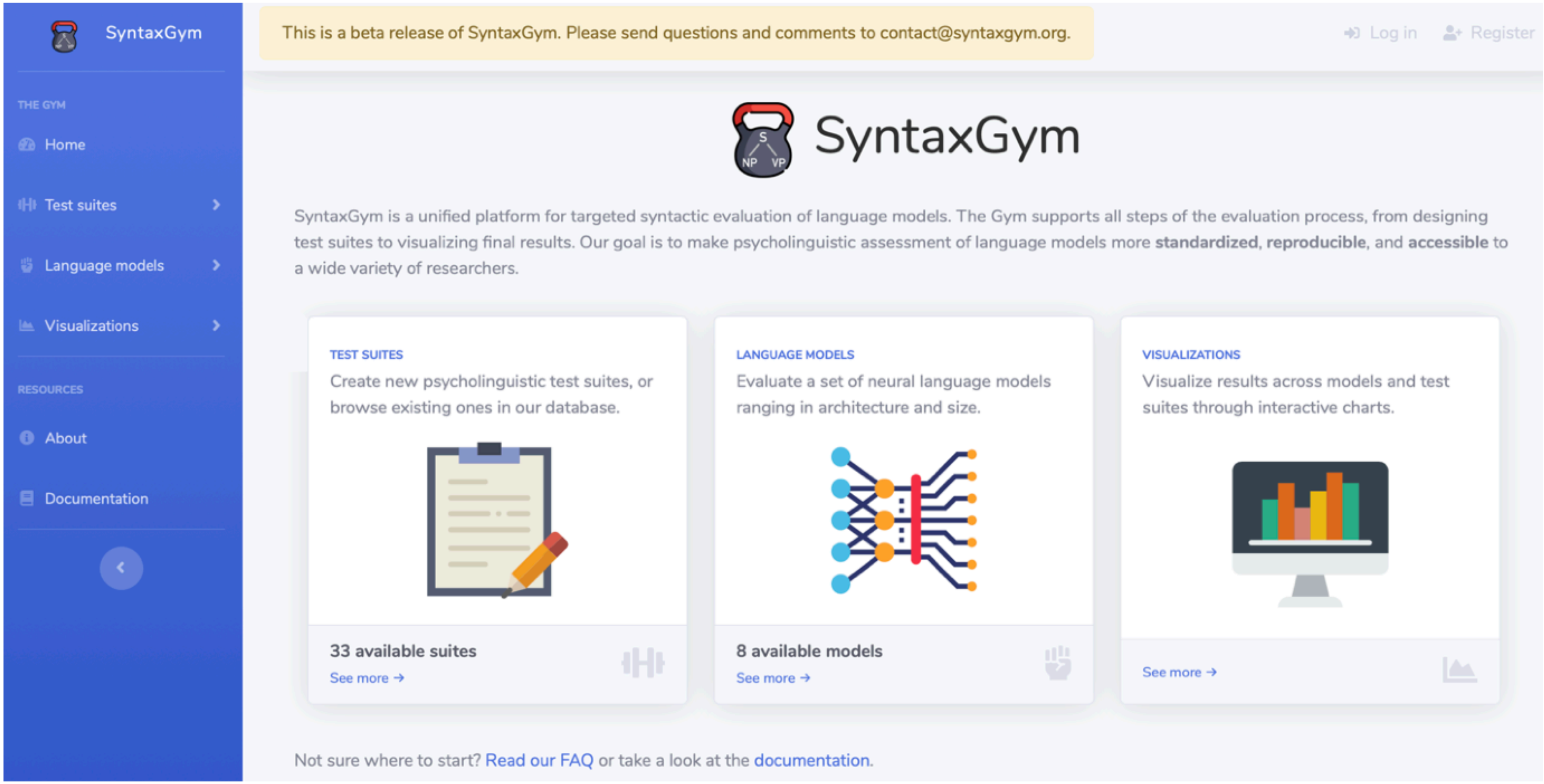



*(Goldstein et al., 2022)*

# In-class exercise: explore GPT-2 word predictions

# Psycholinguistic tests of AI language models




**http://syntaxgym.org**

*(Gauthier et al., 2020)*

# References


Boyce, V., Futrell, R., & Levy, R. (2020). Maze made easy: Better and easier measurement of incremental processing difficulty. Journal of Memory and Language, 111, 1{13.

