What Makes a Sentence Hard to Read? Disentangling the Contributions of Syntax

Diego Alves (diego.alves@uni-saarland.de)[1]

[1]Department of Language Science and Technology, Saarland University

Syntactic complexity has long been associated with cognitive processing difficulty. Psycholinguistic accounts have extensively investigated the mechanisms underlying syntactic processing, including dependency formation, retrieval demands, and locality constraints. At the same time, measures used to predict reading difficulty often emphasize lexical properties, motivating further investigation of how sentence-level syntactic complexity contributes to processing difficulty. In this study, we investigate how sentence-level syntactic complexity predicts reading difficulty, operationalized as total fixation time at the sentence level, computed as the sum of fixation durations for all tokens in the sentence, in the OneStop English eye-tracking corpus [1]. We compare constituency-based measures, including tree depth, number of nodes, average branching factor, and entropy reduction [6], with dependency-based measures such as dependency length [3, 5], computed using both Universal Dependencies (UD) [2] and Surface-Syntactic Universal Dependencies (SUD) [4]. Our objective is to determine which syntactic measures best explain sentence-level reading difficulty beyond lexical frequency, surprisal, and sentence length.

We analysed 1,580 sentences parsed with Stanza [8] and evaluated the contribution of each complexity measure using separate sentence-level linear mixed-effects regression models. For each complexity measure, we compared a model including the measure against a baseline model without the complexity measure, while controlling for lexical frequency, GPT-2 surprisal, word length, sentence length, and sentence position. Model performance was assessed using 10-fold cross-validation and changes in log-likelihood relative to the baseline. Entropy reduction was computed from a probabilistic context-free grammar induced from the Penn Treebank [7], while dependency length was calculated from both UD and SUD dependency parses.

Our results show that syntactic complexity contributes to sentence-level reading difficulty, but not all measures are equally informative. Entropy reduction, number of nodes, and dependency length based on UD significantly improved prediction of reading time, whereas tree depth, average branching factor, and dependency length based on SUD did not as shown in Figure 1. Furthermore, among the complexity measures showing positive contributions in the individual models, we evaluated whether additional measures provided incremental predictive value beyond entropy reduction (Figure 2). Adding UD dependency length to the entropy reduction model provided complementary predictive information, whereas adding number of nodes did not improve the model, suggesting substantial overlap with entropy reduction. These findings support a multi-dimensional view of syntactic complexity and highlight the importance of combining structural and information-theoretic measures when modelling sentence-level cognitive processing difficulty.

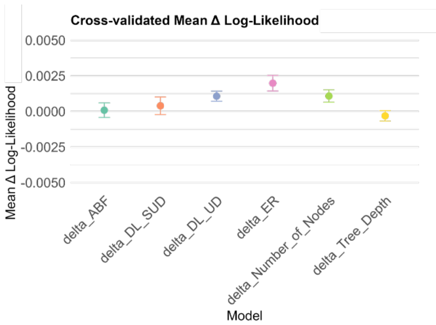

Figure 1: Mean cross-validated delta log-likelihood per sentence for each extended model relative to the baseline. Error bars indicate ±1 standard error across sentences, showing the improvement in predicting reading time provided by each complexity measure.

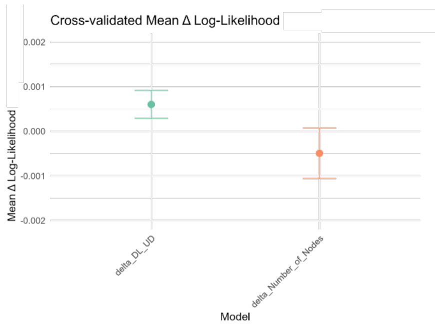

Figure 2: Mean cross-validated delta log-likelihood per sentence for models including DL_UD or Number_of_Nodes relative to the baseline model with entropy reduction. Error bars indicate ±1 standard error across sentences, reflecting the additional improvement in predicting reading time provided by each complexity measure.

References

[1] Yevgeni Berzak, Jonathan Malmaud, Omer Shubi, Yoav Meiri, Ella Lion, and Roger Levy. Onestop: A 360-participant english eye tracking dataset with different reading regimes. *Scientific Data*, 2025.

[2] Marie-Catherine De Marneffe, Christopher D Manning, Joakim Nivre, and Daniel Zeman. Universal dependencies. *Computational linguistics*, 47(2):255–308, 2021.

[3] Richard Futrell, Kyle Mahowald, and Edward Gibson. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112 (33):10336–10341, 2015.

[4] Kim Gerdes, Bruno Guillaume, Sylvain Kahane, and Guy Perrier. Sud or surface-syntactic universal dependencies: An annotation scheme near-isomorphic to ud. In *Proceedings of the second workshop on Universal Dependencies (UDW 2018)*, pages 66–74, 2018.

[5] Edward Gibson. Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68(1): 1–76, 1998.

[6] John Hale. Uncertainty about the rest of the sentence. *Cognitive science*, 30(4):643–672, 2006.

[7] Mitch Marcus, Beatrice Santorini, and Mary Ann Marcinkiewicz. Building a large annotated corpus of english: The penn treebank. *Computational linguistics*, 19(2):313–330, 1993.

[8] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. Stanza: A python natural language processing toolkit for many human languages. In *Proceedings of the 58th annual meeting of the association for computational linguistics: system demonstrations*, pages 101–108, 2020.