Syntactic Node Count as Index of Predictability

Kohei Kajikawa[1], Shinnosuke Isono[2]

[1]Georgetown University, [2]National Institute for Japanese Language and Linguistics (NINJAL)

kk1571@georgetown.edu

Recent neurolinguistic studies of sentence comprehension often use *Node Count* (NC, see Figure 1), the number of syntactic nodes built at each word, as the linking hypothesis connecting word-by-word syntactic structure building and neural activity [e.g., 1]. The assumption behind the use of NC is that syntactic structure building incurs cognitive effort. If so, larger NC should lead to increased processing difficulty. However, several studies using English and Japanese datasets have reported that higher dependency-based NC is associated with *faster* reading times [2, 3, 4]. These findings challenge the interpretation of NC as an index of processing effort and instead point toward an alternative explanation in terms of predictability or anti-locality effects [5], because *NC is correlated with the amount of linguistic context that is strongly related to the current word syntactically and/or semantically.* Building on these findings, we ask the following questions: (i) Is the effect of NC on reading times negative across datasets and grammar formalisms? (ii) If so, do NCs based on different formalisms have independent effects? (iii) If syntactic Node Count indexes predictability, is it independent from large language model surprisal?

Material and Methods. Eye-tracking data from Dundee and self-paced reading data from Natural Stories are analyzed. Reading times are averaged across participants, and spillover from two preceding regions is considered. Five variants of NC are considered based on widely adopted grammars and strategies: Penn Treebank-style phrase structure grammar (PSG) using top-down, bottom-up, and left-corner parsing strategies; Universal Dependencies-style dependency grammar; and incremental parses of Combinatory Categorial Grammar (CCG). Unlike previous works studying CCG NC [e.g., 4], we only count binary operations since the status of unary rules as syntactic operations is debated. To answer (i), we first conduct nested model comparison using linear regression models given in (1). For each variant of NC, its contribution to the model's goodness of fit is evaluated by 10-fold cross-validation comparing a baseline model (2) with the target model (3) containing the NC. Significance of improvement in log likelihood is evaluated by a paired permutation test (4). We then look at the signs of the coefficients of the significant NCs. To address (ii), the best model is searched using AIC, starting from a model containing all variants of NC found to be significant in (i) plus the baseline predictors. For (iii), NCs that are significant in (i) are re-evaluated against a baseline augmented with GPT-2 surprisal.

Results. (i) We find significant effects of NCs based on PSG bottom-up, dependency, and CCG (Figure 2). Significant NCs consistently exhibit *negative* coefficients (Figure 3) in the w_i and w_{i-1} regions. Top-down and left-corner NCs failed probably because predictive structure building makes NCs less directly linked to the amount of contextual information. The best models selected in (ii) show that NCs based on different grammars can have independent effects, with Dependency being the most pronounced (Table 1). (iii) Controlling for GPT-2 surprisal had little impact on the results (Figure 2).

Discussion. Our results indicate that NC functions as a proxy for predictability, rather than structure-building complexity as usually assumed. In addition, that NCs from different grammar formalisms explain unique variance in different timing supports the view that structural processing is shaped by multiple representations. Moreover, the persistence of some effects after controlling for GPT-2 surprisal suggests that structural cues contribute to predictive processing under cognitive resource constraints.

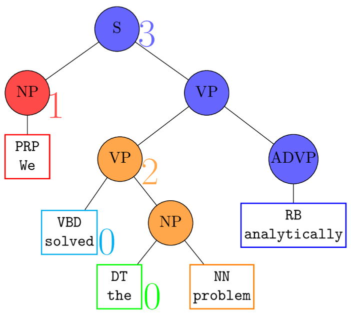

Figure 1: Example of Node Count under a phrase structure bottom-up parsing.