Federmeier, K. D., Wlotko, E. W., Ochoa-Dewald, E. D., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. Brain Research, 1146, 75{84.

Forster, K. I., Guerrera, C., & Elliot, L. (2009). The maze task: Measuring forced incremental sentence processing time. Behavior Research Methods, 41(1), 163{171.

Futrell, R., Wilcox, E., Morita, T., Qian, P., Ballesteros, M., & Levy, R. (2019). Neural language models as psycholinguistic subjects: Representations of syntactic state, In Proceedings of the 18th Annual
Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.

Gauthier, J., Hu, J., Wilcox, E., Qian, P., & Levy, R. P. (2020). SyntaxGym: An online platform for targeted evaluation of language models, In Proceedings of the 58th annual meeting of the Association for Computational
Linguistics.

Goldstein, A., Zada, Z., Buchnik, E., Schain, M., Price, A., Aubrey, B., Nastase, S. A., Feder, A., Emanuel, D., Cohen, A., Jansen, A., Gazula, H., Choe, G., Rao, A., Kim, C., Casto, C., Fanda, L., Doyle, W.,
Friedman, D., Dugan, P., Melloni, L., Reichart, R., Devore, S., Flinker, A., Hasenfratz, L., Levy, O., Hassidim, A., Brenner, M., Matias, Y., ... Hasson, U. (2022). Shared computational principles
for language processing in humans and deep language models. Nature Neuroscience, 25(3), 369{380.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model, In Proceedings of the second meeting of the north american chapter of the Association for Computational Linguistics, Pittsburgh, Pennsylvania. Heilbron, M., Armeni, K., Schoffelen, J.-M., Hagoort, P., & Lange, F. P. d. (2022). A hierarchy of linguistic predictions during natural language comprehension. Proceedings of the National Academy of Sciences,
119(32), e2201968119.

Hu, J., Gauthier, J., Qian, P., Wilcox, E., & Levy, R. P. (2020). A systematic assessment of syntactic generalization in neural language models, In Proceedings of the 58th annual meeting of the Association
for Computational Linguistics.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. Science, 207(4427), 203{205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. Nature, 307, 161{163.

Levy, R. (2008). Expectation-based syntactic comprehension. Cognition, 106(3), 1126{1177.

Schrimpf, M., Blank, I. A., Tuckute, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive
processing. Proceedings of the National Academy of Sciences, 118(45).

Shain, C., Meister, C., Pimentel, T., Cotterell, R., & Levy, R. P. (2022). Large-scale evidence for logarithmic effects of word predictability on reading time. PsyArXiv.

