Lexical Retrieval and Probabilistic Inference Dissociate in Naturalistic Reading

Cory Shain, MIT

Debate exists about why less frequent and less predictable words are harder to read [3, 13, 11, 24, 17, 7]. *Representational* theories emphasize structure-building in working memory, a process in which lexical retrieval and prediction are distinct processes that independently modulate processing difficulty [24]. This position predicts an additive relationship between frequency and predictability. *Inferential* theories emphasize probabilistic inference in which an interpretation is a probability distribution over possible sentence meanings that must be updated word by word [13, 11]. This position predicts that apparent frequency effects reduce to predictability.

Two previous lines of evidence support a frequency-predictability dissociation (and thus, the representational view). *First*, several studies have shown additive frequency and predictability effects on overall reading time [1, 16, 10]. However, this evidence is mostly based on small samples, unnatural tasks (reading isolated constructed sentences), and implausible modeling assumptions (a discrete-time, linear, additive, stationary, and/or homoscedastic reader), which raises the question: do frequency and predictability dissociate in ordinary language comprehension, such as story reading? Two prior naturalistic reading studies addressed this question but came to opposite conclusions [17, 7]. *Second*, two prior studies collectively support dissociable effects of frequency and predictability on the *distribution* of reading times under an exGaussian model, since frequency modulated both location and skewness in one study [25], whereas predictability modulated only location in another [23]. However, this difference across studies has never been directly tested.

This work simultaneously revisits both of these lines of reasoning at scale by analyzing a large collection of naturalistic reading data (six datasets, $>$ 2.2M datapoints) [9, 22, 4, 12, 5, 2] using recent nonlinear regression techniques that can capture effects on distinct distributional parameters within a rich mixed-effects design [20]. Word-by-word predictability estimates are computed by advanced statistical language models (GPT-2 surprisal, [14]) that yield better psychometric performance than human cloze norms [26, 19], and frequency (unigram surprisal) estimates are derived from a 6.5B word corpus comparable to GPT-2's training data [6]. Methods otherwise follow [19].

Fig 1 shows frequency and predictability effects on both mean reading time and the three parameters of the exGaussian distribution. Plots show the estimated change in reading time (vertical axis) as a function of two variables, either frequency/predictability and *delay*, (representing an *impulse response function*, IRF) or frequency vs. predictability at delay 0 (representing an interaction). As shown, in most datasets, frequency and predictability each significantly modulate mean reading time with no significant interaction between them. However, contrary to [25, 23], no clear dissociations emerge at the level of distributional parameters: the bulk of both frequency and predictability effects are driven by their effect on the skewness parameter β (sometimes called τ), rather than on location or dispersion.

In summary, despite the use of strong predictability estimates and flexible regression models, results converge with earlier studies in supporting additive (dissociable) frequency and predictability effects, as predicted by representational theories and contrary to the notion that frequency effects reduce to predictability. Results additionally clarify that this dissociation primarily arises in aggregate measures, rather than emerging between distributional parameters. This outcome does not undermine the claim of inferential theories that probabilistic inference plays a major role in language processing [22, 28, 19]. Instead, results support the existence of an additional cognitive demand (lexical retrieval) that is not explained by standard inferential theories. In this way, this study joins recent evidence that intralexical priming [21], dependency locality [18], and garden-path constructions [27] modulate processing demand over and above predictability. These results motivate increased attention to theoretical accounts that can accommodate these two sets of facts, e.g., by capturing the demands of representation-building within a probabilistic framework [15, 8].

References

[1] Altarriba, J., Kroll, J. F., Sholl, A., and Rayner, K. 1996.

[2] Boyce, V. and Levy, R. P. 2022.

[3] Coltheart, M., Rastle, K., Perry, C., Langdon, R., and Ziegler, J. 2001.

[4] Cop, U., Dirix, N., Drieghe, D., and Duyck, W. 2017.

[5] Futrell, R., Gibson, E., Tily, H. J., Blank, I., Vishnevetsky, A., Piantadosi, S. T., and Fedorenko, E. 2020.

[6] Gokaslan, A. and Cohen, V. OpenWebText Corpus.

[7] Goodkind, A. and Bicknell, K. 2021.

[8] Hahn, M., Futrell, R., Levy, R., and Gibson, E. 2022.

[9] Kennedy, A. and Pynte, J. 2005.

[10] Kretzschmar, F., Schlesewsky, M., and Staub, A. 2015.

[11] Levy, R. 2008.

[12] Luke, S. G. and Christianson, K. 2018.