$$y_i = \mathbf{x}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2), \quad y_i = \mathsf{RT}_i \tag{1}$$

$$\begin{aligned}\mathbf{x}_{i,\text{base}} = [&\text{zone}_i, \text{position}_i, \text{wlen}_i, \text{unifreq}_i, \text{wlen}_i \cdot \text{unifreq}_i,\\ &\text{wlen}_{i-1}, \text{unifreq}_{i-1}, \text{wlen}_{i-1} \cdot \text{unifreq}_{i-1},\\ &\text{wlen}_{i-2}, \text{unifreq}_{i-2}, \text{wlen}_{i-2} \cdot \text{unifreq}_{i-2}]\end{aligned} \tag{2}$$

$$\mathbf{x}_{i,\text{targ}} = [\mathbf{x}_{i,\text{base}}; \text{NC}_i, \text{NC}_{i-1}, \text{NC}_{i-2}] \tag{3}$$

$$\Delta\text{loglik}_i = \log P(y_i \mid \hat{y}_{\text{targ},i}, \hat{\sigma}_{\text{targ}}) - \log P(y_i \mid \hat{y}_{\text{base},i}, \hat{\sigma}_{\text{base}}) \tag{4}$$

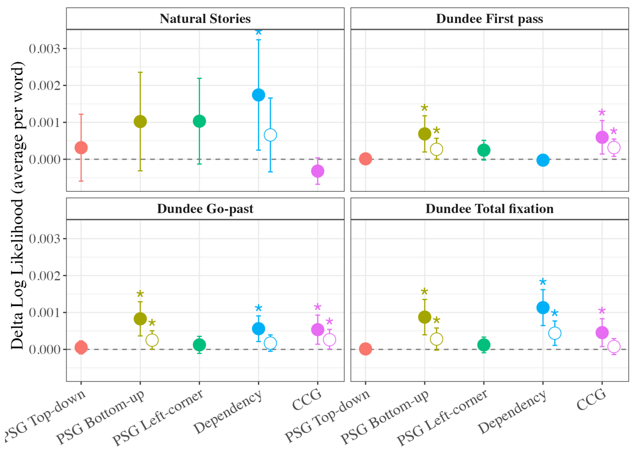

Figure 2: Predictive power of each Node Count. White-filled points indicate results after controlling for GPT-2 surprisal. The star means the statistical significance, where $\alpha = 0.05$ with correction accross datasets.

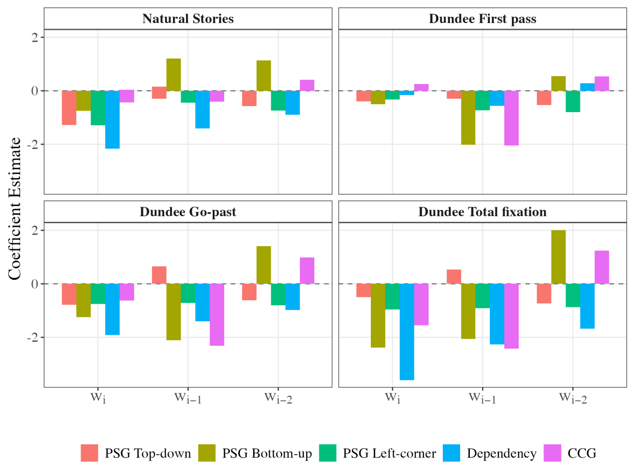

Figure 3: Regression coefficients of each Node Count, averaged over 10-fold CV. Negative values indicate shorter reading times for higher Node Count.

	PSG Bottom-up	Dependency	CCG
Natural Stories	–	*** (w_i)	–
Dundee First-pass	*** (w_{i-1})	–	n.s.
Dundee Go-past	n.s.	** (w_i)	*** (w_{i-1})
Dundee Total fixation	** (w_i) / * (w_{i-2})	*** (w_i) / *** (w_{i-1}) / *** (w_{i-2})	** (w_{i-1})

Table 1: Significance of Node Counts in the best models on held-out test data.

References

[1] Brennan, J. R., Stabler, E. P., Wagenen, S. E. V., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, *157–158*, 81–94.

[2] Asahara, M., Ono, H., & Miyamoto, E. T. (2019). BCCWJ-EyeTrack. *Gengo Kenkyu*, *156*, 67–96.

[3] Isono, S. (2024). Category locality theory: A unified account of locality effects in sentence comprehension. *Cognition*, *247*, 105766.

[4] Kajikawa, K., Yoshida, R., & Oseki, Y. (2024). Dissociating Syntactic Operations via Composition Count. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *46*, 297–305.

[5] Konieczny, L. (2000). Locality and parsing complexity. *Journal of Psycholinguistic Research*, *29*, 627–645.