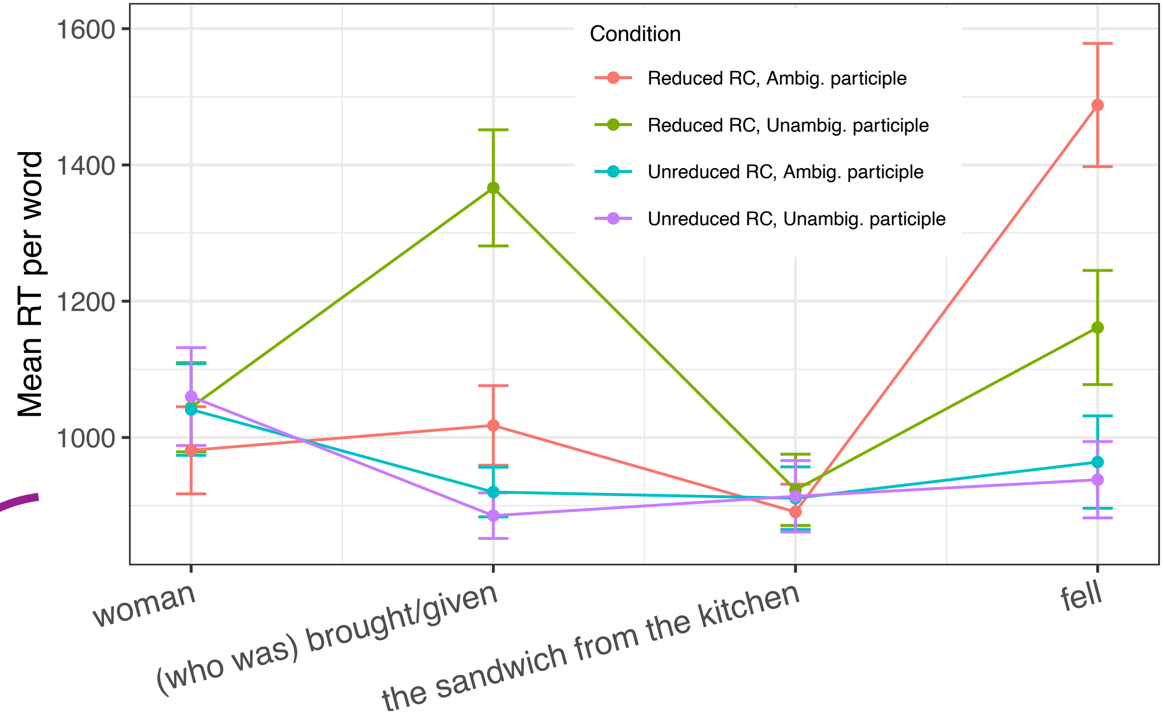
Shannon, C. (1948). A mathematical theory of communication. Bell Systems Technical Journal, 27(4), 623{656.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3), 302{319.

Szewczyk, J. M., & Federmeier, K. D. (2022). Context-based facilitation of semantic access follows both logarithmic and linear functions of stimulus probability. Journal of Memory and Language, 123, 104311. Van Schijndel, M., & Linzen, T. (2021). Single-stage prediction models do not explain the magnitude of syntactic disambiguation difficulty. Cognitive Science, 45(6), e12988.

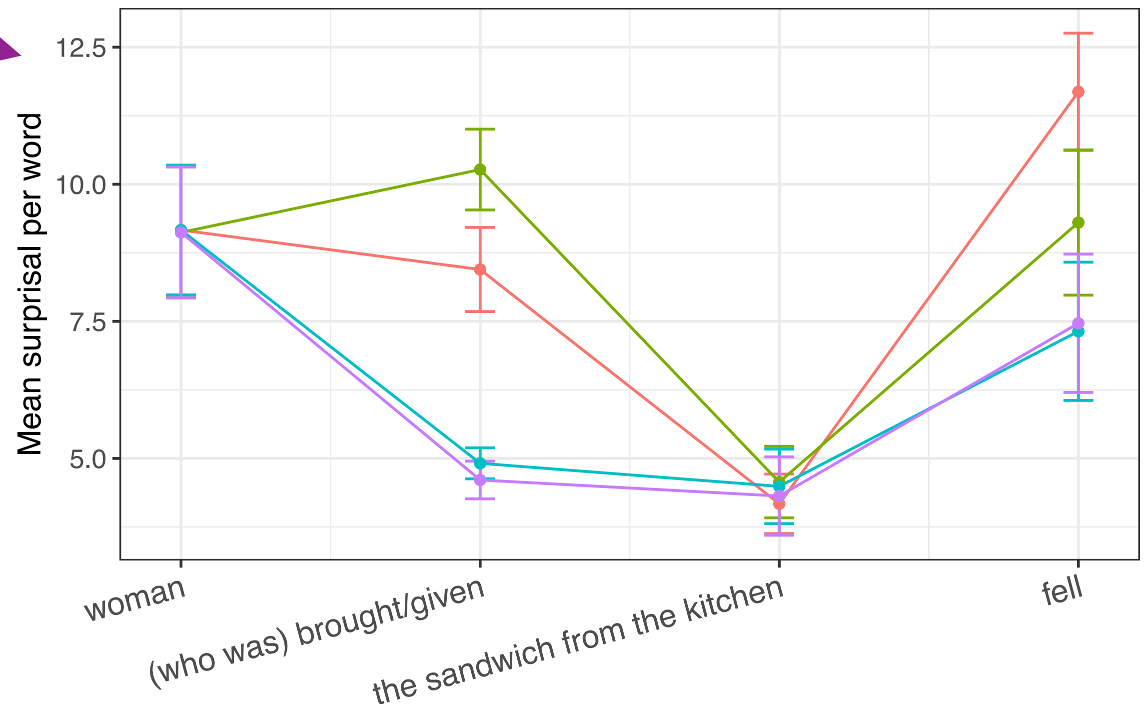
Vani, P., Wilcox, E. G., & Levy, R. P. (2021). Using the Interpolated Maze task to assess incremental processing in English relative clauses, In Proceedings of the 43rd annual meeting of the Cognitive Science
Society.