[13] Norris, D. 2006.

[14] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. 2019.

[15] Rasmussen, N. E. and Schuler, W. 2018.

[16] Rayner, K., Ashby, J., Pollatsek, A., and Reichle, E. D. 2004.

[17] Shain, C. 2019.

[18] Shain, C., Blank, I. A., Fedorenko, E., Gibson, E., and Schuler, W. 2022.

[19] Shain, C., Meister, C., Pimentel, T., Cotterell, R., and Levy, R. P. 2022.

[20] Shain, C. and Schuler, W. 2022.

[21] Smith, N. J. and Levy, R. 2011.

[22] Smith, N. J. and Levy, R. 2013.

[23] Staub, A. 2011.

[24] Staub, A. 2015.

[25] Staub, A., White, S. J., Drieghe, D., Hollway, E. C., and Rayner, K. 2010.

[26] Szewczyk, J. M. and Federmeier, K. D. 2022.

[27] Van Schijndel, M. and Linzen, T. 2021.

[28] Wilcox, E. G., Gauthier, J., Hu, J., Qian, P., and Levy, R. 2020.

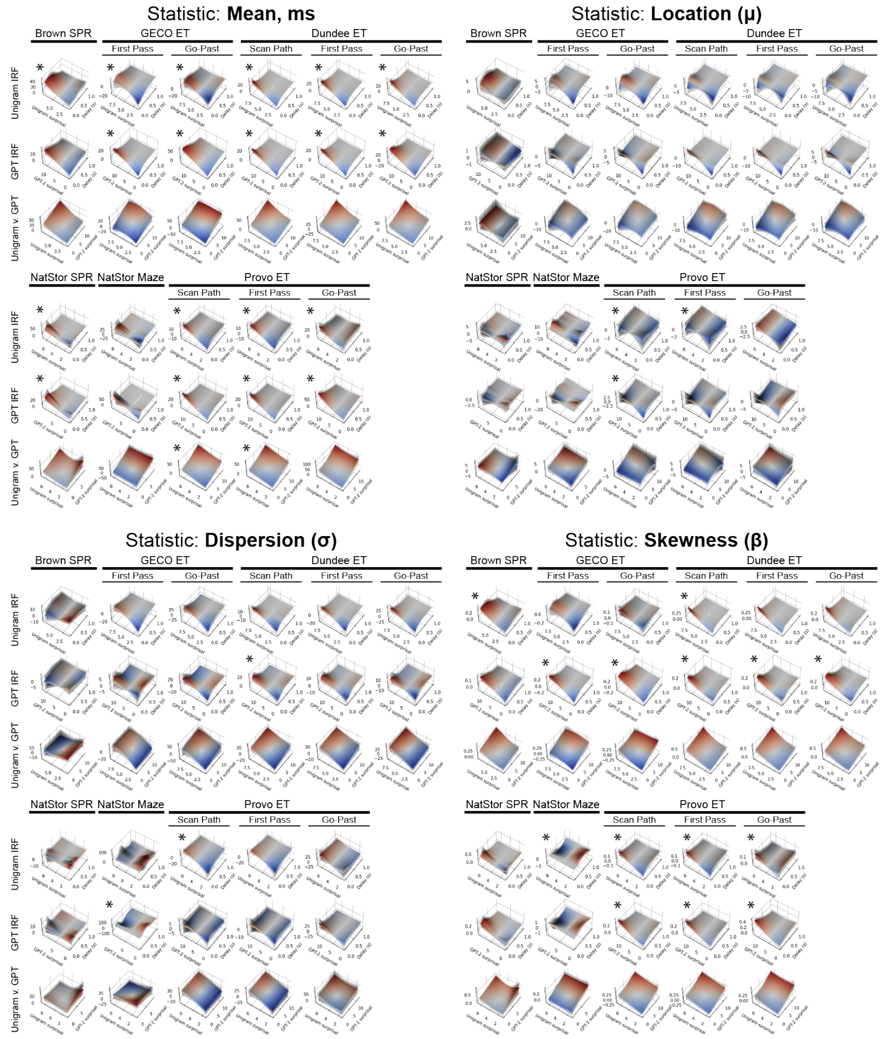

Figure 1: Model estimates across datasets for four exGaussian distributional statistics: mean (top left), location (μ, top right), dispersion (σ, bottom left), and skewness (β, bottom right). Plots show 3D surfaces representing the estimated change in the critical parameter (vertical axis) as a function of two key variables (horizonal axes), with 95% credible intervals plotted as gray bars. Significant unique effects are marked with $*$.