Sentence and Clause Boundary Effects Challenge Strict Incrementality in Sentence Comprehension

Ondřej Drobil (ondrej.drobil@ling.uni-stuttgart.de)[1], Titus von der Malsburg[1]
[1]Institute of Linguistics, University of Stuttgart

Sentence processing is widely assumed to be aggressively incremental, with each word integrated upon encounter. This paradigm is most prominently represented by surprisal-based accounts, which formalize processing difficulty as a function of a word's conditional probability in context [3]. Although earlier models assuming that integration occurs at syntactic boundaries have been empirically rejected, the "wrap-up effect" – an increase in processing times at the end of clauses or sentences – persists as a "specter" that leads researchers to exclude sentence-final data or avoid placing target regions in final positions [4]. Yet, direct empirical evidence for wrap-up remains sparse and its underlying causes are poorly understood [7]. Recent findings from Chinese even challenge the very existence of sentence wrap-up [8]. In this study, we analyzed the PoTeC corpus [2], a large-scale eye-tracking dataset of German scientific texts (N=12) read by 75 students to establish: (A) whether the wrap-up effect exists in German passages reading and (B) what variance it explains (if any) over and above surprisal and other word properties.

Method: Using linear mixed-effects models, we analysed reading times on all individual words in the corpus. We modeled the effects of word position (four binary predictors: clause-final vs. not, clause-initial vs. not, sentence-final vs. not, and sentence-initial vs. not) on gaze durations and go-past times while controlling for GPT-2 surprisal. Additional controls included word-level (length, frequency, content vs. function words, expert term), and sentence-level (index), and text-level variables (line breaks), as well as spillover effects (fixations, gaze duration and the difference between gaze duration and go-past-time on the previous word) and reader expertise.

Results: Gaze durations were significantly longer on sentence-final words (β=0.067, t=2.341) and clause-initial words (β=0.067, t=3.080). In contrast, durations on sentence-initial words did not differ significantly from the average, neither did clause-final words. The impact of sentence-final position on gaze duration was further modulated by reader expertise. Results for go-past times revealed a more pervasive wrap-up effect, appearing at both sentence ends (β=0.297, t=7.598) and clause ends (β=0.141, t=5.132). Go-past times were also significantly longer at clause-initial positions (β=0.060, t=2.301). Finally, GPT-2 surprisal significantly predicted both gaze durations (β=0.088, t=13.832) and go-past times (β=0.113, t=15.107).

Discussion: In addition to the previously attested sentence wrap-up effects in naturalistic reading of connected texts [5], we observed delays even at clause boundaries, not only at the ends but also at the beginnings which we suggest may reflect ahead-of-time structure building. Clause and sentence boundary effects explained significant variance in reading times above and beyond GPT-2 surprisal, therefore challenging strict incrementality. While previous work [1, 6] already demonstrated limitations of surprisal in explaining processing times in garden-path sentences, our findings extend the limits of surprisal to the much more pervasive processing of syntactic clause boundaries. Whether wrap-up reflects a qualitatively distinct process triggered by syntactic boundaries or a non-linear accumulation of incremental load remains a question for future experimental inquiry.

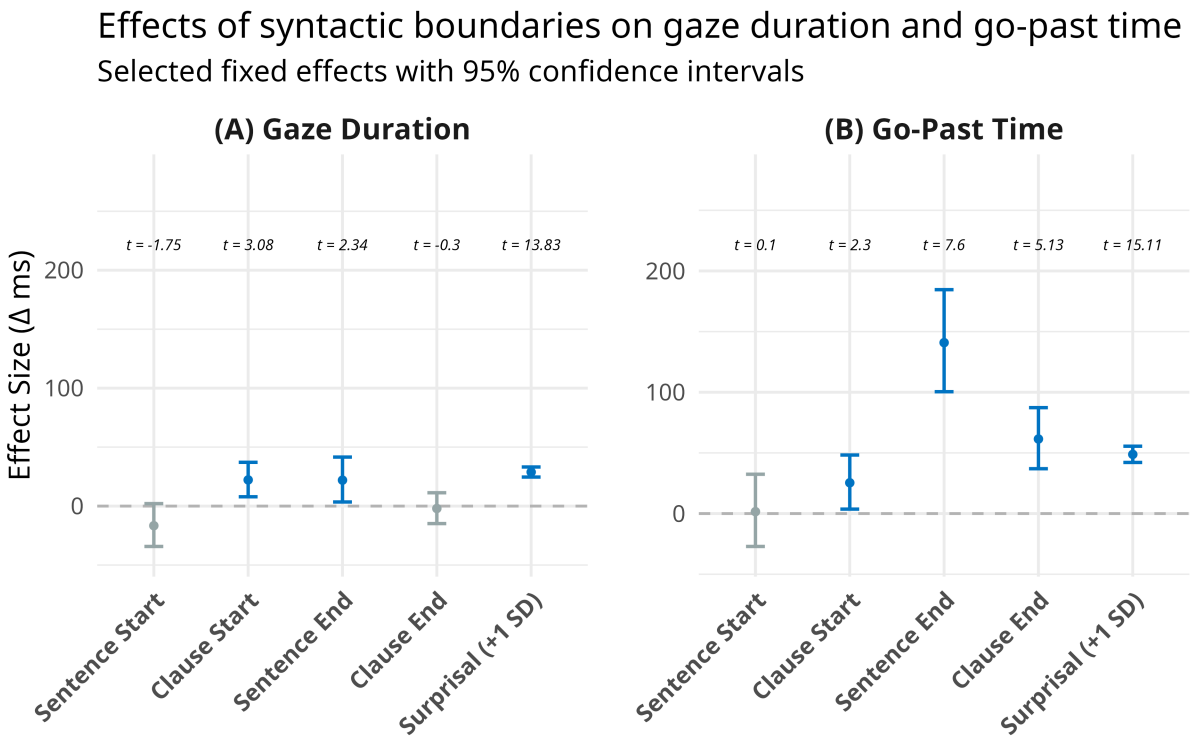

Figure 1: Panel (A) displays effects on gaze duration, while Panel (B) shows effects on go-past time. Points represent estimated fixed effects and their sizes in ms relative to the model baseline (the dashed line at 0 ms), with 95% confidence intervals, where blue markers indicate statistically significant effects and grey markers indicate non-significant effects. Effect sizes in ms were derived by back-transforming log-linear coefficients. In addition to the reported effects, the following covariates were included in the models: line break, previous word's fixation, expert technical term, general technical term, content word, previous word's gaze duration, previous word's difference between go-past time and gaze duration, word index in sentence, sentence index in text, word length, lemma frequency, and level of expertise and its interactions with boundary effects). Contrasts for binary predictors were set to -0.5 and 0.5. Continuous variables were centered and scaled.

References: [1] Huang et al., 2024, JML; [2] Jakobi et al., 2025, BRM; [3] Levy, 2008, Cognition; [4] Stowe et al., 2018, Cognition; [5] Tiffin-Richards & Schroeder, 2018, JEP:LMC; [6] Van Schijndel & Linzen, 2021, CogSci; [7] Warren et al., 2009, Cognition; [8] Xue et al., 2026, QJEP