# Human reaction times




*Pooling many controlled experiments, regress human RTs against model surprisal and examine* ***residual***

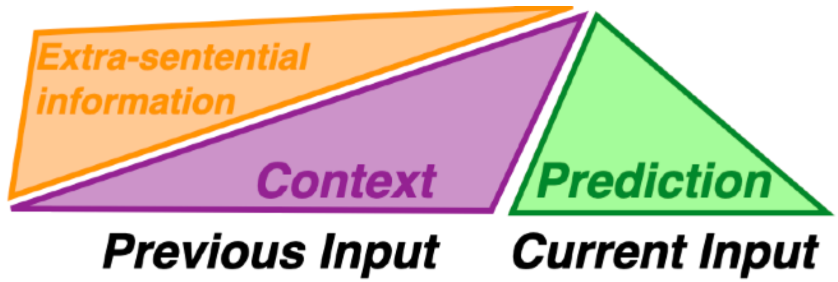
# GPT-2 Surprisal



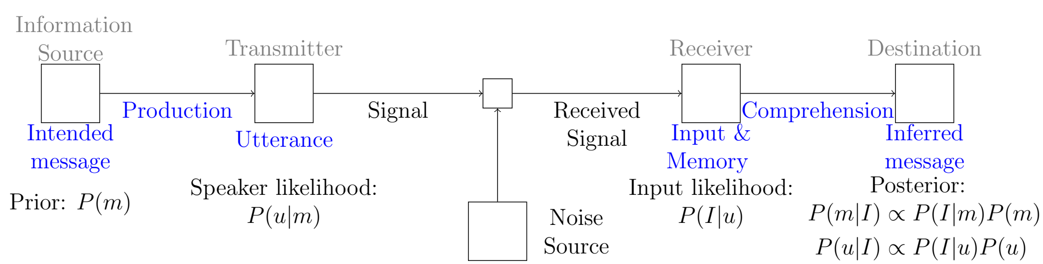

*(Wilcox et al., 2021; see also van Schijndel & Linzen, 2021)*

# Ingredients for theory of human language comprehension

- Ubiquitous expectation-based inference, including prediction/surprisal

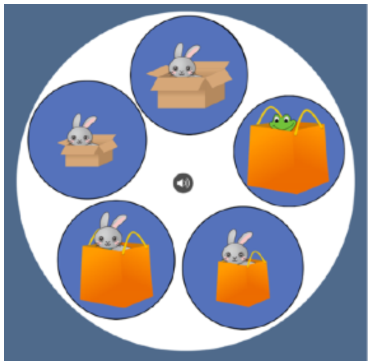



- Noisy-channel mechanisms for error detection & robustness (Levy 2008, Gibson et al., 2013, Futrell et al., 2020)




- **And of course:** Incremental semantic representations evaluable in context (Jacobson 1999, Aparicio et al. in prep)

*Click on the rabbit in the big...*



*Mary loves and John hates...*

**λx[LOVE(x)(mary) ∧HATE(x)(john)]**